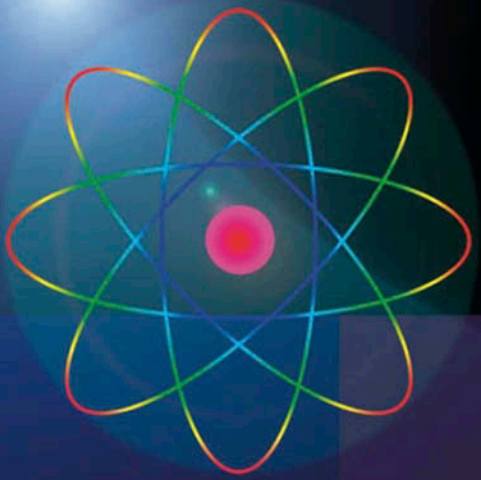
Nuclear Electronics

Superconducting Detectors and Processing Techniques

Vladimir Polushkin



WILEY

Nuclear Electronics

Superconducting Detectors and Processing Techniques

Vladimir Polushkin
High Wycombe, UK

John Wiley & Sons, Ltd

Nuclear Electronics

Nuclear Electronics

Superconducting Detectors and Processing Techniques

Vladimir Polushkin
High Wycombe, UK

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some of the content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Polushkin, Vladimir.
Nuclear electronics : superconducting detectors and processing techniques /
Vladimir Polushkin.
p. cm.
Includes bibliographical references and index.
ISBN 0-470-85759-5 (cloth : alk. paper)
1. Superconductors. 2. Nuclear counters. 3. Molecular electronics.
4. Nanotechnology. I. Title.
TK7872.S8P65 2004
539.7′7—dc22

2004004944

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-85759-5

Typeset in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by TJ International Ltd, Padstow Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

The nuclear electronics is often associated primarily with the fundamental research conducted in nuclear laboratories and astrophysics. Yet many would be quite surprised to find out an extent it directly affects our everyday life. The food safety, the security at airports, medical examinations, doping tests, forensic investigations and so on are all provided with the nuclear analytical instrumentation. The technological advance in the life science, biology, ecology, medicine, semiconductors, remote sensing and many others relay heavily upon the information which comes uniquely and, what is particularly important, non-destructively from nuclear sensors. This information can be either a chemical composition of a material dealt with, or a crystal or atomic structure, or the thickness of individual layers in a multilayer formation of a complex shape, or the mass of molecules, or a DNA structure. The number of possible applications is wide and expanding rapidly almost on a daily basis. Besides, even in traditional areas of the research and development specialists need to observe smaller features, a lesser percentage of contaminations, etc., but, at the same time, scanning larger areas or volumes. Therefore, the nuclear electronics as a discipline must continuously evolve in its search for new detection methods or more elaborate processing techniques in order to deliver much needed systems with a better resolution, accuracy, reproducibility and the speed of operation.

A big step in the radical advance of the nuclear analytical instrumentation will undoubtedly be the commercialisation of cryogenic particle and radiation detectors. Leading representatives of these devices are the Superconducting Tunnel Junction (STJ) and quantum calorimeters. Both types have already demonstrated a sub 10 eV energy resolution FWHM (the full width at the half maximum) at 5.9 keV (MnKα) as opposite to 125 eV achieved with the best semiconductor detectors. Initial laboratory

experiments have proven the new class of detectors to be very accurate, stable, adequately fast and well suited for the high-density multipixel integration. The only disadvantage associated with the superconducting technology was a need in liquid cryogens as the only means to ensure the sub 1 K temperature environment required for their operation. Besides the inconvenience and the maintenance cost, the presence of a cryogen or its evaporated gas could interfere adversely with some of technological processes which instruments were supposed just to monitor. A recent commercial introduction of "dry" pulse-tube coolers, however, has addressed the problem and, in fact, changed priorities in the high-precision nuclear measurements. One can see that the industry started diverting a part of the development budget into advancing these new technologies. Despite the extra cost associated with cryogenics, an added value such as a more than one order of magnitude increase in resolving capabilities fully justifies it.

To summarise this brief preface, it is easy to predict that an increasing number of scientists, engineers and post-graduate students is expected to be involved in the nuclear measurements based on cryogenic detectors who may wish to update or refresh their skills on the subject. Whereas the information on sensors and the physics behind them is scattered in thousands of papers, some system integration issues have not been addressed so far at all. The prime goal of this book is to systematise the published material on the subject of interest and possibly fill in existing gaps. We will avoid reprinting complex expressions of the microscopic superconductivity or the information theory. For those interested in gaining a deeper background in these and some other relevant issues, appropriate references will be given throughout the book.

The book is primarily based on research which was done in collaboration with my colleagues at the Joint Institute for Nuclear Research (Dubna), Physikalisch-Technische Bundesanstalt (Berlin) and Thin Film Group of Oxford Instruments (Cambridge) as well as a result of multiple discussions of the matter with my friends and colleagues from other institutions worldwide. I would like to express my gratitude to all of them hoping the book will be useful in their further research.

1

Detection Methods with Cryogenic Particle and Radiation Sensors

INTRODUCTION

There are two types of general detection methods that are commonly utilised in the majority of nuclear measurements. The first one involves a variety of the quasiparticle detectors. A good example of such a device in semiconductors is an ionisation Silicone Drift Detector (SDD) [1]. The operation principle of all quasiparticle detectors is generally based upon counting a number of charged quasiparticles, for instance electrons, created by the energy of an incident radiation in the sensitive volume. The number of freed quasiparticles is assumed to be proportional to an amount of the energy absorbed by a sensor. Another class of devices represents various types of calorimeters. All the calorimeters incorporate three major components: an absorber, a thermometer and a link to the heat sink. The function of the absorber is to transform an incident energy into heat. A temperature increment, ΔT, due to this added heat is detected by a thermometer. Unlike in the particle detectors, the amount of the energy absorbed by calorimeters can be restored only by the integration of the output of the thermometer during the period of time from the moment when the radiation event had started the interaction with the absorber until the system has come back to its original equilibrium state. Therefore, they are normally a bit slower in the operation

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin
 ISBN: 0-470-85759-5

than their quasiparticle counterparts, but at the same time potentially more accurate being free from statistical limitations always associated with random discrete processes.

Both types of detection methods exist also in the superconductor nuclear electronics. The Superconducting Tunnel Junctions (STJs) represent the quasiparticle detection method. Together with the Quantum Calorimeters based on the Transition Edge Sensor (TES) they constitute the most mature of all known cryogenic detection techniques, which are suitable for the immediate implementation in the industrial analytical instrumentation. Therefore, the first three sections of the chapter will be concerned with the principles of these two devices. The fourth section introduces some other types of cryogenic detectors which are still at an early development stage, but promise to become the device of choice for some particular applications.

1.1 QUASIPARTICLE DETECTORS: INTERACTION OF NUCLEAR RADIATION WITH SUPERCONDUCTORS

The prime goal of nuclear measurements is to derive energy E, which an incident particle or a radiation deposits into the sensitive volume of a detector with the best possible accuracy within an established time constraint. In the case of quasiparticle detectors, E generates a number N_0 of discrete charged carriers (ions in gas chambers, electrons and holes in semiconductors, quasiparticles in STJs). Provided that an average energy portion ε spent on the creation of a single quasiparticle (or a pair of them) is known and it is fairly constant in the whole energy range of interest, the measurement procedure results in collecting and counting these N quasiparticles. The creation of particles is obviously a random process. Therefore, the number N_0 is the mean value characteristic of this process. Statistical fluctuations in this number define the intrinsic limiting resolution of nuclear measurements with this kind of detectors.

Let us assume that the random process follows the Poisson statistics. The standard deviation of N_0 has the form $\sigma = \sqrt{N_0}$. The intrinsic statistical resolution of a detector is given by [2]

$$R = \frac{2.35}{\sqrt{N_0}} \tag{1.1.1}$$

The equation (1.1.1) shows that the ultimate accuracy of a quasiparticle detector depends on its ability to produce as many charged carriers as possible. One can state that the whole development history of this type of sensors involves the persistent search for materials with the smallest possible excitation energy ε in order to maximise N_0.

In the early 1970s, scientists were thrilled by an opportunity to reduce ε from approximately 20 eV in the gas proportional counters to about 3 eV in solid-state semiconductor devices. Superconductor materials reduce the figure even further by as much as three orders of magnitude. This leads to exciting possibilities of achieving an unprecedented accuracy of nuclear measurements, as shown in Table 1.1.

In the "normal" state, that is, above the superconducting transition temperature T_c, all electrons behave according to the conventional Fermi statistics. They occupy single particle Bloch states determined by their momentum and the spin orientation according to the Pauli exclusion principle. When the temperature drops below T_c the physical picture changes dramatically. The thermal fluctuations of the lattice reduce to a level when its oscillations (phonon waves) produced by the interaction of electron clouds with positively charged lattice ions become pronounced. Each phonon wave can be picked up by another electron with the opposite spin, so that both of them fall for some period of time into a synchronism. The phenomenon thereby forms the so-called Cooper pair. The pairing of electrons is energetically favourable. This can be seen in Figure 1.1 where the energy diagrams of a superconductor are presented at a temperature $T = 0\,\mathrm{K}$ and at $0 < T < T_c$, correspondingly.

Table 1.1 Properties of materials that are utilised in different types of quasiparticle detectors: proportional counters, semiconductor solid state ionisation detectors and superconductor detectors. ε is the average energy to create one charged carrier in the sensitive volume of a detector, N_0 is a mean number of carriers created by an X-ray photon with energy 5.9 keV, the Fano factor is equal to $\frac{\mathrm{Var}(N_0)}{N_0}$

Material	ε, eV	N for MnKα (5.9 keV)	Fano factor	Intrinsic Energy resolution (eV)
Argon gas	26	30	0.2	1000
Si	3.65	6000	0.14	115
Nb	1.5×10^{-3}	2.2×10^{6}	0.22	4

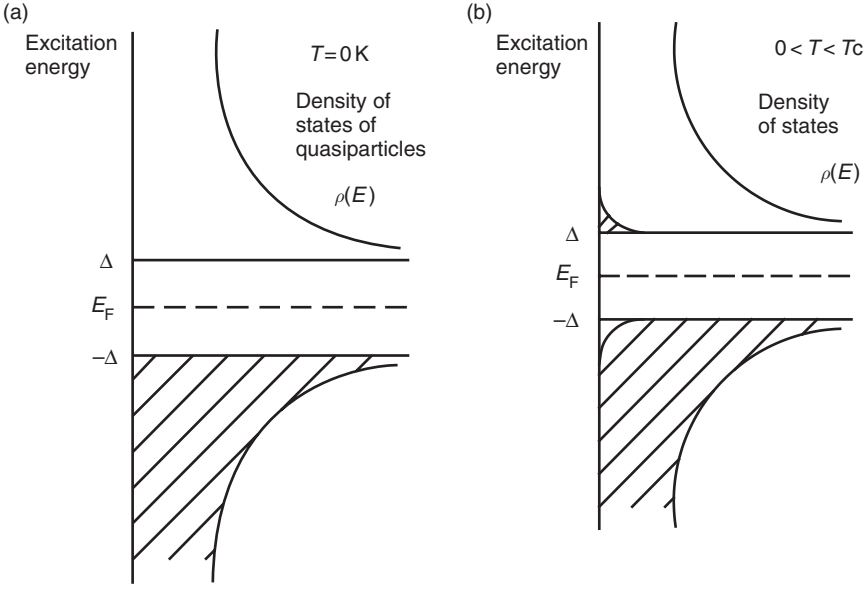


Figure 1.1 Formation of Cooper pairs is an energetically favourable state. In the energy diagrams of a superconductor, which correspond to temperatures $T = 0\,\mathrm{K}$ (a) and $0 < T < T_c$ (b), the dashed areas represent occupied states. In the finite temperature range, the pair-breaking process is continuously going on due to thermal excitations. The process creates quasiparticles with energies above $E_F + \Delta$, where Δ is the energy gap. In superconductors, Cooper pairs represent a fundamental state, whereas the presence of quasiparticles corresponds to an excited non-equilibrium state. Without artificial excitations, the average number of quasiparticles in a superconductor is proportional to $\exp(-\alpha/k_b T)$, where α is a factor depending on the energy gap of the material and k_b is the Boltzmann constant

The dashed area represents occupied states, which are below the Fermi level, E_F, and the energy gap Δ is given by

$$\Delta = h\omega_c e^{-\frac{1}{N(0)V_{ip}}} \tag{1.1.2}$$

where h is the Planck constant, $h\omega_c$ represents the Debye maximum phonon energy, $N(0)$ is the density of states of electrons at the Fermi level above T_c and V_{ip} denotes the interaction potential between two electrons forming the Cooper pair.

An essence of the detection principle for a superconductor particle detector is the fact that an incident particle or a radiation photon injects its energy portion into a superconducting absorber. This energy breaks a certain number of Cooper pairs through the ionisation process and the

phonon relaxation process. A number of created quasiparticles is proportional to the amount of the deposited energy in the limit of the statistical error. The value of the deposited energy can be recovered if a primary sensor of the detector unit is capable to distinguish between free electrons and those remaining in pairs. Figure 1.1a suggests the basic detection principle. It involves a potential barrier, which is set up on the way of diffused charged carriers. If the energy height of the barrier is just below $E_F + \Delta$, it becomes transparent for single quasiparticles, whereas pairs with the energy $E_F - \Delta$ will be filtered out. Figure 1.2 illustrates the process schematically. The total amount of charge integrated on the right side of the barrier is the measure of the deposited energy.

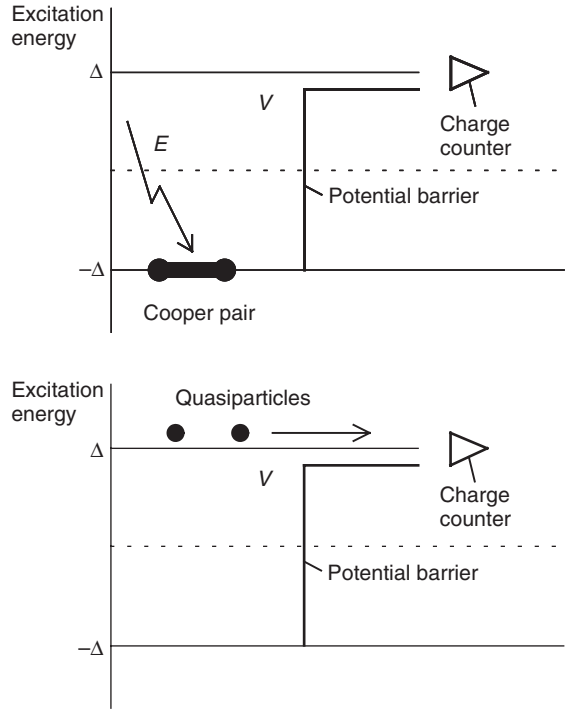


Figure 1.2 The basic detection principle of cryogenic quasiparticle detectors. An unperturbed Cooper pair does not have enough energy to climb over the potential barrier, whereas the quasiparticles formed by a radiation event propagates freely above the barrier towards a charge counter. The number of counted quasiparticles is proportional to the amount of energy deposited into the superconductor absorber. The temperature of operation must be such that the number of particles created as a result of temperature excitations is much lower compared to the number of quasiparticles generated as a result of interaction of the absorber with a phonon

A distinctive feature of the superconductor quasiparticle detectors compared to their semiconductor ionisation counterparts is the ability to respond to phonon excitations. It is known that an incident event (whether it is a radiation or an impact with a particle) mostly dissipates its energy via the interaction with the lattice. The energy of produced phonons lies in the range of a fraction of meV to a maximum Debye value of $E_d = 60\,\text{meV}$. All created phonons would not have any impact on the output of a semiconductor ionisation detector due to its much larger forbidden energy gap (for instance, Germanium has $\Delta_{Ge} = 0.665\,\text{eV}$). An amount of wasted energy can well exceed tens of percents in the range of "soft" X-rays.

In the case of superconductors, phonons carry the pair-breaking capability as long as their energy is larger than 2Δ. This particular property creates entirely new opportunities in the nuclear measurements, which are uniquely attributed to cryogenic detectors. One such application is the time-of-flight mass-spectrometry of large mass molecules. As the mass of an analysed molecule (it can also be a cell or a microorganism) increases, semiconductor ionisation detectors (for instance, microchannel plates), which are currently used in this type of instrumentation, struggle to maintain their efficiency. In contrast, the number of quasiparticles generated in the cryogenic phonon detectors does not depend on the size and mass of the molecule. They remain being 100% efficient even in the mass range where microchannel plates do not produce any response at all. In fact, the only limitation on the maximum particle mass would be incurring a permanent mechanical damage to the absorber. An opportunity to register the arrival moment of a molecule and actually measure its kinetic energy with a high accuracy adds a new dimension to the mass spectrometry. This initiated the development of a new advanced class of instruments, which will be discussed later in the Chapter 5.

In what follows, we will summarise briefly the energy cascade in a superconductor. This will be helpful for performing the numerical simulations of particle detectors attached to superconductor absorbers. In the case of "soft" X-rays, the interaction with the sensitive volume of a detector undergoes primarily through the photoelectron absorption process [3,4]. In this process, X-ray photon deposits all its energy into an atom that ejects a photoelectron from an appropriate shell. The energy of the photoelectron is given by $E_{phe} = h\nu - E_b$, where $h\nu$ is the energy of the incident X-ray photon and E_b is the binding energy of the photoelectron in its original shell.

The primary electron emitted by the atom relaxes within 0.1 ps through the electron–plasmon interaction as well as the ionisation of other atoms if its energy is sufficient to do so. The secondary electrons and plasmons lose their energies in approximately 1 ps by emitting phonons, which break a number of Cooper pairs into separate quasiparticles. The relaxation as well as the recombination of quasiparticles create more phonons that break further still a larger amount of Cooper pairs. In about 2 ns, the absorber contains only the phonons which have no pair-breaking capability. They also relax on quasiparticles for some time, but as far as the detection process is concerned 2 ns means the end of the energy cascade of interest. This dynamics is summarised in the Rothwarf–Taylor (R–T) equations describing differential rates for quasiparticles $\left(\frac{\partial n}{\partial t}\right)$ and phonons $\left(\frac{\partial n^*}{\partial t}\right)$ [5]:

$$\frac{\partial(n_{\mathrm{th}} + n_{\mathrm{ex}})}{\partial t} = \tau_{\mathrm{i}}^{-1} + 2\tau_{\mathrm{b}}^{-1}(n_{\mathrm{th}}^* + n_{\mathrm{ex}}^*) - \tau_{\mathrm{r}}^{-1}(n_{\mathrm{th}} + n_{\mathrm{ex}})^2 - \tau_{\mathrm{q}}^{-1} n_{\mathrm{ex}}$$

$$\frac{\partial(n_{\mathrm{th}}^* + n_{\mathrm{ex}}^*)}{\partial t} = \frac{\tau_{\mathrm{g}}^{-1}(n_{\mathrm{th}} + n_{\mathrm{ex}})^2}{2} - \tau_{\mathrm{c}}^{-1}(n_{\mathrm{th}}^* + n_{\mathrm{ex}}^*) - \tau_{\mathrm{p}}^{-1} n_{\mathrm{ex}}^* \qquad (1.1.3)$$

where n_{th} and n_{ex} represent the number density of thermal and excess quasiparticles respectively, n_{th}^* and n_{ex}^* are responsible for the number density of thermal and excess photons respectively, τ_{i}^{-1} is the rate at which quasiparticles are being produced by the primary interaction of the radiation with an absorber, τ_{r}^{-1} is the rate of the quasiparticle recombination back into the Cooper pairs, τ_{g}^{-1} is the rate of the phonon generation by the recombined quasiparticles, τ_{c}^{-1} and τ_{b}^{-1} denote the rate of generation of phonons having the pair-breaking capability and the rate of the quasiparticle production by these phonons, τ_{q}^{-1} and τ_{p}^{-1} are escape rates for quasiparticles (including tunnelling) and phonons respectively. A schematic illustration of these equations is shown in Figure 1.3 [6].

The coefficients of the R–T equations starting from a certain value of the incident energy (depending of the absorber design) become increasingly non-linear. This is due to non-equilibrium superconductivity effects, which are likely to develop in the vicinity of the spot where the primary interaction occurs. During the initial stage of the cascade the energy density and the quasiparticle concentration can be high enough to partially suppress the local energy gap. The gap suppression causes two parasitic effects: first, the variation in the number of initially created

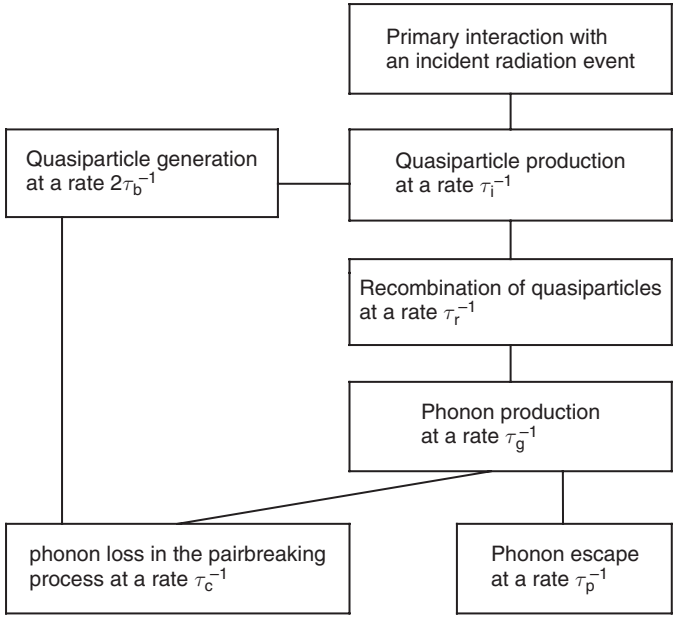


Figure 1.3 A block diagram illustrating schematically Rothwarf–Taylor equations. The energy cascade ends in approximately 2 ns after an incoming photon has deposited its energy into an absorber

quasiparticles; second, the trapping of quasiparticles with energy less than the unperturbed Δ. An example of such a trap is shown in Figure 1.4.

Quasiparticle traps were originally understood as a parasitic phenomenon. Later on, however, N. Booth proposed to design them intentionally

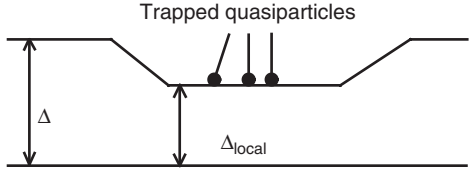


Figure 1.4 Local heat, non-equilibrium processes, defects and so on can partially suppress a local energy gap. Quasiparticles relaxed to Δ_{local} do not have enough energy to diffuse into the unperturbed superconductor. In other words, they become trapped in volumes with a lower energy gap. In general, this is a parasitic phenomenon, which may lead to the charge deficit in the detector response. However, designed intentionally just before the tunnel junction, it works as a carrier concentrator speeding up the tunnelling time (see section 1.2 for details)

[7]. This idea made a crucial impact on the practicality of superconductor radiation detectors, as we will see in section 1.2.

The probability of the photon absorption per atom is proportional to $\sim\frac{Z^5}{E^{3.5}}$. Ta with $Z = 73$ and $M_\alpha = 1.7\,\text{keV}$ is one of the most attractive materials for designing a superconducting absorber for quasiparticle X-ray detectors. The required thickness t_{abs} of the absorber can be found from the attenuation factor of the radiation intensity

$$\frac{I}{I_0} = \mathrm{e}^{-t_{\text{abs}}/\lambda_{\text{abs}}} \tag{1.1.4}$$

where I_0 and I represent intensities of the incident radiation and the radiation passed through the absorber respectively and λ_{abs} is the absorption length given in Figure 1.5 [8].

To conclude this section, we will evaluate the ultimate intrinsic energy resolution of a superconducting quasiparticle detector at FWHM. To do so we will assume that all quasiparticles created by an incident event can

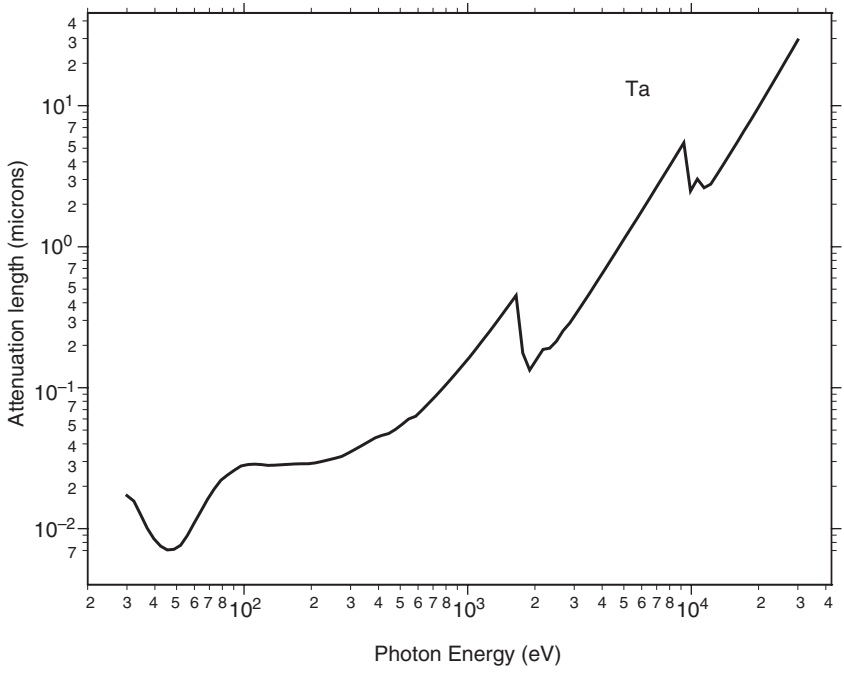


Figure 1.5 X-ray absorption length as a function of the photon energy of Ta. Ta with an atomic number $Z = 73$ and critical temperature $T = 4.5\,\text{K}$ is a very attractive material for designing highly efficient superconductor absorbers

be counted. Thus, on absorption of a photon or a phonon with energy E, the mean number of excess quasiparticles is given by

$$N_0 = \frac{E}{\varepsilon} \tag{1.1.5}$$

where ε is the mean energy required for creating two charge carriers of a Cooper pair. Provided that the energy cascade follows a Poisson statistics, a spectral response to a repetitive deposition of E with sufficient statistics is expected to be a Gaussian distribution as shown in Figure 1.6. FWHM is defined as the energy width of the Gaussian spectral response coming from the monochromatic radiation at a half-height relative to the maximum energy. For any quasiparticle detector, the energy resolution FWHM is defined as

$$\Delta E_{\mathrm{FWHM}} = 2.35 \cdot \sqrt{\varepsilon F E} \tag{1.1.6}$$

where $F = \frac{\mathrm{Var}(N_0)}{N_0}$ is the Fano factor. The Fano factor takes into consideration the fact that the energy cascade contains partially correlated events and, strictly speaking, the entire process does not follow precisely the Poisson statistics. It is common practice fitting measured spectra

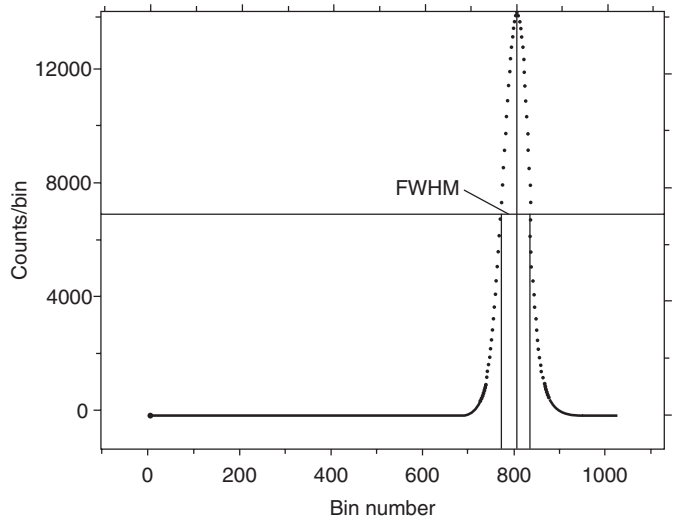


Figure 1.6 A spectral response of a detector to the repetitive deposition of the photon energy E into its absorber. With sufficient statistics, the response has a quasi-Gaussian distribution. The FWHM is defined as the energy width of the Gaussian spectral peak produced by the monochromatic radiation at the half-height relative to the maximum energy.

Table 1.2 Properties of materials, which are commonly used in fabrication of superconductor absorbers

Material	NB	TA	AL
Atomic number	41	73	13
Energy gap, Δ (meV)	1.53	0.74	0.18
Average quasiparticle creation energy, ε (meV)	2.52	2	0.31
Intrinsic energy resolution at 5.9 keV (MnKα) FWHM (eV)	4	3.5	1.5
Fano factor	0.2	0.2	0.22
Superconducting transition temperature, K	9.25	4.48	1.2

with the Gaussian distribution curve and using ΔE_{FWHM}. In doing so however, one needs to always remember the error present in this procedure.

In Table 1.2, we present the intrinsic energy resolution together with some other useful parameters for materials widely used in the fabrication of quasiparticle detectors.

1.2 SUPERCONDUCTING TUNNEL JUNCTION DETECTORS

As we discussed in section 1.1, the number of excess quasiparticles in the superconducting absorber is the measure of the incident photon or phonon energy E. These quasiparticles exist during a certain period of time, called lifetime, at the end of which they tend to recombine back into the Cooper pairs. In order to restore a value of E, we require an efficient technique to identify quasiparticles, separate them from the Cooper pairs and to deliver them to a read-out circuit. One such technique utilises an S–I–S (Superconductor–Insulator–Superconductor) tunnel junction (the STJ). Figure 1.7 demonstrates schematically a possible detector configuration involving a relatively large Ta absorber attached to a Ta–Al–AlO_x–Al–Ta STJ. An epitaxial Ta absorber is deposited and patterned on a SiO_x substrate. The next step in the manufacturing process is to sputter and pattern an Al layer on top of it. The area of Ta absorber is covered by an Al layer that constitutes the base electrode of the junction. In order to form a high quality, 1–2-nm thick isolator (I) barrier, the Al layer is then oxidised in well-controlled conditions in terms of the oxygen pressure, temperature and time. The procedure has a strong effect on the quality of the junction as a whole. Any pinhole, the non-uniformity and stresses in the dielectric will

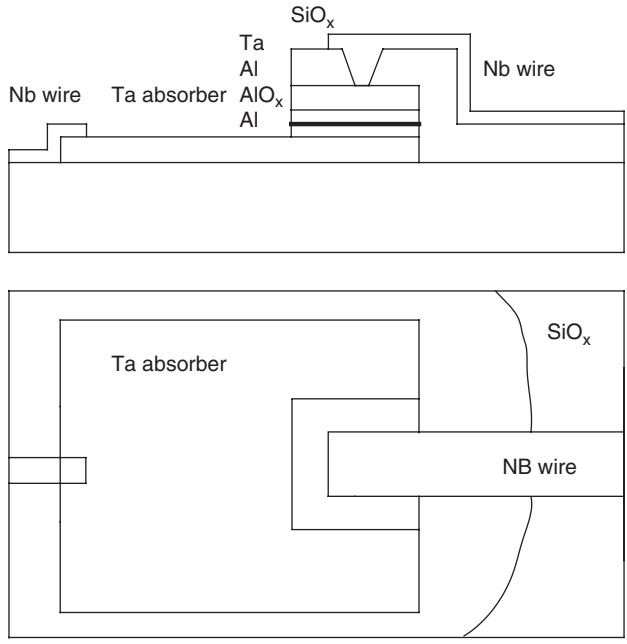


Figure 1.7 Schematic representation of a possible detector configuration based on the Superconducting Tunnel Junction (STJ). The device consists of an epitaxial Ta absorber, which together with a lower Al layer forms the base electrode of the junction, a high quality, 1–2 nm thick AlO_x isolator barrier and the Al/Ta top electrode. Two Nb leads attached to the absorber and the top electrode are necessary to apply the bias voltage and for collecting the output charge produced as a result of interactions of incoming photons/phonons with the absorber

degrade the leakage current, noise and the collection efficiency. The top counter-electrode of the STJ consists of Al and Ta layers to make the junction symmetrical. Other materials can also be utilised. The device has two Nb leads attached to the absorber and the top electrode, which are necessary to apply the bias voltage V_b as well as for collecting the output charge.

The energy diagram of the detector with Nb wires is shown schematically in Figure 1.8. Quasiparticles are generated in the Ta absorber as a result of its interaction with incoming photons/phonons. After the energy cascade has been completed, quasiparticles relax to an energy level close to Δ_{Ta}. Since the energy gap of Nb is larger than that of Ta, carriers cannot escape into the left wire. Instead, they diffuse into the Al layer with even lower energy gap Δ_{Al}, where they can further relax by the emission of phonons with $h\nu \sim (\Delta_{Ta} - \Delta_{Al})$. Similar to Ta–Nb interface,

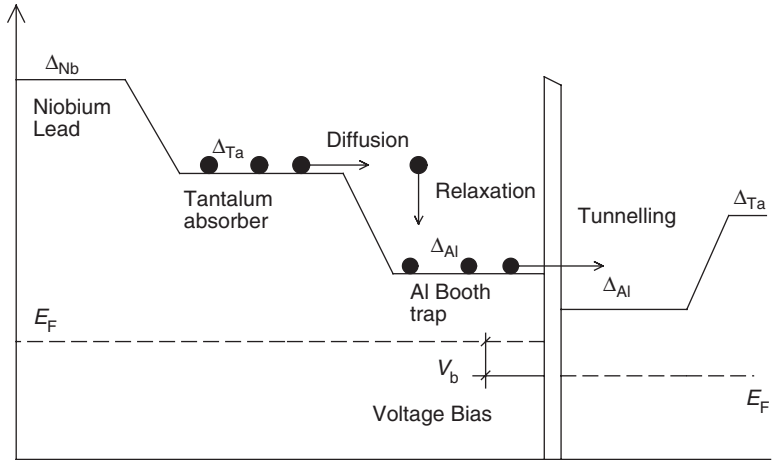


Figure 1.8 The energy diagram of the detector with Nb wires. Quasiparticles are generated in the Ta absorber as a result of its interaction with incoming photons/phonons. After the energy cascade has been completed, quasiparticles relax to an energy level close to Δ_{Ta}. Since the energy gap of Nb is larger than that of Ta, carriers cannot escape into the left wire. Instead they diffuse into the Al layer with an even lower energy gap Δ_{Al}, where they can further relax by the emission of phonons with $h\nu \sim (\Delta_{Ta} - \Delta_{Al})$. Similar to Ta–Nb interface, once relaxed carriers become trapped in Al. From this moment on they have a choice either to recombine back into pairs or tunnel through the barrier into the top electrode to be integrated by a charge transconductance amplifier

once relaxed carriers become trapped in Al. From this moment on they have a choice either to recombine back into pairs or tunnel through the barrier into the top electrode to be integrated by a charge transconductance amplifier.

The presence of Ta electrodes modifies the energy gap of the Al layer due to the proximity effect. Thus, the real energy diagram will most definitely have a more complex shape compared to one shown in Figure 1.8, which, in addition, can be affected by the quality of interfaces between layers. We will not discuss here the microscopics of the proximity effect in interfaces. Interested readers can find detailed information on variations in the energy gap as a function of the Al layer thickness, for instance, in [9].

The volume of the Booth trap v_t is normally designed to be much lower compared to the volume of the whole absorber V_{abs}. As a result, an attempt frequency for the each trapped quasiparticle to tunnel through the barrier rises proportionally to $\frac{V_{abs}}{v_t}$, improving the count rate capability of the detector. Engineering of an efficient trap is a crucial

step in the junction design. Carriers escaped back into the absorber or recombined before tunnelling contribute directly to the charge collection deficit and result in worsening the energy resolution FWHM.

There are two tunnelling mechanisms in the tunnel junction [10,11]. Figure 1.9 illustrates both of them. In the first process, sometimes called the "electron-like" tunnelling, the barrier is transparent for particles excited during the energy cascade above $E_F + \Delta_{Ta}$ (Figure 1.9b). In the second, "hole-like" one, the incident energy E creates a quasiparticle in

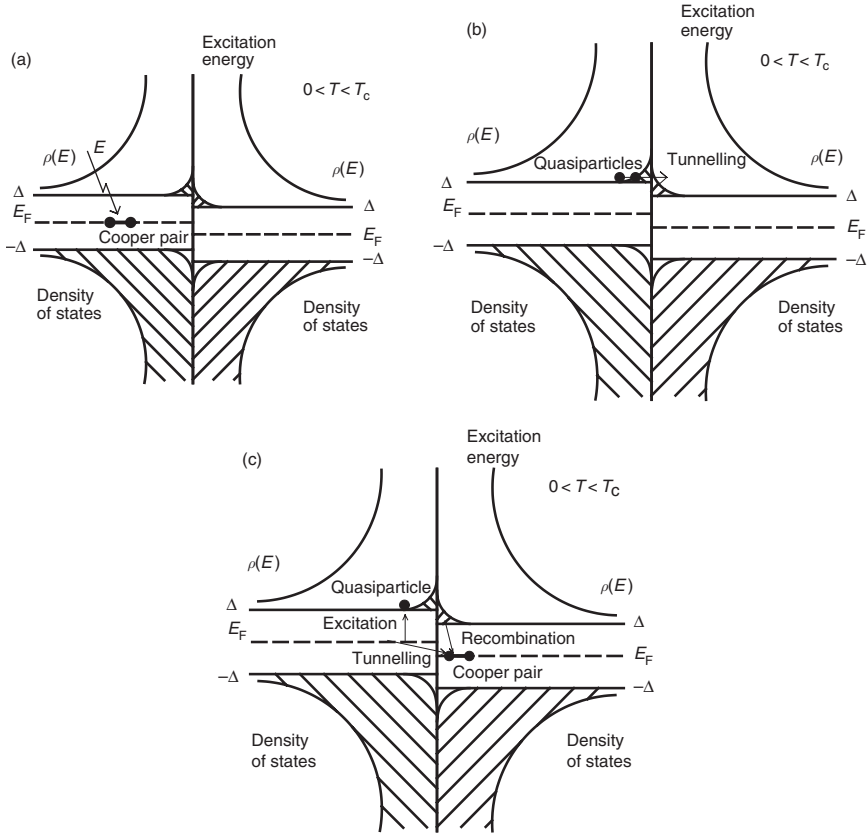


Figure 1.9 Energy diagrams explaining two tunnelling mechanisms in the superconducting tunnel junction. After an incoming photon/phonon with energy E has broken a Cooper pair, the barrier becomes transparent for created quasiparticles, as shown in the diagram (b). The process is called "electron-like" tunnelling. Alternatively, the incident energy E creates a quasiparticle in the left electrode, whereas the second electron forms a new Cooper pair with the recombined electron in the right electrode. This process is called a "hole-like" tunnelling and illustrated in diagram (c)

the left electrode, whereas the second electron forms a new Cooper pair with the recombination electron in the right electrode (Figure 1.9c). With a typical voltage bias V_b less than $\frac{2\Delta}{e}$ (e is the electron charge), it is obvious that both mechanisms can proceed in either direction. A quasiparticle, should it have a sufficient energy to do so, can tunnel forward and back several times breaking on its way more Cooper pairs in the left electrode. This makes more quasiparticles available for tunnelling and thus increases the total charge output in response to the incident energy E. This desirable phenomenon however comes at a price of the higher noise signal. The expression for the intrinsic energy resolution of the tunnel junction that takes into account the multiple tunnelling can be written as

$$\Delta E_{\mathrm{FWHM}} = 2.35[\varepsilon E(F+G)]^{1/2} \tag{1.2.1}$$

where $G = \left[\frac{P_1 - P_1^2 + 3P_1P_2 + P_1^2P_2}{(P_1 + P_1P_2)^2}\right]$ is the Goldie factor introduced in [12], $P_1 = \frac{\Gamma_{\mathrm{tun1}\to 2}}{\Gamma_{\mathrm{tun1}\to 2} + \Gamma_{D1}}$ and $P_2 = \frac{\Gamma_{\mathrm{tun2}\to 1}}{\Gamma_{\mathrm{tun2}\to 1} + \Gamma_{D2}}$ are probabilities of the forward and back tunnelling respectively, Γ_{tun} denote tunnelling rates from the electrode 1 to the electrode 2 (subscript $1 \to 2$) and back (subscript $2 \to 1$), Γ_{D1} and Γ_{D2} represent quasiparticle loss rates in the left and right electrodes respectively. According to [13], a good quality STJ should be able to demonstrate a value of G as low as 1.

The qualitative microscopic picture of the energy cascade, the quasiparticle diffusion and the tunnelling process presented so far, allows one to formulate three major design criteria for the primary sensor. These are as follows:

1. the thickness of the absorber should be sufficient to captivate incoming photons in the entire energy range of interest;
2. the maximum size of the absorber is to be selected such that a distance from a spot of the primary interaction to the Booth trap is less than the quasiparticle loss length of the absorber material; and
3. the concentration of quasiparticles in the Booth trap should not exceed a certain limit above the thermal equilibrium throughout the entire measurement procedure.

The third criterion takes care of the fact that a large density of quasiparticles per unit volume may suppress the energy gap of the trap and also lead to higher recombination rate. Both effects will result in

uncontrolled non-linear effects on the detector response or the absence of any response all together.

Criterion 1 The thickness of the absorber required is primarily linked to the quantum efficiency of a detector. The quantum efficiency as a function of the incident photon energy is defined by

$$Q_{\mathrm{eff}}(E_{\mathrm{ph}}) = \left(1 - e^{-\frac{t_{\mathrm{abs}}}{\lambda_{\mathrm{abs}}}}\right) \times 100\% \tag{1.2.2}$$

At $E_{\mathrm{ph}} = 10\,\mathrm{keV}$, the characteristic absorption length $\lambda_{\mathrm{abs}}[\mathrm{Ta}] = 5\,\mu\mathrm{m}$ (Figure 1.4). A quantum efficiency of 90% can be provided by an absorber thickness $t_{\mathrm{abs}} = 12\,\mu\mathrm{m}$.

Criterion 2 Here we would like to evaluate the maximum size of the absorber, in which quasiparticles can diffuse without a substantial loss in their number density n. Let us assume that X-ray photon interacts with the absorber at a spot with coordinates $r_0 = \{x_0, y_0, z_0\}$. The propagation of particles due to the density gradient can be described adequately by the diffusion equation given by [14]

$$\frac{\partial n(r)}{\partial t} = D\nabla^2 n(r) - \frac{n(r)}{\tau_{\mathrm{loss}}} \tag{1.2.3}$$

with the diffusion coefficient

$$D = \frac{1}{3}\sqrt{\frac{2k_{\mathrm{b}}T_{\mathrm{qp}}}{\pi\Delta}}\nu_{\mathrm{F}}\lambda \tag{1.2.4}$$

where T_{qp} denotes the temperature of quasiparticles, $\boldsymbol{\tau}_{\mathrm{loss}}$ is the quasiparticle lifetime, $\nu_{\mathrm{F}} = 0.38 \times 10^6\,\mathrm{m/s}$ is the Fermi velocity and λ represents the mean free pass. In practice, λ can be estimated from the expression $\lambda = RRR \cdot \lambda_{300}$. In this equation, the residual resistivity ratio is defined as $RRR = \frac{\rho_{\mathrm{n}}(293\,\mathrm{K})}{\rho_{\mathrm{n}}(T_{\mathrm{c}})}$, where $\rho_n(293\,\mathrm{K})$ and $\rho_n(T_{\mathrm{c}})$ denote the normal state resistivity at room temperature and just before the transition into the superconductive state respectively. $\lambda_{300} = 7\,\mathrm{nm}$ [15] is the mean free path of the Ta epitaxial film at room temperature.

Provided that the diffusion time is much longer than the duration of the energy cascade, the initial condition for (1.2.4) takes the following form:

$$n(r,0) = N \cdot \delta(r - r_0) \tag{1.2.5}$$

where $N = \frac{E}{\varepsilon} \pm \sqrt{\frac{E}{\varepsilon} F}$ and the boundary condition near the Booth trap $D\left(\frac{\partial n}{\partial x}\right)_{\text{neartrap}} = n_{\text{neartrap}} (D\tau_t)^{-\frac{1}{2}}$.

The three-dimensional (3D) diffusion picture has yet to be developed. Some results of 2D numerical computation of (1.2.4) can be found in [15,16].

According to [16], the solution of 2D case of (1.2.3) with zero boundary conditions is written as

$$n(r,t) = \frac{1}{4\pi D t} e^{\left(\frac{-r^2}{4Dt} - \frac{t}{\tau_{\text{loss}}}\right)} \tag{1.2.6}$$

provided that $t_{\text{abs}} \ll l_{\text{loss}}$. From this expression, the average distance for a quasiparticle to propagate before it disappears can be evaluated as

$$l_{\text{loss}} = \frac{\pi}{2} \sqrt{D \cdot \tau_{\text{loss}}} \tag{1.2.7}$$

This gives the upper limit for the square absorber area $A_{\text{abs}} \leq \frac{\pi^2}{4} D\tau_{\text{loss}}$.

Reliable experimental data on the quasiparticle dynamics have been obtained so far for sub-1-mm long and 0.1-mm wide thin film Ta strips. In this case, the single dimensional diffusion process becomes applicable. Fitting the experimental data with the solution of 1D (1.2.3) resulted in a diffusion factor $D = 8.2\,\text{cm}^2/\text{s}$, a lifetime $\tau_{\text{loss}} = 83\,\mu\text{s}$ and a loss length l_{loss} as low as 408 μm for polycrystalline films with $RRR = 17$ [17]. In the epitaxial films of a better quality ($RRR = 48$) [15], being slightly thinner than those described in [17], $D = 17\,\text{cm}^2/\text{s}$ and $\tau_{\text{loss}} = 50\,\mu\text{s}$ were measured. With these figures we derive the loss length $l_{\text{loss}} = 471\,\mu\text{m}$.

An upper theoretical limit for the diffusion factor of the pure monocrystalline Ta is estimated to be $63\,\text{cm}^2/\text{s}$ [15] and $\sim 40\,\text{cm}^2/\text{s}$ [18]. In the both papers, predicted figures are factor of 5 higher than the experimental data available in literature. One possible explanation for it is the fact that quasiparticles are being pushed to the surface of the superconducting absorber during the diffusion process by the Meissner effect where they scatter faster on impurities. If this is true, the quality of

the surface itself as well as the value and the uniformity of the surface energy gap define the effective diffusion factor of the absorber rather than the properties of the bulk material.

It is useful to calculate the fundamental limitations on absorber dimensions in order to draw a scope of possible applications of STJ detectors. It is derived from a measure of the longest possible quasiparticle lifetime, which is defined by the recombination rate in the bulk material [19]:

$$\tau_{\mathrm{rec}}^{-1} \approx \frac{1}{\tau_0}\pi^{1/2}\left(\frac{2\Delta}{k_{\mathrm{b}}T_{\mathrm{c}}}\right)^{5/2}\left(\frac{T}{T_{\mathrm{c}}}\right)^{1/2} e^{-\frac{\Delta}{k_{\mathrm{b}}T}} \tag{1.2.8}$$

with the characteristic time constant of the material $\tau_0[\mathrm{Ta}] = 1.8\,\mathrm{ns}$. In the equation (1.2.8), T is the effective quasiparticle temperature. The relaxation depends on T, which, in turn, changes as the population of quasiparticles evolve. At $T \ll T_{\mathrm{c}}$, the maximum value of the loss time for Tantalum is expected to be $\tau_{\mathrm{loss}}[\mathrm{Ta}] = 1\,\mathrm{ms}$. Therefore, the size of the square absorber cannot exceed in principle a value of $l_{\mathrm{loss,\,max}} = 4\,\mathrm{mm}$ with a single tunnel junction attached to its side.

For convenience, characteristic time constants of some other materials are summarised in Table 1.3.

Criterion 3 The linearity of the detector response to the photon energy E depends strongly on the dynamics of the behaviour of quasiparticles in the Booth trap and tunnelling itself. In any case, one needs to design this part of the detector in such a way that the tunnelling rate is faster than or equal to the trapping rate. Assuming the whole volume of the trap ν_{t} being in a state close to the equilibrium and neglecting the proximity effect, the trapping time can be found from [7] as

$$\tau_{\mathrm{t}} = \tau_{\mathrm{s}}\frac{V_{\mathrm{abs}}}{\nu_{\mathrm{t}}} \tag{1.2.9}$$

Table 1.3 Characteristic time constants, V_{F}, density of states in selected superconducting materials

Material	τ_0 (ns)	τ_{ph} (ns)	V_{F} ($10^6\,\mathrm{ms}^{-1}$)	$N(0)$ (10^{31}states/eV/cm^3)
Nb	0.149	0.004	1.37	31.7
Ta	1.78	0.023	1.5	40.8
Al	438	0.242	2.03	12.2

Here V_{abs} represents the volume of the absorber and $\tau_s = \tau_0 \frac{\Delta_{Al}}{\Delta_{Ta}} \approx 1.56\,\text{ns}$ denotes the scattering time of quasiparticles in the trap.

Expression (1.2.9) is very useful as a qualitative illustration of the trapping phenomenon. In practice, however, the physical picture turns out to be much more complex. The proximity effect is obviously pronounced in the whole v_t as the high-quality epitaxial films with long coherence lengths are utilised. In this case, the energy gap exists in a gradient form and an average trapping time can be derived only from the volume integration of local distribution of $\frac{d\tau_t}{dv_t}$. Non-equilibrium effects become stronger as the thickness of the absorber and the volume of the trap increase. Partially this can be explained by the fact that in thinner devices a large portion of phonons emitted by trapped quasiparticles escape from the trap without transferring their energy to other quasiparticles. As the volume v_t becomes larger the probability of phonon–particle interaction rises, so does the average temperature of the trapped quasiparticles T_{qp}. This may lead to a larger escaping rate of quasiparticles from the trap and eventually a reduced charge collection. An immediate idea to operate large-area STJ detectors with thick absorbers at a lower base temperature may not resolve the problem as it further slows down the thermalisation process, contributing more to non-stationary effects. Clearly, a compromise needs to be found.

An average tunnelling time for trapped quasiparticles to pass through the barrier is given by

$$\tau_{tun} = k_1 e^2 N(0) \nu_t R_N \tag{1.2.10}$$

Here k_1 is the proportionality factor introduced in [20]. In general, k_1 is a function of the energy gap, temperature and the voltage bias. Its value is close to 1 at $T \ll T_c$ and V_b is just below Δ/e. $N(0)$ is the normal density of states of the trap material at the Fermi level and R_N represents the normal resistance of the insulating barrier of the tunnel junction at $V_b > \Delta/e$. The normal resistance can be found from the junction I–V curve. Figure 1.10 shows a schematic example of such an I–V curve. R_N provides a measure of the tunnelling probability and is proportional to $\exp\left(\frac{2b}{A}\right)$, where b is the thickness of the insulating barrier and A is the area of the superconducting tunnel junction. The equation (1.2.10) suggests that the tunnelling rate is independent of temperature when $T \ll T_c$.

Due to the DC Josephson effect the Cooper pairs can also propagate without a resistance through the barrier. This results in the existence of

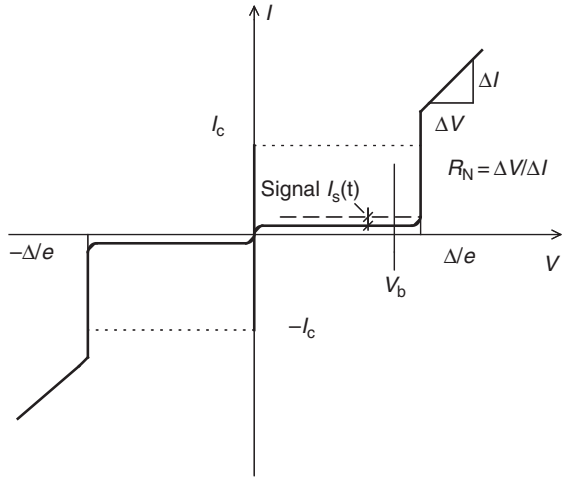


Figure 1.10 Schematic representation of the *I–V* curve of the tunnel junction. V_b is the voltage bias applied to the junction electrodes. R_N is called the normal resistance of the insulating barrier. It provides a measure of the tunnelling probability and is proportional to $\exp\left(\frac{2b}{A}\right)$, where b is the thickness of the insulating barrier and A is the area of the tunnel junction

the supercurrent I_c, which can flow in both directions. In order to suppress it, a small magnetic field is normally applied parallel to the barrier as shown in Figure 1.11.

The magnetic field dependence of the critical current has a Fraunhofer-like periodicity. For the square shape STJ, it can be written as

$$I_c(B) = I_c(0)\left|\frac{\sin(\pi\Phi/\Phi_0)}{\pi\Phi/\Phi_0}\right| \tag{1.2.11}$$

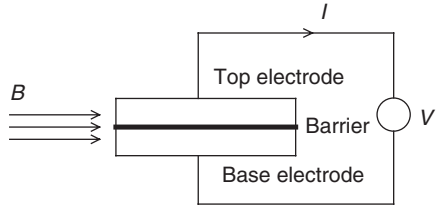


Figure 1.11 Voltage-biased superconducting tunnel junction. A small magnetic field B must be applied parallel to the insulation barrier in order to suppress a supercurrent caused by the Josephson effect

where $\Phi = Ba_J(2\lambda_L + d)$ is the magnetic flux in the barrier, a_J denotes the size of the junction, λ_L is the London penetration depth, d is the thickness of the barrier, $\Phi_0 = 2.07 \times 10^{-15}$ Wb is the magnetic flux quantum and $I_c(0) = \frac{\pi}{2}\frac{\Delta}{R_N e}$ is the zero-field critical current for the symmetric junction (both electrodes are made of the same material). An example of an ideal Fraunhofer pattern is shown in Figure 1.12. The presence of trapped Abrikosov vortices, pinholes in the barrier, nonuniformity of magnetic field and so on cause variations in the shape of the pattern. In any event, magnetic field must maintain the operating point at one of minimums of I_c(B). The problem becomes more complex when biasing an array of junction with one solenoid. We will discuss it in more details in Chapter 5.

The accuracy in the suppression of the supercurrent is also very important in damping the so-called Fiske current steps on the *I–V* curve. These current steps originate from the AC Josephson effect, which prescribes the generation of the AC-current (electromagnetic field) by the voltage-biased tunnel junction. The voltage–frequency transfer factor is expressed through the fundamental atomic constants:

$$f_J = \frac{2e}{\pi\hbar} V_b \tag{1.2.12}$$

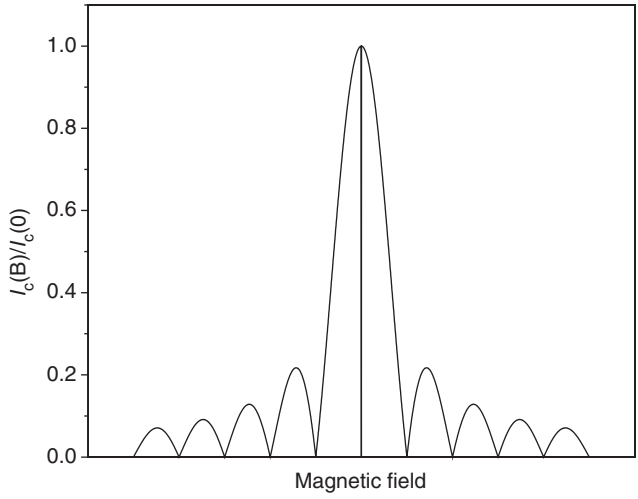


Figure 1.12 The magnetic field dependence of the Josephson supercurrent

with $\hbar$ being the Planck constant. The Fiske current steps appear at bias voltages, where f_J coincides with the major cavity resonance of the junction as well as its harmonics. They are merely the result of the detection of Josephson oscillations by the junction non-linearity. In the square junction, Fiske steps are expected to appear at the voltage bias values given by [21]:

$$V_b^n = \frac{\hbar \pi c n}{2 e a_J} \sqrt{\frac{d}{\varepsilon_r (2\lambda_L + d)}} \tag{1.2.13}$$

where n is the harmonic number, c is the velocity of light and ε_r is the relative permittivity of the barrier.

A theoretical estimate of the magnetic field at which DC and AC Josephson currents become effectively suppressed has a form [22]:

$$B \geq \sqrt{\frac{I_c Q_f \Phi_0^2 e^2 \Delta}{2A(2\lambda_L + d)\pi^3 n^2 \varepsilon_0 \hbar c^2 (2\pi \Delta k_b T)^{1/2} e^{\frac{-\Delta}{k_b T}}}} \tag{1.2.14}$$

For square-shaped junctions with a quality factor of the cavity $Q_f = 11, B \sim 200$ Oe. For some shapes, other than the square-like one, the same degree of suppression can be achieved at half this value [22].

One of the most important indicators of the junction quality, similar to semiconductor ionisation detectors, is the leakage current ΔI_l (Figure 1.10). ΔI_l includes primarily the current of thermally excited quasiparticles (Figure 1.1b):

$$I_{th} = 2e \frac{N_{th}}{\tau_{tun}} \tag{1.2.15}$$

where the number of thermally excited quasiparticles in the absorber and the Booth trap is given by

$$N_{th} = (V_a + \nu_t) \sum_{i=1}^{2} 4 N_i(0) \Delta_i \sqrt{\frac{\pi k_b T}{2\Delta_i}} e^{-\frac{\Delta_i}{k_b T}}$$

where $i = 1, 2$ represent the materials of the absorber and the trap respectively. The expression (1.2.15) does not take into account carriers that tunnel from the counter-electrode back into the base electrode, as well as the recombination and escape of particles.

In real devices, the total leakage current may not be entirely defined by the sub-gap thermal current I_{th}. Pinholes or/and the non-uniformity of the barrier as well as the edges normally contribute a substantial part of it.

An amount of the charge that can be collected from an STJ detector is given by

$$Q_{col} = Q_0 \frac{1}{1 + \frac{\tau_\Sigma}{\tau_{rec}} + \frac{\tau_\Sigma}{\tau_l}} \tag{1.2.16}$$

where $\tau_\Sigma = \tau_{tun} + \tau_d + \tau_t$. In this expression, τ_l^{-1} represents the loss rate caused by all temperature-independent mechanisms. Examples of such mechanisms involve the diffusion of quasiparticles from the junction absorber area into the current leads, phonon escape into the substrate and partially into the top counter-electrode, quasiparticle trapping in the volumes with the suppressed energy gap apart from the intentionally designed Booth traps. Three types of parasitic traps may be present in the STJ sensor:

1. traps caused by imperfection of material;
2. traps at edges and surfaces of the absorber and electrodes; and
3. traps created around the Abrikosov vortices [23].

The first two types of traps can be addressed by tuning the design, the technological process or by the collimation of the X-ray beam. The last one places separate requirements on the design stage of the STJ detector packaging and the operation procedure. Energetically, Abrikosov vortices enter a high-quality superconductor II mostly during the cool down near its T_c. Therefore, the sensor should be surrounded by a compact, permanently installed, combined cryoperm/superconductor shield. Alternatively, the whole detector package can be placed in a screened enclosure with a small residual magnetic field less than 1 nT during the cooling down. The shielding and the magnetic field bias will be discussed again in more detail in Chapter 5.

Following the logistics of [13], the time dependence of the charge output can be derived from the following equation

$$Q(t) = Q_0 \frac{\tau_r}{\tau_{tun}} \left[1 - e^{\frac{t}{\tau_r}}\right] \tag{1.2.17}$$

Here τ_r is the rise time of the charge pulse. According to equation (1.2.17), it can be defined as a time during which a 1/exp fraction of the total

charge is collected. The charge rise time can be found experimentally using a fast current read-out of the detector, based normally on high-speed super conducting quantum interference device (SQUID) amplifiers. The output current can be approximated by

$$i_s(t) = i_s(0)e^{-\frac{t}{\tau_r}} \tag{1.2.18}$$

where $\tau_r \approx \tau_l$ is transformed into the decay time of the current pulse and the initial current $i_s(0) \approx Q_0 e \tau_{tun}$, Q_0 is the initial number of created quasiparticles.

The state of art in the STJ detector technology is such that expressions given above work reasonably well up until the energy of the incident photons $E < 1\,\text{keV}$. In the energy range from 1 to 5 keV, ΔE is found somewhat larger compared to simulations based on the extended Rothwarf–Taylor equations and the classical diffusion process. Predicted and measured values of ΔE are given in Table 1.4. One of the most interesting commercial applications of the STJ detectors is the electron-probe nanoanalysis in combination with high-resolution field emission gun scanning electron microscope (FEG-SEM). The FEG-SEM operates with low-beam energy $E_0 < 5\,\text{keV}$. Performance of the current state STJ detectors is already adequate for this type of application.

As we mentioned before, the incomplete thermalisation of quasiparticles prior to tunnelling is responsible for the broadened peaks of energy spectrum. Models adequately describing the quasiparticle dynamics and the energy redistribution in the non-equilibrium superconductors with a complex energy gap profile are still at the development stage. The latest results on this topic with respect to the date of writing this book can be found in [27–29]. One needs to remark that linearity of detectors can be significantly improved by attaching several STJs to an absorber connected in parallel or in series. This is the right way to build devices with large-volume absorbers capable of operating in the whole energy range of "soft" X-rays.

Table 1.4 The energy resolution as a function of the incident photon energy

	1 keV	**3.8 keV**	**5.9 keV**
Intrinsic ΔE (eV)	6.3 [24]	12 [24]	15.3 [24] 13 [26]
Predicted ΔE (eV)	3.0 [24], 4.6 [25]	7.3 [24]	11.6 [24] 8.9 [17]

1.3 MICROCALORIMETERS BASED ON TRANSITION EDGE SENSOR

An X-ray calorimetric detector in its classical form is shown schematically in Figure 1.13. It consists of an absorber with which incoming photons interact, a thermometer to measure the temperature dynamics of the absorber and finally a controlled thermolink providing the power flow deposited by the photons from the absorber to the heat sink. The principle of operation of the calorimeter is very simple. After a photon has transferred its energy E into the absorber, the temperature of the latter increases by $\Delta T = \frac{E}{C}$. Here $C = cV$ with c and V representing the specific heat capacity and the volume of the absorber respectively. The rise time of the output of the thermometer τ_r is defined by the duration of the energy redistribution in the whole absorber plus the interaction time between the absorber and the thermometer.

The decay time of the calorimeter is given by

$$T(t) = \Delta T e^{-\frac{t}{\tau_d}} \tag{1.3.1}$$

where the decay time constant $\tau_d = \frac{cV}{G}$.

The phonon noise in the absorber determines the thermodynamic limit of the energy resolution that can be obtained in principle with any type of calorimeters. The average number of thermally activated phonons in the incoming energy E (the signal-to-noise ratio) is

$$\overline{N} = \frac{E}{\overline{E}_{ph}} = \frac{CT}{k_b T} \tag{1.3.2}$$

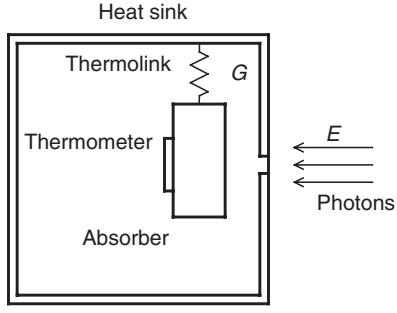


Figure 1.13 General representation of a calorimetric detector intended for X-ray measurements. Equally, it is capable of measuring the energy of incoming phonons. The calorimeter consists of an absorber, a thermometer and a controlled thermolink providing the power flow from the absorber to the heat sink

Assuming again the Poisson statistics, the expression for the energy resolution takes the form

$$\Delta E_{\mathrm{FWHM}} = 2.35\sqrt{\overline{N}}\,\overline{E}_{\mathrm{ph}} = 2.35\zeta\sqrt{k_{\mathrm{b}}T^2C} \qquad (1.3.3)$$

where the factor ζ takes into account an outcome of any mechanism which influences the ultimate signal-to-noise ratio in the calorimeter.

Similar to particle detectors, the formation of thermally activated phonons is not an entirely random process. However, the correlation between events is caused by entirely different reasons [30]. As a result, the equation (1.3.3) must include a correction factor analogous to the Fano factor in the STJ detectors.

A selection of a temperature-measurement method plays a crucial part in the optimisation of the desirable energy resolution versus the count rate capability. Originally, X-ray microcalorimeters intended for operation at sub-1 K temperature range were assembled with doped semiconductor thermistors. The most popular choice was either neutron transmission-doped germanium or ion-implanted silicon. The detectors achieved a very impressive energy resolution of 8 eV for 5.89 keV MnK_α at a temperature of 70 mK [31]. Yet at this temperature, semiconductor thermometers suffer from two fundamental limitations caused by the hot electron effect. As the temperature of the detector is lowered, electrons become increasingly isolated from the crystal lattice. This slows down the detector response due to a longer thermalisation time as well as worsens the energy resolution by the reduced temperature sensitivity of the thermistor [32].

An important step forward in designing the ultra low temperature (ULT) calorimeters was the combination of the TES and a normal metal absorber [33]. The energy cascade in the normal metal starts with the ejecting of a high-energy inner core electron by the incoming X-ray photon. The photoelectron loses its energy through the excitation of the shell electrons in other atoms as well as via the scattering of other free electrons. The early stage of the cascade ends with the formation of a hot spot containing high-energy electrons and de-excitation phonons. The heat of the spot further diffuses through three scattering mechanisms:

1. electron–electron interaction;
2. electron–phonon interaction; and
3. phonon–phonon interaction.

As the temperature decreases, rates of all three power-distribution mechanisms have downward trends, but with different slopes. Below a certain threshold temperature (~0.5 K), the electron–electron scattering rate becomes the fastest. At a temperature of 0.1 K, which is a characteristic operation point for typical TES microcalorimeters, the electron–electron interaction is already so much faster compared to the electron–phonon interaction that the electron system can be treated as independent from phonons. In other words, the energy, E, deposited by an incoming event, is first absorbed by the electron system and the temperature of "hot" electrons carries precise information about E. To take advantage of the fast electron–electron interaction rate, we need a thermometer which is capable of directly measuring the temperature of the "hot" electron system. It has to perform the measurement within a time constraint, before the energy is absorbed by the lattice. With a non-inertial thermometer, the response time of a "hot" electron microcalorimeter compared to its electron–phonon counterpart scales as the ratio of the electron–electron interaction rate to the electron–phonon interaction rate, $\Gamma_{e-e}/\Gamma_{e-ph}$. That can be several orders of magnitude depending on the operating temperature and absorber/thermometer materials utilised.

The TES is a thermometer that changes its resistance in response to the temperature of the hot electrons. The sensor is made of a superconducting strip operating near the middle of its temperature transition curve, between the normal and superconducting states. An example of such a curve is shown in Figure 1.14. The quality of the TES is

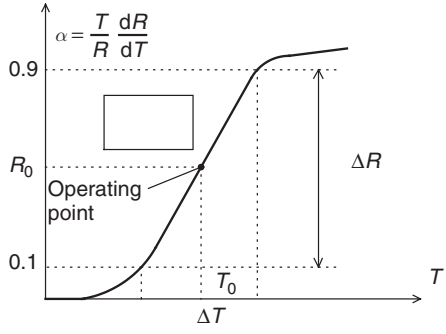


Figure 1.14 An example of the temperature transition curve of the TES between normal and superconducting states. The dynamic range is shown as ΔR and ΔT. The quality of a TES at its operating point is characterised by the sensitivity $\alpha = \frac{T_0}{R_0}\frac{dR}{dT}$

characterised by the linearity of the dynamic range $\Delta R(\Delta T)$ and the sensitivity is defined as

$$\alpha = \frac{T_0}{R_0}\frac{\mathrm{d}R}{\mathrm{d}T} \tag{1.3.4}$$

The manufacture of a TES with a predictable transition temperature T_{ct}, dynamic range ΔT and sensitivity is a quite formidable task in itself. The control of parameters becomes even more complex if the following deposition steps require heating of the substrate or the exposure to chemicals. There are three major techniques allowing tuning T_{ct}. The first approach is to use superconductor/normal metal bilayers, in which the critical temperature of the superconductor is lowered due to the proximity effect [34]. In the second one, T_{c} decreases by implanting magnetic atoms that destroy the local superconductivity [35]. The mechanism can be treated as a sort of the lateral proximity effect. Therefore, care must be taken to arrange magnetic centres uniformly across the film. In the third approach, the most robust from the manufacturing point of view, the transition temperature is adjusted by proper selection of the substrate material and the film growth conditions. T_{c} changes apparently due to the stress in the film developed by the differential contraction between the film and the substrate. In fact, it can be enhanced as well as decreased compared to the critical temperature of the bulk material.

In the bilayer technology, the transition temperature and the temperature dynamic range, ΔT, depend strongly on the thickness of each layer and the interface between the layers. T_{c} can be found using the Usadel theory [36] as

$$T_{\mathrm{ct}} = T_{\mathrm{c}}\left[\frac{\mathrm{d_v}}{\mathrm{d_0}}\frac{1}{1.13\left(1+\frac{1}{\alpha}\right)}\frac{1}{t}\right]^{\alpha} \tag{1.3.5}$$

with $\frac{1}{d_0} = \left(\frac{\pi}{2}\right)k_{\mathrm{b}}T_{\mathrm{c}}\lambda_{\mathrm{f}}^2 n_{\mathrm{s}}$, $\alpha = \frac{d_{\mathrm{n}}n_{\mathrm{n}}}{d_{\mathrm{s}}n_{\mathrm{s}}}$, d_{n} and d_{s} denote the thickness of the normal and superconducting film respectively, n_{n}, n_{s} represent the density of electronic states in the normal and superconducting films respectively, λ_{f} is the Fermi wavelength and t is the transmission factor. The transmission factor was found in experiments with Mo/Cu bilayers. The best fit between the theory and the experiment occurs at $t = 0.21$ [36].

Another popular combination of materials is Au and Ir. A detailed discussion of properties of this bilayer can be found in [34]. In the same paper, authors present their own microscopic model based on a modified de Genes–Werthammer theory.

The manufacture steps of TES must ensure a fine control over the quality of materials, interfaces, the sharpness of edges and the precision of the layer thickness. The normal metal layer should well overlap the superconductor to provide the uniformity of T_{cs}. At room temperature, some interdiffusion process between layers takes place, modifying T_{cs} itself as well as ΔT with time. In order to avoid these problems, microcalorimeters of the latest design tend to incorporate monolayer transition-edge thermometers, for instance, Mo with implanted ferromagnetic centres or, alternatively, Ir with W, whose T_c is tuned by applying a stress.

One possible configuration of TES primary detector is shown schematically in Figure 1.15. It includes an Au absorber deposited on top of Au/Ir bilayer thermometer. The whole detector is fabricated on a freely suspended silicon-nitride membrane. The thickness and the design of the membrane define the thermal link G between the TES and the silicon substrate attached to the heat bath with a minimum Kapitza resistance. The membrane also helps to prevent escaping of the incoming phonons from the absorber before they fully thermalised on hot electrons. Al leads connect the sensor to the biasing and read-out electronic circuits. In the superconducting state, leads virtually do not conduct heat out of the sensor due to a small number of thermally activated, conducting, normal electrons at $T \ll T_{cAl}$.

Conventional electron–phonon calorimeters exhibit a rather poor count rate capability, insufficient for the majority of applications in the analytical instrumentation. It is caused by a relatively long decay time, τ_d. The decay time, however, can be reduced by operating the calorimeter

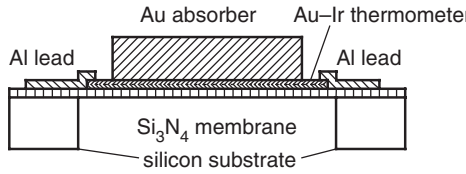


Figure 1.15 An example of one possible configuration of TES primary detector. It includes an Au absorber deposited on top of the Au/Ir bilayer thermometer. The whole detector is fabricated on a freely suspended silicon-nitride membrane. The thickness and the design of the membrane define the thermal link G between the TES and the silicon substrate attached to the heat bath with a minimum Kapitza resistance. The membrane also helps to prevent escaping of the incoming phonons from the absorber before they fully thermalised on hot electrons. Al leads connect the sensor to the biasing and read-out electronic circuits. In the superconducting state, leads virtually do not conduct heat out of the sensor due to a small number of thermally activated, conducting, normal electrons at $T \ll T_{cAl}$

in a faster bolometric mode with a negative electro-thermal feedback. Bolometers measure an incoming energy of a single photon E as a power $P(t) = \frac{dE(t)}{dt}$. The temperature of the absorber and TES is maintained elevated well above the temperature of the cold finger by applying a power bias $P_b(t) = P_0$. The power flows continuously from the absorber through the TES, through the SNi membrane into the Si substrate. If the lattices of the TES and Si_3Ni_4 are well matched, the heat conduction is fulfilled by the electron–phonon mechanism, which can be described as

$$P = WV_{\Sigma}(T^5 - T_s^5) \tag{1.3.6}$$

where W is the Wellstood factor [37], V_{Σ} is the volume of the sensor and T_s denotes the temperature of the substrate. The presence of a mismatch between lattices of the TES and SNi membrane creates a Kapitza resistance acting in series with the electron–phonon coupling. The power flow through the Kapitza resistance is given by [38]

$$P_K = \frac{(T^4 - T^4)A_f}{4R_K T_s^3} \tag{1.3.7}$$

Here A_f represents the area of the TES and $R_K = 2 \times 10^{-3}\,\mathrm{Km^2/W}$ is the Kapitza resistance per unit area for Cu–Si interface [39].

In order to keep the power flow constant, when incident events release their energy into the absorber, the bolometer includes an electro-thermal feedback which reduces the bias P_b by an amount of the incoming event power $P_i(t)$ so that the following condition is met

$$P_b(t) + P_i(t) = P_0 \tag{1.3.8}$$

In this case, the feedback electric signal becomes essentially the output of the detector. The original energy of the incoming photon can be recovered from linear proportionality between the feedback system and the incoming power by the integration of the latter one as follows

$$E = \int_0^t [P_b(t) - P_0]dt \tag{1.3.9}$$

Further, we will also see that the deposited event energy can be removed from the sensor much faster in the bolometric mode with electro-thermal feedback compared to the decay time defined by the equation (1.3.1).

There are two ways to introduce the electro-thermal feedback:

1. a current-biased TES with an external electro-thermal feedback incorporating a separate, specially designed heating element into the absorber; and
2. a voltage-biased TES with an internal electro-thermal feedback.

The first approach is shown schematically in Figure 1.16. The sensor is biased initially by a constant current I_0 and a power $P_b(0)$ from an external heating element. Once the energy of an incident photon is absorbed, the hot electrons start diffusing into the TES, thereby increasing its resistance and the voltage across it. A SQUID voltage amplifier senses this voltage increment. Its output is used as a feedback signal to partially reduce the current flowing through the heating element until the TES returns to the initial unperturbed operating point. The energy E can be found as

$$E = -I_0 \int_0^t \Delta V(t) \mathrm{d}t \tag{1.3.10}$$

This method has several unique merits, which make it attractive for single-pixel applications. These are as follows:

– the key component of the detector, the non-linear TES, operates as a zero indicator, suggesting excellent prospects in terms of the output linearity;

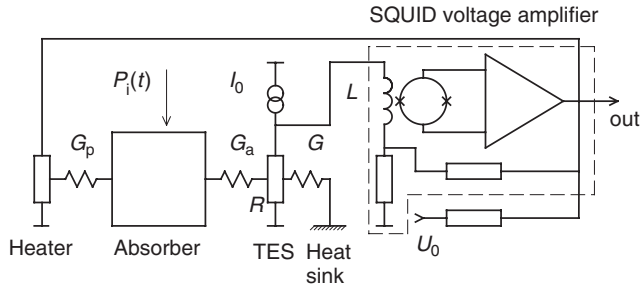


Figure 1.16 Schematic diagram of the current-biased TES microcalorimeter. The sensor is biased by a constant current I_0 and a power $P_b(0)$ from an external heating element. Once the energy of an incident photon is absorbed, the hot electrons start diffusing into the TES, thereby increasing its resistance and the voltage across it. A SQUID voltage amplifier senses this voltage increment. Its output is used as a feedback signal to partially reduce the current flowing through the heating element until the TES returns to the initial unperturbed operating point

– virtually any value of the electro-thermal feedback strength can be set up electronically just with a single limiting condition to make the whole system remain stable;
– the heat flows in the same desirable direction from absorber through the TES towards the heat sink over the whole measurement time of the incoming photon/phonon energy;
– the bias current remains unchanged, eliminating its influence on the TES operating point and the output of the detector as a whole.

The detector with the current-biased TES and the external electro-thermal feedback also has a greater potential when operating at a higher temperature range. Its figures of merit such as the energy resolution and the count rate capability depend to a lesser degree on the sensitivity of the TES itself [40].

The other operation technique of the bolometric calorimeter prescribes using the voltage-biased thermometer with a passive electro-thermal feedback [41] presented in Figure 1.17. In this mode, the TES performs simultaneously the functions of a heater, providing the power self-bias, and a thermometer. A Joule heating coming into the unperturbed system amounts to $P_{\mathrm{b}}(0) = \frac{V_{\mathrm{b}}^2}{R_0}$, where R_0 is the resistance of the TES at the nominal operating point in the middle of the transition curve, as shown in Figure 1.14. After a photon has interacted with the absorber, hot electrons diffuse into the TES, thereby increasing its electron temperature and the resistivity by ΔR and, at the same time, reducing the

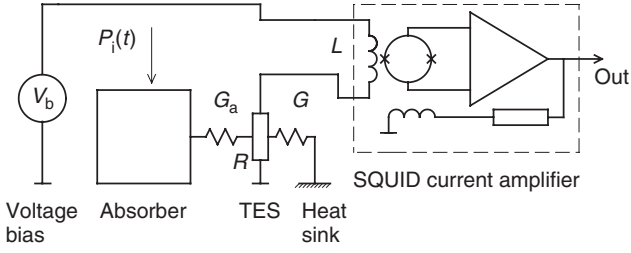


Figure 1.17 The voltage-biased TES microcalorimeter with an internal electro-thermal feedback. In this mode, the TES performs simultaneously the functions of a heater, providing the power self-bias, and a thermometer. A Joule heating coming into the unperturbed system amounts to $P_{\mathrm{b}}(0) = \frac{V_{\mathrm{b}}^2}{R_0}$. After a photon has interacted with the absorber, hot electrons diffuse into the TES, thereby increasing its electron temperature and the resistivity by ΔR and, at the same time, reducing the bias power to a level of $P_{\mathrm{b}}(t) = \frac{V_{\mathrm{b}}^2}{R_0 + \Delta R(t)}$. The absorbed energy from an incoming event is removed from the absorber over a decay time, τ_{etf}, that allows the system to return to its initial (R_0, I_0) state

bias power to a level of $P_{\mathrm{b}}(t) = \frac{V_{\mathrm{b}}^2}{R_0+\Delta R(t)}$. The absorbed energy from an incoming event is removed from the absorber over a decay time, τ_{etf}, that allows the system to return to its initial (R_0, I_0) state. The current variation $\Delta I(t) = \frac{V_b}{R_0+\Delta R(t)}$ is amplified by a SQUID current amplifier. The deposited energy E is recovered by the integration of the current over the pulse duration as follows

$$E = -V_{\mathrm{b}} \int_0^t \Delta I(t)\mathrm{d}t \tag{1.3.11}$$

In the voltage-biased TES, temperature of the electron system in the absorber takes a role of the null indicator. Therefore, the output does depend on the linearity of the thermometer, transfer function of which can also be affected by variations of the current flowing through the device. Despite these drawbacks, self-stabilising calorimeters have gained a wider acceptance as primary detectors for the X-ray analytical instrumentation. Without an external electro-thermal feedback, the amount of wires connecting the ULT stage with the room temperature biasing, read-out and processing electronics can be significantly reduced. This simplifies the detector assembly and cuts down the thermal load on the cold finger. Obviously, the mode of operation becomes particularly attractive when devices are combined into multipixel imaging arrays. An array of the voltage-biased TES microcalorimeter, read out with the analogue SQUID, requires half the amount of room temperature leads compared to the number of wires that would be needed for the current-biased devices. An implementation of digital SQUIDs performing the function of ADC in combination with a fast superconductor signal processor eliminates the need to control each individual pixel from outside the cryogenic environment all together. Such a system will be able to deliver a fully processed spectrum through a single cable in the digital form to a host computer for storage and visualisation.

The dynamics and noise in a TES microcalorimeter can be described by a system of two linearised equations written in the frequency domain with the help of the Fourier transform [42]:

$$\begin{pmatrix} |\delta I_\omega|^2 \\ |\delta T_\omega|^2 \end{pmatrix} = |D(\omega)^2| \begin{pmatrix} \omega^2 + \tau_{\mathrm{th}}^{-2} & A^2 \\ B^2 & \omega^2 + \tau_{\mathrm{el}}^{-2} \end{pmatrix} \begin{pmatrix} \left|I_{JN}(\omega)\right|^2 \\ |T_{PN}(\omega)|^2 \end{pmatrix} \tag{1.3.12}$$

where $D(\omega)^{-1} = \det\begin{pmatrix} j\omega + \tau_{\text{th}}^{-1} & A \\ B & j\omega + \tau_{\text{el}}^{-1} \end{pmatrix}$, $A = \frac{I_0}{L_0}\left[\frac{\partial R(I,T)}{\partial T}\right]_0$, $B = \frac{1}{C(T_0)}\left[\frac{\partial P_J(I,T)}{\partial I}\right]$, I_0 and L_0 denote the current flowing through the TES and the input inductance of the SQUID current amplifier at an initial operating point. The input inductance can change substantially due to the SQUID back action when the non-compensation error in the SQUID amplifier feedback loop is present. We will discuss the back action in Chapter 4 in more detail. $P_J(I,T) = I^2R(I,T)$ is the Joule heating due to the voltage bias, $|I_{JN}(\omega)|^2 = \frac{4k_bT_0R_0}{L_0}$ and $|T_{PN}(\omega)|^2 = \frac{2k_b(T_0^2+T_{\text{heat sink}}^2)G(T)}{C(T)}$ represent the power spectra of the Johnson noise and thermal noise of the TES respectively.

$$\tau_{\text{th}} \approx \left[\frac{\partial}{\partial T}\left(\frac{P_{(K)}(T)}{C(T)}\right)\right]_0 \left\{\frac{1}{n}\frac{T_0}{R_0}\left[\frac{\partial R(I,T)}{\partial T}\right]_0\left(1 - \frac{T_{\text{heat sink}}^n}{T_0^n}\right) - 1\right\} \text{ and}$$

$$\tau_{\text{el}} \approx \left\{\frac{1}{L_0}\left(R_0 + I_0\left[\frac{\partial(I,T)}{\partial I}\right]_0\right)\right\}^{-1}$$

are thermal and electrical time constants respectively.

The Bode criterion of the feedback stability with a $\frac{\pi}{6}$ margin in the phase shift requires the following ratio between the major time constants of the microcalorimeter

$$(\tau_{\text{th}}^{-1} + \tau_{\text{el}}^{-1}) \geq 4\tau_{\text{etf}}^{-1} \tag{1.3.13}$$

Here $\tau_{\text{etf}}^{-1} = \sqrt{AB}$.

The energy resolution of the TES microcalorimeter can be estimated as [42,43]

$$\Delta E_{\text{FWHM}} \approx 2.35\sqrt{\frac{4k_bT_0^2C(T)}{\alpha}\frac{\sqrt{n}}{2}\sqrt{1 + \frac{T_{\text{heat sink}}^2}{T_0^2}}} \tag{1.3.14}$$

The type of feedback does not influence the energy resolution as long as the SQUID read-out does not introduce extra noise and additional non-linearity. Therefore, the equation (1.3.14) is valid for the TES micro-calorimeter with the external electro-thermal feedback as well. Only in that case α is no longer the sensitivity of the TES. It has to be derived from the loop amplification comprising the TES, the SQUID voltage amplifier, heater and absorber. Details of these calculations can be found in [44].

In both modes of operation, the electro-thermal feedback helps to speed up the detector response and also reduces its sensitivity to the drift and fluctuations of the base temperature. Both the factors are improved by a factor of

$$1 + \Im(j\omega) = 1 + \frac{P_b \alpha}{GT(1 + j\omega\tau_{etf})} \tag{1.3.15}$$

where $\Im(\omega)$ is the loop amplification of the electro-thermal feedback.

To summarise the section, Table 1.5 compares theoretical predictions about the energy resolution of different types of TES microcalorimeters with experimental results obtained by some research groups not long before the date of this publication.

1.4 OTHER CRYOGENIC DETECTORS

1.4.1 NIS Hot Electron Microcalorimeters

As we mentioned in the previous section, hot electron microcalorimeters can employ some other types of thermometers, apart from TES, which can equally measure temperature of the electron system before the electron–phonon thermalisation takes place. The normal metal–insulator–superconductor (NIS) tunnel junction represents a practical example of such a thermometer. A schematic representation of the NIS-based detector is shown in Figure 1.18. It includes an Ag–Bi sandwich absorber deposited on top of an Si_3N_4 membrane. The right side of the absorber is attached to a superconductor counter-electrode through a 1–2-nm thick Al_2O_3 insulator barrier. Contacts to the detector are provided by superconducting Al leads, which carry electrical current but do not conduct heat.

If the NIS junction is voltage biased below the energy gap of the superconductor counter-electrode, the current flowing through the junction is given by [48]

$$I = \frac{1}{e^2 R_N} \int_{\Delta}^{\infty} \frac{1}{\left(e^{\frac{E - eV_b}{k_b T_{el}}} + 1\right)} \frac{E}{\sqrt{E^2 - \Delta^2}} \, dE \tag{1.4.1}$$

where E_{el} and T_{el} represent the excitation energy and the temperature of electrons respectively and R_N is the normal state resistance that is derived from the current to voltage characteristics.

Table 1.5 Comparison of theoretical and measured energy resolution in TES microcalorimeters

Energy resolution (eV)	2 eV at 1.5 keV	4.5 eV at 5.9 keV	4.5 eV at 5.9 keV	3.9 eV at 5.9 keV	6.8 at 5.9 keV
Theoretical energy resolution (eV)	–	–	2.6 eV at 5.9 keV	1.5 eV at 5.9 keV	–
Material of absorber	Bi	Bi	Cu	Cu	Cu
Size of absorber	$0.4 \times 0.4 \times 0.3\,mm^3$	$0.4 \times 0.4\,mm^2$	$150 \times 150 \times 2\,\mu m^3$	$100 \times 100 \times 4\,\mu m^3$	$240 \times 240 \times 0.8\,\mu m^3$
Sensitivity of TES (α)	–	–	30	–	–
Temperature of operation	70	70	–	–	104
Response time, μs or count rate (cps)	500 cps	700 μs	100 μs	100 μs	–
Reference	[45]	[46]	[47]	[47]	[47]

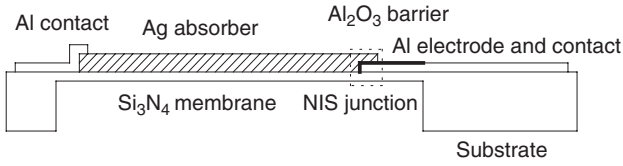


Figure 1.18 A schematic diagram of the NIS-based detector. The device includes a Ag–Bi sandwich absorber deposited on top of an Si_3N_4 membrane. The right side of the absorber is attached to a superconductor counter-electrode through a 1–2-nm thick Al_2O_3 insulator barrier. Two Al leads attached to the absorber and the top electrode are necessary to apply the bias voltage and for collecting the output charge produced as a result of interactions of incoming photons/phonons with the absorber

When the conditions $T \ll T_c(\mathrm{Al}), k_b T \ll eV, k_b T \ll \Delta$ are met, the equation (1.4.1) can be written in a simplified form as follows

$$I \approx \frac{\sqrt{2\pi k_b T_{el} \Delta}}{2eR_N} e^{-\frac{\Delta - eV_b}{k_b T_{el}}} \tag{1.4.2}$$

Thus, an increment in the tunnelling current is the measure of the electron temperature rise caused by the interaction of an X-ray incoming photon with the absorber. A SQUID current preamplifier transforms the current pulses into voltage suitable for further digitisation, shaping and processing.

The rise time of the electron temperature depends on the material of the absorber, its volume and the temperature of operation. In real terms, it is well under 1 μs, so that the slew rate of a SQUID current amplifier defines the rise time of the output voltage. The decay time is determined by the electronic heat capacity of the absorber, C, and the thermal conductance between electrons in the absorber and a heat sink, similar to all types of calorimeters, is given as

$$T_{el}(t) = \frac{E}{C} e^{-\frac{t}{\tau_d}} \tag{1.4.3}$$

with the thermal relaxation time constant $\tau_d = \frac{C}{G}$.

Assuming that the tunnelling of hot electrons through the NIS removes all deposited energy from absorber, thereby cooling the remaining electron condensate, τ_d is equal to the tunnelling time τ_{NIS}. The

thermal conductance of NIS junction as a heat link has been derived in [49] in the following form

$$G = \frac{dP}{dT_{el}} = \frac{k_b I}{e}\frac{T_b}{T_{el}}\left(\frac{1}{2} + \frac{T_b}{T_{el}}\right) \tag{1.4.4}$$

with $T_b = \frac{\Delta - eV_b}{k_b} \approx 2T_{el}$.

The equation (1.4.4) suggests that G depends on the voltage bias of the junction V_b. Adjusting V_b allows one to find a compromise between the energy resolution and the pulse duration at the output of the detector.

In practice, the electron–phonon interaction contributes to cooling the electron system in the absorber. If we neglect some correlation between the electron–phonon rate and the tunnelling rate, the total relaxation time can be derived from the sum of these individual rates as follows

$$\tau_d = \frac{\tau_{NIS}\tau_{e-p}}{\tau_{NIS} + \tau_{e-p}} \tag{1.4.5}$$

Due to the finite diffusion time of "hot" electrons from a spot of interaction in the absorber to the base electrode of the NIS junction, τ_{dif}, the detector will exhibit a position-sensitive charge collection deficit. The loss factor is proportional to $\exp\left(-\frac{\tau_{diff}}{\tau_{e-p}}\right)$ [50]. The diffusion time constant, D, and the electron–phonon scattering time, τ_{e-p}, were evaluated in experiments with 200-nm thick Ag film to be 0.08 m^2/s and 2.8 μs respectively, at a temperature of 0.1 K. An upper tolerable limit for the charge collection deficit defines the maximum dimensions of the absorber, which can be derived from the following expression

$$< x >= (2\tau_{dif}D)^{\frac{1}{2}} \tag{1.4.6}$$

The energy resolution of the NIS detector is given by [51]

$$\Delta E = \zeta(k_b T_{el}^2 C)^{1/2}\left[1 + a\left(\frac{S_I^{SQUID}}{S_I^{NIS}}\right)\right]^{1/2} \tag{1.4.7}$$

where $a \approx \frac{1}{2+\frac{T_{el}}{T_b}}$, S_I^{SQUID} and $S_I^{NIS} = (k_b T_b)^2 \frac{2I}{e} H(\omega)\Delta\omega$ denote the power of the SQUID amplifier noise and the power of shot noise in the NIS tunnel junction, respectively, and $H(\omega)$ is the transfer function of the pulse-shaping filter in the frequency range of interest, $\Delta\omega$.

A complete schematic diagram of the detector based on the NIS microcalorimeter is shown in Figure 1.19. The NIS junction is embedded in the input circuitry of the fast SQUID preamplifier. The voltage source, V_b, provides the bias for the junction. Incoming photon–phonon energy E generates the tunnelling current, which is transformed by the SQUID into the output voltage.

According to equation (1.4.7), the energy resolution of NIS hot electron microcalorimeters can be as good as those of any other types of calorimeters including those based on the TES. At the same time, they preserve a relatively high count rate capability. The electrical time constant does not influence it because the normal resistance of the NIS tunnel junction is several hundred mΩ. From the practical point of view, the NIS detector does not require magnetic field due to the absence of the Fiske steps. They conveniently operate at a wider temperature range of the cold finger without a power bias and an electro-thermal feedback. Attachment of several junctions in parallel with the absorber (distributed array) will proportionally speed up the overall response due to the faster pumping out rate of created "hot" electrons. It is also important that NIS junctions are fully compatible with the multiplexing techniques developed for the voltage-biased TES detector.

Progress towards the thermodynamic limit with NIS hot electron microcalorimeters will require further advance in processing techniques of the primary sensor itself as well as improvement in the noise performance of SQUID current amplifiers to a level below $0.1\,\frac{\mathrm{pA}}{\sqrt{\mathrm{Hz}}}$. The Chalmers group have recently obtained some very promising results in this research [52].

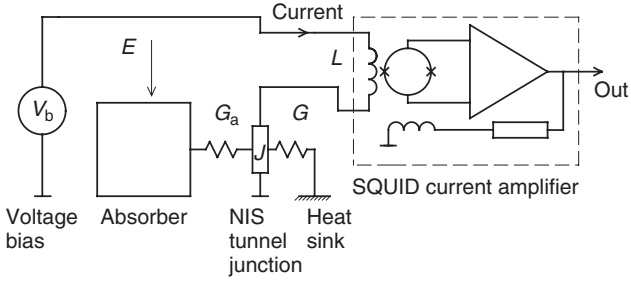


Figure 1.19 A complete schematic diagram of the detector based on NIS microcalorimeter. The NIS junction is embedded in the input circuitry of the fast SQUID preamplifier. The voltage source, V_b, provides the bias for the junction. Incoming photon/phonon energy E generates the tunnelling current, which is transformed by SQUID into output voltage

A practical X-ray detector with a 0.25 mm^2 area, 1 μm thick bismuth absorber was described in [53]. It demonstrated an energy resolution of 30 eV FWHM at MnKα and a pulse time constant of 20 μs. Earlier, the same group measured $\Delta E = 20$ eV and a decay time of 15 μs for a detector with an area of 0.01 mm^2, opposite of the best 100 μs achieved in TES microcalorimeters [54]. The quoted figures for the energy resolution are approximately a factor of 5 times the theoretical prediction given by (1.4.7). The discrepancy can partially be explained by the position sensitivity of detectors that affect the pulse shape of the tunnelling current. Readers should refer to [55], where a topic relevant to heat diffusion in the normal metal thin films is discussed in detail. Here we will only notice that an improved detector design, which, for instance, comprises a number of NIS tunnel junctions forming a distributed array, as well as implementation of an intelligent pulse-processing technique will further improve the energy resolution and the response time.

1.4.2 Phonon Detectors with Inductive Thermometers

In the previous two sections, we discussed "hot" electron bolometers that utilise resistive temperature transducers. This type of thermometers worsen the energy resolution of the whole detector by contributing the Johnson thermal fluctuations or/and some extra noise caused by various macroscopic discrete processes. In the TES, working in the mixed mode between the normal and superconducting states, this extra noise originates from the flux slippage or the Abrikosov vortex motion. NIS junctions suffer from shot noise and non-equilibrium superconducting state near the barrier due to the high concentration of phonons produced by the recombination of quasiparticles, further increasing temperature of tunnelled electrons. All these noise components can be eliminated in the inductive types of thermometers.

The kinetic inductance of a superconducting microstrip line separated from a superconductor ground plane by a very thin dielectric is given by [56]

$$L_k(T) = \mu_0 \frac{l_s}{w_s} \frac{\lambda_{gp}^2}{t_{gp}} \tag{1.4.8}$$

where μ_0 is the magnetic permeability of the free space, l_s and w_s denote the length and the width of the microstrip line respectively, t_{gp} and

$\lambda_{gp}(T) = \frac{\lambda_{gp}(0)}{\sqrt{1-\left(\frac{T}{T_c}\right)^2}}$ are the thickness and the magnetic penetration depth of the ground plane superconductor respectively, with T being the temperature of the ground plane, T_c, the critical temperature of the ground plane material and $\lambda_{gp}(0)$, the penetration depth at zero temperature.

An electric circuit diagram of the kinetic inductance thermometer coupled to a SQUID current amplifier is shown in Figure 1.20. The principle of operation of this system is based on storing a magnetic flux Φ in a superconducting loop, which cannot be changed unless the loop is unlocked. In this case, the current in the input circuit depends on the kinetic inductance as follows

$$I_k(T) = \frac{\Phi}{L_k(T) + L_{SQUID} + L_0} \tag{1.4.9}$$

where L_{SQUID} and L_0 represent the SQUID inductance and the residual geometrical inductance of the thermometer respectively.

Since the input circuitry does not include resistive components, the noise and the electrical time constant of the thermometer are determined by the energy resolution and the 3-dB bandwidth of the SQUID preamplifier, respectively. According to [56], a temperature resolution better than $1\,\frac{nK}{\sqrt{Hz}}$ should be possible for DC-biased circuitry.

Despite the good potential energy resolution, the inductive thermometer did find widespread applications in cryogenic detectors. Being insensitive directly to the electron temperature of the absorber, they can utilise a relatively slow chain of electron–phonon, phonon–phonon and

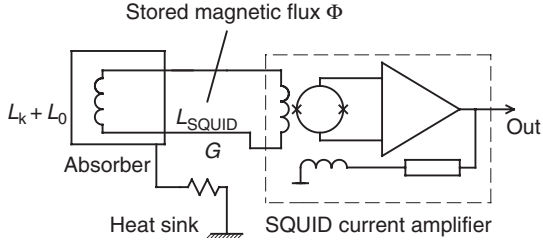


Figure 1.20 An electric circuit diagram of the kinetic inductance thermometer coupled to a SQUID current amplifier. The principle of the system is based on storing a magnetic flux Φ in a superconducting loop, which cannot be changed unless it is unlocked. The current in the loop varies with the kinetic inductance. Since the circuit does not include resistive components, the noise and the electrical time constant of the thermometer are defined by the SQUID amplifier

phonon–electron interactions. This puts them into the same range of microcalorimeters read out by conventional semiconductor thermistors.

REFERENCES

[1] L. Strueder. "High resolution imaging X-ray spectrometers". *Nuclear Instruments and Methods in Physics Research* **A454**, 73–113 (2000).

[2] G. Knoll. *Radiation Detection and Measurement*. John Wiley & Sons, New York, 1989, p. 116.

[3] N. Rando, A. Peacock, A. van Dordrecht, C. Fogen, R. Engelhardt, B. Taylor, P. Gare, J. Lumley and C. Pereira. *Nuclear Instruments and Methods in Physics Research* **A313**, 173 (1992).

[4] A. Zehnder. *Phys. Rev*. **B52**, 12858 (1995).

[5] A. Rothwarf and B. Taylor. *Phys. Rev. Lett*. **19**, 27 (1967).

[6] A. Barone. *Nuclear Physics B (Proc. Suppl.)* **44**, 645 (1995).

[7] N. Booth. *Appl. Phys. Lett*. **50**, 28 (1987).

[8] http://www-cxro.lbl.gov/optical_constants/atten2.html

[9] G. Brammertz, A. Golubov, A. Peacock, P. Verhoeve, D.J. Goldie and R. Venn. *Physica C* **350**, 227 (2001).

[10] M. Kurokado. *Nuclear Instruments and Methods in Physics Research*. **196**, 275 (1982).

[11] L. Li, C.M. Wilson and D.E. Prober. *Applied Superconductivity Conference* 2002.

[12] D. Goldie, P. Brink, C. Patel, N. Booth and G. Salmon. *Appl. Phys. Lett*. **64**, 3169 (1994).

[13] N. Booth and D. Goldie. *Supercond. Sci. Technol*. **9**, 493 (1996).

[14] L. Li, L. Frunzio, K. Segall, C.M. Wilson and D.E. Prober, A.E. Szymkowiak and S.H. Moseley. *Nuclear Instruments and Methods in Physics Research* **A444**, 228 (2000).

[15] R. den Hartog, A. Kozorezov, D. Martin, G. Brammertz, P. Verhoeve, A. Peacock, F. Scholze and D. Goldie. Large Format Distributed read Out Imaging Devices for X-ray Imaging Spectroscopy. www: http://astro.estec.esa.nl.

[16] M. van den Berg. Development of a High Resolution X-ray Spectrometer based on Superconductive Tunnel Junctions. PhD Thesis, 1999.

[17] L.Li, L. Frunzio, C.M. Wilson and D.E. Prober. *J. Appl. Phys*. **90**, 3645 (2001).

[18] S.B. Kaplan, C.C. Chi, D.N. Langenberg, J.J. Chang, S. Jafarey and D.J. Scalapino. *Phys. Rev*. **B14**, 4854 (1976), **B15**, 3567(E) (1977).

[19] S. Friedrich, K. Segall, M.C. Gaidis *et al. Appl. Phys. Lett*. **71**, 3901 (1997) and V. Narayanamurti, R. Dynes, P. Hu, H. Smith and W. Brinkman. *Phys. Rev*. **B14**, 4854 (1976).

[20] P. Brink. PhD Thesis. Oxford University, 1995 (unpublished).

[21] I.O. Kulik. *JETP Letters* **2**, 84 (1965).

[22] J. le Grand, M. Bruijn, M. Frericks *et al. Proceedings of the 4th Workshop on Low Temperature Detectors for Neutrinos and Dark Matter*, Edition Frontieres, 265 (1992).

[23] M. Ohkubo, K. Tanabe, K. Suzuki *et al*. 8th International Workshop on Low Temperature Detectors, Munich (1997), 1999.

[24] R. den Hartog, D. Martin, A. Kozorezov *et al*. Distributed Read-Out Imaging Devices for X-ray Imaging Spectroscopy. ESA Preprint, ESLAB2000/018/SA.

[25] D. Prober, C. Wilson, L. Li *et al. Private communication* (2002).
[26] L. Li, L. Frunzio, C. Wilson *et al. IEEE Trans. Appl. Supercond.* **11**, 685 (2001).
[27] K. Segall, C. Wilson, L. Frunzio *et al. Appl. Phys. Lett.* **76**, 3998 (2000).
[28] R. den Hartog, A. Kozorezov, J. Wigmore *et al.* 9th International Workshop on Low Temperature Detectors, Munich (1997).
[29] A. Kozorezov, J. Wigmore, A. Peacock *et al.* 9th International Workshop on Low Temperature Detectors, Munich (1997).
[30] D. McCammon, R. Almy, E. Apodaca *et al. Nuclear Instruments and Methods in Physics Research* **A436**, 205 (1999).
[31] E. Silver, M. Le Gros, J. Beeman and E. Haller. *X-ray Spectrometry* **25**, 115 (1996).
[32] J. Zhang *et al. Phys. Rev.* **B57**, 4472 (1998).
[33] W. Seidel, G. Foster, W. Christen *et al. Phys. Lett.* **B236**, 483 (1990), M. Frank, D. Dummer, S. Cooper *et al. Nucl. Instr. And Meth.* **A345**, 367 (1994), F. Proebst, M. Frank, S. Cooper *et al. J. Low Temp. Phys.* **100**, 69 (1995).
[34] V. Nagel, A. Nowak, H. Gebauer *et al. J. Appl. Phys.* **78**, 4262 (1994).
[35] B.A. Young *et al. J. Appl. Phys.* **86**, 6975 (1999).
[36] J. Martinis *et al. Nuclear Instruments and Methods in Physics Research* **A444**, 23 (2000).
[37] F. Wellstood, C. Urbina and J. Clarke, *Appl. Phys. Lett.* **54**, 2599 (1989).
[38] E. Swartz and R. Pohl. *Rev. Mod. Phys.* **61**, 606 (1989).
[39] M. Roukes, M. Freeman, R. Germain *et al. Phys. Rev. Lett.* **55**, 422 (1985).
[40] O. Meier, S. Uchaikin, F. Proebst and W. Seidel. *Int. Phys. Conf.*, ser. 158, 735 (1997).
[41] K. Irvin. *Appl. Phys. Lett.* **66**, 1998 (1995).
[42] S. Labov, M. Frank, J. le Grant *et al.* 7th International Workshop on Low Temperature Detectors, Munich (1997).
[43] K. Irvin and B. Cabrera. US Patent 5,641,961 (1997).
[44] O. Meier, M. Bravin, M. Bruckmayer *et al. Nuclear Instruments and Methods in Physics Research* **A444**, 350 (2000).
[45] D. Wollman, S. Nam, D. Newbury *et al. Nuclear Instruments and Methods in Physics Research* **A444**, 145 (2000).
[46] K. Irvin, G. Hilton, J. Martinis *et al. Nuclear Instruments and Methods in Physics Research* **A444**, 184 (2000).
[47] W. Bergmann Tiest, H. Hovers, W. Mels *et al. AIP Conf. Proc.* 605 (9th Int. Workshop on Low Temp. Detectors, Madison, Wisconsin), 199 (2002).
[48] L. Solymar. Superconducting Tunnelling and Applications. Wiley, NY, 1972.
[49] N. Nahum, T. Eiles and J. Martinis. *Appl. Phys. Lett.* **93**, 733 (1993).
[50] H. Netel, J. Jochum, S. Labov *et al. Applied Superconductivity Conference*, 1996, Pittsburgh, LLNL Preprint, UCRL-JC-125226.
[51] M. Nahum, J. Martinis and S. Castles. *J. Low Temp. Phys.* **93**, 733 (1993).
[52] L. Kuzmin. *Private communication* 10 June, 2002.
[53] J. Martinis. *Nuclear Instruments and Methods in Physics Research* **A370**, 171 (1996).
[54] M. Nahum and J. Martinis. *Appl. Phys. Lett.* **66**, 3203 (1995).
[55] M. Roukes, M. Freeman, R. Germain *et al. Phys. Rev. Lett.* **55**, 422 (1985).
[56] D. McDonald. *Appl. Phys. Lett.* **50**, 775 (1987).

2

Front-end Read-out Electronic Circuits

INTRODUCTION

The major objective of the analogue read-out electronics is to preserve information contained in the detector response to the incident photon–phonon energy. For preamplifiers, specifically, this means that the output voltage pulse must bear a known relationship with respect to the charge or the current amplitude of the input pulse. It is also important that this relationship remains stable over the data acquisition time. The influence of temperature fluctuations, the thermal noise, the variable count rate, the pile-up, the electromagnetic interference and the back reaction on the primary sensor, all these and other adverse effects should be reduced to a level well below the intrinsic energy resolution of a detector. In practice, the low energy resolution and throughput of detector systems is often limited by the analogue channel of spectrometers rather than by the primary sensor. Thus, evolution of nuclear analytical instrumentation requires parallel advancing in all technologies involved such as the nucleonics, system architecture and software of spectrometers, not forgetting about a very important link between them: the fast low-noise analogue–digital interface.

As the energy resolution of a spectrometer improves, the specification requirements for the analogue channel become more stringent. For instance, according to the previous chapter, the cryogenic detectors intended for X-ray microanalysis are capable to resolve spectral peaks

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin
 ISBN: 0-470-85759-5

with a width better than 0.1% FWHM. For such a resolution, the noise linewidth (NLW) of the whole system should add no more than 0.02% to it, as opposite to a figure of 0.4% in the best semiconductor detectors. The stability of gain factor should be of the same order of magnitude.

In addition to the energy resolution, some applications like the time-of-flight mass-spectrometry or multiplexed read-out schemes place very strict constraints for the timing accuracy. Thus, the slew rate of a preamplifier may need to be such that the detector pulse rise time is not extended by more than 1% [1]. Low noise helps also in improving the time resolution.

In principle, SQUID-based preamplifiers are capable to read out all superconductor detectors. Whereas they are arguably the best choice to read out NIS, TES calorimetric devices, the STJs can also be matched to either the JFET or the radio-frequency single electron transistor (RF SET). In the latter case, however, a SQUID would still be required to amplify the signal of the RF SET. Unlike semiconductor quasiparticle detectors, STJs with a 1–2-nm thick tunnel barrier have a much larger capacitance per area, typically of several μFcm^2. As an example, measurements of a Nb-based STJ with an area of $178 \times 178\,\mu m^2$ yielded a capacitance as large as 1.8 nF [2]. Therefore, matching the tunnel junctions with an FET transconductance amplifier proves to be quite a problem.

There is a number of other constrains imposed on the design of preamplifiers working with superconductor detectors. They mostly relate to the ultralow temperature cryogenics, in particular, a comparatively low cooling power of refrigerators, available at a temperature below 100 mK and perceptivity of cryogenic detectors to thermal radiation photons. Both the reasons do not allow the accommodation of active semiconductor components on/near the cold finger itself. As a result, a distance in excess of 1 m may separate an STJ (250 mK) and, say, a first-stage JFET (150 K) of preamplifier. Connecting cables can be twice as long as that distance because they have to be properly heat-sinked to a metallic frame often attached to the common ground. The combination of the separation, the large junction capacitance, plus a stray cable capacitance affect adversely the rise time and the settling time of preamplifiers giving rise to a substantial back reaction on the bias stability of the detector.

A SQUID current preamplifier is widely considered as an alternative to the FET counterpart. Indeed, it operates close to the primary sensor at the same temperature, dissipates a small amount of power, provides a condition for an ideal DC voltage bias and, finally, practically does not

throw any noise back into the junction. Being a two-terminal device, the SQUID can easily be incorporated into any grounding arrangement. The last feature is particularly valuable in a large format multiplexer configuration. At high frequencies, however, the inductive impedance of the SQUID has some effect on the loadline of the STJ. Besides, some additional damping must be foreseen in the input circuitry to critically suppress the L_{SQUID} C_{STJ} resonance.

Recently developed RF SET seem to combine the positive qualities of both the JFET and the SQUID. They are adequately fast, operate in a mK temperature range and dissipate very low power. So far, they have not been widely utilised in practical instrumentation. First, this is due to the involvement of the expensive nanometer technology in their manufacture. Second, the RF SET requires properly terminated cables for the transmission of the high-frequency signal from the room temperature environment to the cold finger and back. This makes it rather difficult to organise large format multiplexers on their base right now. In the future, however, in combination with superconductor SQUID analogue-to-digital converters, when the room temperature is eliminated, they may become an alternative solution to JFETs and analogue SQUID current amplifiers targeting applications involving fast STJ detector arrays. A detailed description of RF SET devices is given in [3].

The present chapter examines mostly JFET and SQUID preamplifier techniques. We will deal with their dynamics and the noise properties. Each section will include practical examples of circuit diagrams, results of their simulations or calculations in the analytical form and some application notes, which readers may find useful in the development of analogue channels in their detector systems.

2.1 FET TRANSCONDUCTANCE PREAMPLIFIERS

2.1.1 Principles of JFET Transconductance Amplifiers

The shape, magnitude and the noise content of a signal delivered at the input of a digital signal processor are determined by the detector as much as the preamplifier. All superconductor detectors with a reasonably large absorber have one important specific feature. Due to a relatively slow diffusion velocity of carriers, the shape of the charge output impulse produced in response to, say, X-ray photon depends on a position on the absorber, where the interaction occurred. Provided that losses are negligible, it is the charge or the integral of the total current rather than

an instantaneous amplitude of the current that bears the proportionality to a deposited energy E. The integration of the charge is normally performed by the feedback capacitor C_F of the charge sensitive preamplifier (CSA), which is normally used with all types of quasiparticle detectors, including the STJ.

Charge sensitive amplifiers intended for the nuclear instrumentation can be classified into two types depending on which reset method of the voltage across the feedback capacitor is utilised: a continuous or pulsed one. Figure 2.1 gives two examples of preamplifiers with the continuous reset. In the circuit diagram shown in Figure 2.1(a) the reset time constant is fixed by feedback components $R_F C_F$. The presence of R_F adds a current noise $I_R^2 = \frac{4k_b T}{R_F}$ to the input. In order to reduce it to a tolerable level, a value of the resistor should be chosen such that the following condition is met

$$R_F > \frac{2k_b T}{q(I_d + I_f)} \tag{2.1.1}$$

where I_d, I_f denote the leakage current of the detector and the JFET.

On the other hand, the time constant $R_F C_F$ must be adequate to fulfil its direct designation preventing the saturation of the preamplifier. Therefore, in practice, R_F is defined by the maximum frequency of incoming events (the count rate) and by the leakage current of the detector, rather than by the noise consideration (2.1.1), It is easy to show that the discharging resistor of any reasonable value placed at a room temperature would represent the dominating noise source of the

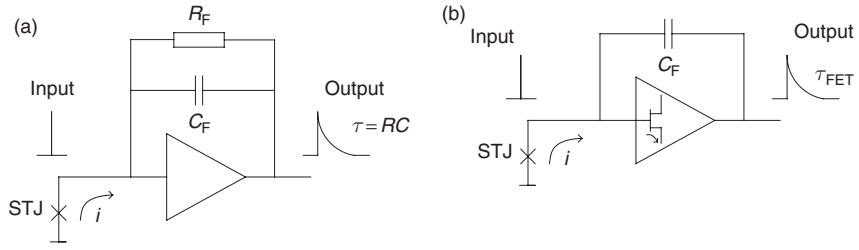


Figure 2.1 Schematic representation of preamplifiers employing the continuous reset techniques based on the feedback resistor R_F (a) and the forward-biased gate-source junction of the input JFET (b). The forward-biased JFET eliminates the thermal noise current generated by R_F, but increases the shot noise contribution proportionally to the total leakage current flowing into the input terminal of the preamplifier. Both preamplifiers yield ordinary exponentially decaying pulses

superconductor detector system. Of course, this noise can be practically eliminated by cooling R_F to a base temperature of a cold finger on which primary sensor itself is mounted. But this would make the preamplifier unstable due to the cross-talk between two placed alongside wires carrying the input and output signals.

Despite the extra noise, this type of CSA is widely used to read out the signal of the STJ. The benefit comes from the fact that the presence of R_F controls the DC potential of the input terminal of the amplifier. As a result, the junction directly attached to this terminal can be biased and read out with a single wire, thereby operating in the much preferred quasivoltage-biased mode (Figure 2.2). This configuration is particularly attractive for multiplexing the large format arrays of devices.

The circuit diagram, shown in Figure 2.1(b), does not include R_F. The leakage and signal currents of the detector have a DC path through the gate of the input JFET, which works with its gate-to-source junction slightly forward biased [4].

Alternatively, the input-JFET can also operate in an inverse-biased mode. In this case, a potential difference between the gate and the transistor channel is discharged by the impact ionisation in the transistor channel close to the junction formed by the gate and the drain at the end of the channel [5].

Both discharging mechanisms act in such a way that the signal and leakage currents are automatically self-compensated by controlling the gate voltage. Therefore, preamplifiers of this type are designed with a

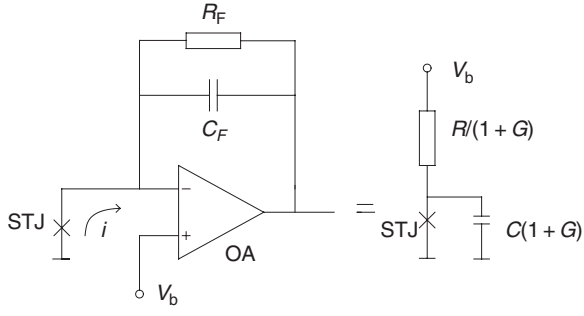


Figure 2.2 Charge-sensitive operational amplifier (CSA). Despite the extra noise, this type of CSA is widely used to read out the signal of the STJ. The benefit comes from the fact that the presence of R_F controls the DC potential of the input terminal of amplifier. As a result, the junction directly attached to this terminal can be biased and read out with a single wire, thereby operating in the much preferred quasi-voltage-biased mode. This configuration is particularly attractive for multiplexing the large format arrays of devices

reduced DC and low-frequency gain providing a dynamic range sufficient for accommodating the input voltage variations without the saturation of the output. Special precaution measures must be taken to prevent these voltage variations affecting the stability of the voltage bias across the superconductor tunnel junction. One of these is to utilise an AC coupling technique.

The DC path in the both types of preamplifiers with the forward- or reverse-biased JFET appears essentially as an extra leakage current generating the shot noise. This doubles the shot noise contribution into the total energy resolution originated from the leakage currents only and adds a component proportional to the average signal current. As the average signal current is a direct function of the input count rate, so becomes the noise linewidth. Extra noise, if it exceeds the shot noise of the primary device, cannot be tolerated. Cryogenic detectors would start losing against best semiconductor counterparts exhibiting a low energy resolution as low as 60 eV. Thus, in order to eliminate extra noise and preserve the high stability of the junction operating point, cryogenic detectors are read out with the charge sensitive preamplifiers operating in the, so-called, "pulse-reset mode". The schematic configuration of the pulse-reset preamplifier is shown in Figure 2.3.

According to Figure 2.3, the technique provides an intermittent path for the charge restoration. The overall preamplifier stage includes two additional components: a base-grounded bipolar transistor and a comparator. Once the output of the preamplifier reaches a predetermined voltage level, V_{ref}, the comparator generates a short pulse injecting a current into the gate of the FET that discharges C_F. The duration of the reset can be as low as 1 μs. This makes the technique compatible with very high count rate applications [6]. A more elaborate design of the fast pulse reset technique was presented in [7]. It further reduces an approximately 1 pF parasitic capacitance introduced by the discrete reset bipolar transistor. This extra capacitance, however, is much lower compared to a capacitance of an STJ, which is normally well above 100 pF, and, therefore, has a negligible effect on its total energy resolution.

The outputs of the pulse-reset preamplifiers and preamplifiers employing the continuous reset techniques have completely different shapes. Whereas the first one produces a steadily increasing voltage "ramp" caused by integration of leakage currents with superimposed on it event sharp steps (Figure 2.3), the latter yields ordinary exponentially decaying pulses. Each technique obviously needs to be matched to a digital spectrometer, which would have an individually tuned hardware architecture as well as dedicated software algorithms. There are, however,

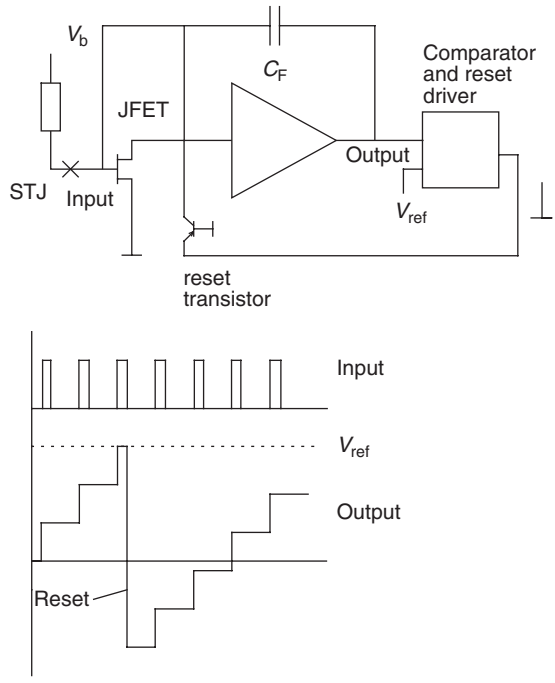


Figure 2.3 Schematic representation of the preamplifier employing the pulse-reset technique. The technique provides an intermittent path for the charge restoration. The preamplifier stage includes two additional components: a base-grounded bipolar transistor and a comparator. Once the output of the preamplifier reaches a predetermined voltage level, V_{ref}, the comparator generates a short pulse injecting a current into the gate of the FET that discharges C_F. The duration of the reset can be as low as 1 μs. This makes the technique compatible with very high count rate applications. The outputs of the pulse-reset preamplifiers yields a steadily increasing voltage "ramp" caused by integration of leakage currents with superimposed on it event sharp steps

commercially available universal spectrometers with the built-in capability to recognise and process both types of signals. An example of a such spectrometer can be found in [8].

2.1.2 Settling Time of a Preamplifier

The settling time is defined as the time elapsed from the application of an ideal instantaneous step input to the time at which the closed loop amplifier output has entered and remained within a specified error band γ [9]. The settling time is a very important parameter for superconductor detectors. This is due to the fact that the shape of their output current

varies with a position on the absorber where the interaction with an incoming photon actually occurs. If the settling time is comparable or longer than the duration of the STJ output current pulse, then the evaluation of the event energy from the pulse height at the preamplifier output may contain a substantial ballistic deficit error. The error becomes more pronounced at shorter peaking times, which deal with high input count rates.

The maximum count rate capability of modern STJ detectors reaches approximately 10 kcps (kilocounts per second). The rate can be handled by a pulse shaping circuit with a peaking time of $\tau_0 = 20\,\mu s$ which produces a few per cents of a dead time. In the first approximation, the pulse-shaping circuit can be represented by a selective filter with a central frequency $f_0 = \frac{1}{2\pi\tau_0} \approx 8\,\text{kHz}$. Preamplifier specifications for the linearity and stability of the gain factor define directly the error band as well as the open loop gain $K_0(f_0)$ at the central frequency f_0. For an error band $\gamma = 0.05\%$, the loop amplification must be in excess of $K_0(f_0) = \frac{1}{\gamma} \approx 2000$. Assuming a 20-dB/dec roll-off of the open loop gain, the gain-bandwidth product or, in other words, the unity gain frequency of the amplifier must be as large as $f_m = K_0(f_0)f_0 = 16\,\text{MHz}$.

The settling time of the preamplifier assembled with a detector is given by

$$t_s^{\min} \approx \frac{C_d + C_k + C_i}{C_F} \frac{1.2}{f_m} \tag{2.1.2}$$

where C_d, C_k and C_i represent capacitance of the detector, connecting cable and the input FET respectively. Presence of parasitic capacitance requires a further increase in the gain-bandwidth product of the amplification channel.

A 20-dB/dec roll-off is quite important to ensure the minimum t_s and avoid undesirable undershoots or overshoots in the output pulse. This roll-off is characteristic for a single-pole minimum phase inertial circuit. Since all practical preamplifiers include a number of stages each with its own pole, a designer should take a particular care to properly match them or, if needed, compensate the action of some of them.

The Laplace transform of the open loop gain of a two-pole amplifier can be represented in the following form

$$G = \frac{1 + mK(f_0)\tau_0 p}{1 + [K_0(f_0) + 1]m\tau_0 p + mK_0(f_0)\tau_0^2 p^2} \tag{2.1.3}$$

where m is the mismatch factor between poles.

For a perfectly matched circuit $m = 1$, $G = \frac{1}{1+\tau_x p}$ produces frequency response with a 20-dB roll-off shown in Figure 2.4 (middle line). Here $\tau_x = \frac{2\pi}{f_m}$. For a mismatch $m = 4$, a loop gain is given by

$$G = \frac{1}{1 + 0.996\tau_x} \frac{1 + 800\tau_x p}{1 + 803\tau_x p} \tag{2.1.4}$$

It is illustrated by a top curve in Figure 2.4. As we can see, the settling time extends to a value of $t_s^{m=4} \approx 3.6 t_s^{\min}$.

In the case of the opposite mismatch, that is, $m = 0.25$, the loop gain in the analytical form is as follows

$$G = \frac{1}{1 + 1.015\tau_x p} \frac{1 + 50\tau_x p}{1 + 49.2\tau_x p} \tag{2.1.5}$$

Its graphical representation is shown in Figure 2.4 (bottom curve). According to the plot, the settling time becomes even longer, $t_s^{m=1/4} = 49.2\, t_s^{\min}$. Figure 2.5 shows step responses of a two-pole amplifier calculated for all three cases under discussion. Each mismatch would substantially reduce the count rate capability of the detector system or, alternatively, give a contribution into the ballistic deficit.

An example of a typical low-noise JFET preamplifier is shown in Figure 2.6. It consists of an input low-noise cascode stage (JFET, T_1; a bipolar transistor, T_2; and a load resistor, R) and an operational

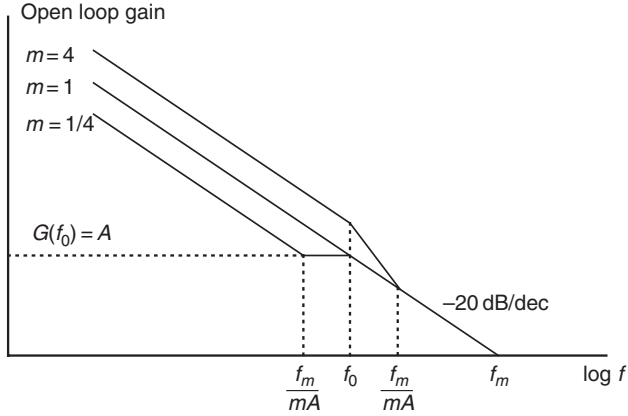


Figure 2.4 The frequency response of a preamplifier with two major poles. Three curves correspond to mis-match factors m: 0.25, 1 and 4 respectively, at a central frequency of the pulse-shaping network. The preamplifier yields the fastest response without overshot or undershot at $m = 1$

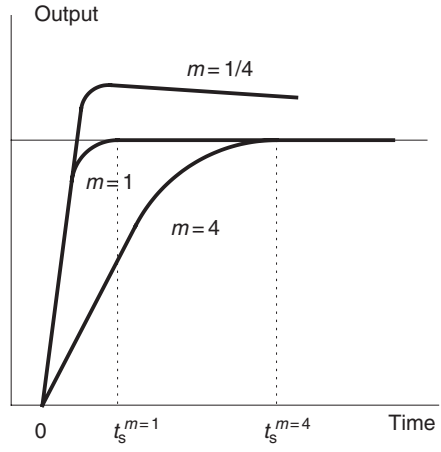


Figure 2.5 The time response of a preamplifier with two major poles to a step-like current impulse. Three curves correspond to mis-match factors m: 0.25, 1 and 4 respectively, at a central frequency of the pulse-shaping network. The preamplifier yields the fastest response without overshot or undershot at $m = 1$

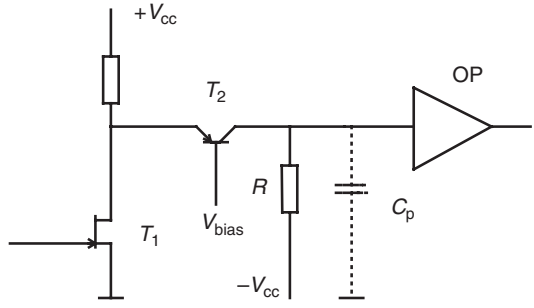


Figure 2.6 An example of practical implementation of the low-noise JFET preamplifier. It consists of an input low noise cascode stage including JFET, T_1; a bipolar transistor, T_2; a load resistor, R; and an operational amplifier, OP. It has two significant open loop poles: the first one, representing the maximum frequency of the 3-dB bandwidth of the OP and, a second one, $\omega_c = (RC_p)^{-1}$ formed by the load resistor and the parasitic capacitor between the T_2 collector and the ground including the input capacitance of the OP

amplifier OP. It has two significant open loop poles: the first one, ω_{op}, representing the maximum frequency of the 3-dB bandwidth of the OP with an open loop gain, $G(f_0)$, and the second one, $\omega_c = (RC_p)^{-1}$, formed by the load resistor and the parasitic capacitor between the T_2 collector and the ground including the input capacitance of the OP. The two-pole preamplifier will provide a maximum gain at the central

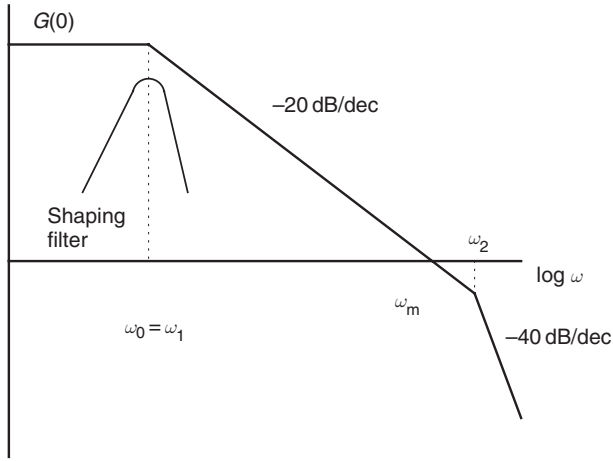


Figure 2.7 The Bode plot of a two-pole preamplifier with the optimised frequency response. The preamplifier will provide a maximum gain at the central frequency of the shaping filter and a minimum settling time without ringing if its closed loop poles are complex conjugate at $p_{1,2} = \alpha_0 \pm j\omega_0$ with $\omega_0 = 2\pi f_0 = \alpha_0$ and $\omega_1 = \omega_0$ [10]. On the Bode plot, these conditions are equivalent to setting the pole ω_2 at a frequency twice of the unity gain frequency $\omega_m = G(0)\omega_1$

frequency of the shaping filter and a minimum settling time without ringing if its closed loop poles are complex conjugate at $p_{1,2} = \alpha_0 \pm j\omega_0$ with $\omega_0 = 2\pi f_0 = \alpha_0$ and $\omega_1 = \omega_0$ [10]. On the Bode plot, these conditions are equivalent to setting the pole ω_2 at a frequency twice the unity gain frequency $\omega_m = G(0)\omega_1$, as shown in Figure 2.7.

2.2 DYNAMICS AND NOISE OF JFET AMPLIFIERS

A detector will demonstrate its intrinsic energy resolution and the peak-to-background ratio only if the noise contribution and the back reaction of a following preamplifier are negligibly small. The back reaction is essentially a non-compensation electrical signal, Δe, at the preamplifier input caused by inertial components present in the direct amplification channel as well as in the feedback (Figure 2.8). It appears during the propagation of the signal pulse from the primary detector and acts via variation of the voltage bias (changes the dynamic loadline) or/and injects extra carriers into the counter-electrode. Both processes alter the dynamics of the tunnelling in the junction distorting energy information containing in the output pulse. In the final energy-histogram, this

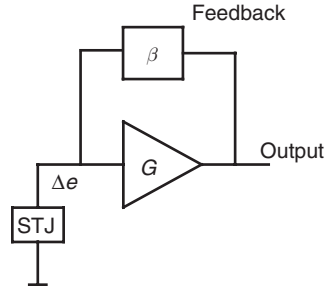


Figure 2.8 An STJ detector attached to the transconductance preamplifier. The circuit diagram shows a non-compensation electrical signal, Δe, at the preamplifier input caused by inertial components present in the direct amplification channel as well as in the feedback. It appears during the propagation of the signal pulse from the primary detector and acts via variation of the voltage bias and injects extra carriers into the counter-electrode. Both processes alter the dynamics of the tunnelling in the junction distorting energy information containing in the output pulse

will appear as the instability of characteristic peaks and the presence of tails (the deviation from the Gaussian distribution form).

According to the previous section, the minimisation of the back reaction requires an adequate gain-bandwidth product as well as a right positioning of poles on the Bode plot for the open loop gain of the amplifier. These requirements, however, may not always lead to the optimum design solution in terms of the noise performance. Although noise of the first stage of a cascaded amplifier is usually dominant, subsequent stages as well as biasing elements make also some contribution. In fact, the faster the preamplifier has to be, the larger the contribution becomes. At the system designing phase, this trade-off must be taken into consideration. Figure 2.9 illustrates the matter. It presents an

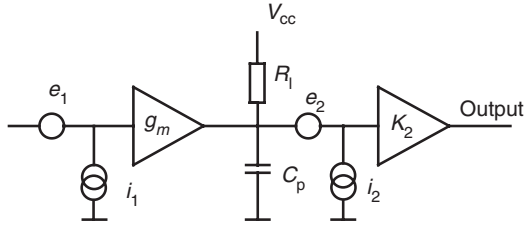


Figure 2.9 An equivalent schematic diagram of a low-noise cascaded preamplifier. The first stage is shown as a transconductance coefficient of an input-JFET loaded by a resistor R_l and a parasitic capacitance C_p. The following stages have an amplification factor K_2. Various current and voltage noise sources of the both stages are summarised by generators i_{n1}, i_{n2}, e_{n1} and e_{n2} respectively

equivalent schematic diagram of a low-noise cascaded preamplifier. The first stage is shown as a transconductance coefficient of an input-JFET loaded by a resistor R_l and a parasitic capacitance C_p. The following stages have an amplification factor K_2. The current and voltage noise of both the stages are summarised by generators i_{n1}, i_{n2}, e_{n1} and e_{n2} respectively.

The total voltage noise referred to the input of a preamplifier is given by

$$e_{ni}^2 = e_{n1}^2 + \frac{e_{n2}^2 + i_{n2}^2 R_l^2}{(g_m R_t)^2} \tag{2.2.1}$$

According to equation (2.2.1) R_l should be maximised in order to reduce contributions of e_{n2} and i_{n2}. This, however, may contradict to optimum positioning of the pole formed by $R_l C_p$ (see the previous section). The trade-off between the noise performance and the gain-bandwidth product rests entirely on the technical specifications of the input-JFET and the OA. Best commercial preamplifiers are still being built with custom designed JFETs and other discrete components as they outperform on-chip integrated devices routinely available from semiconductor mass manufacturers. When the "component quality" resource is exhausted, the further trade-off must involve seeking a compromise between the throughput and the energy resolution, which comes from the specifics of an intended application of the detector system.

Different configurations of the input-JFET stage help to tailor required noise versus slew rate performance. For instance, a cascode and differential stage, shown in Figures 2.10 and 2.11 respectively, have an amplification factor greater than 1. They are likely to be selected for an ultra low noise preamplifier. Figure 2.12 depicts a differential follower stage with a gain below the unity. It enables us to practically eliminate its pole in a frequency range of interest for STJ detectors.

2.2.1 Basic Static and Dynamic Parameters of JFET

The equivalent circuit diagrams that help to describe the static and dynamic behaviour of the JFET are shown in Figure 2.13. The gate and the channel are made of P- and N-type semiconductor material so that along the joint surface they form a PN junction. The PN junction normally operates in a reverse-biased mode (apart from a case of the continuous discharge), thereby providing the high-input gate-to-channel impedance.

The depth and the profile of the depletion region extending from the heavily doped gate electrode control the conductance of the channel.

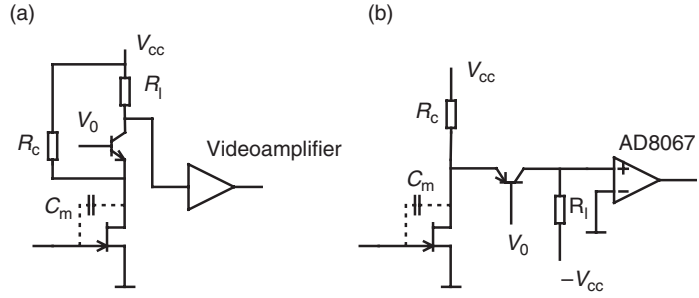


Figure 2.10 A circuit diagram of a preamplifier based on the unipolar cascode (a) and bipolar cascode (b). The cascode represents an amplifier stage consisting of a common source JFET immediately followed by a common base bipolar transistor. It has the particular advantage when an input JFET operates at a lower temperature (optimally $\sim$150 K) with some distance separating it from following stages. A common base bipolar transistor has the low input emitter-base impedance, thereby practically eliminating the pole formed by the Miller effect $R_l C_{gd}(1 + gmR_l)$ and reduces a loss in the amplification factor introduced by a parasitic capacitance of a cable connecting both parts of the stage. Cascodes operate in amplifiers with a single and double power supply respectively. The resistor R_c in the bipolar cascode must carry a much larger biasing current for both transistors. Thus, a positive power supply voltage $+V_{cc}$ must be selected such that a value of R_c is much larger than the input impedance of the common-base bipolar transistors

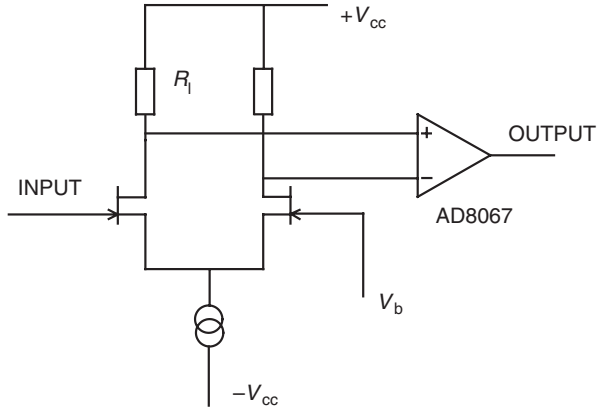


Figure 2.11 Fast low noise amplifier based on the common emitter differential stage. A differential amplifier provides an output proportional to the difference between signals delivered to its input terminals. In this case, the non-inverting channel can be used to indirectly apply a voltage bias to the STJ detector through the feedback resistor, whereas the inverting input terminal accommodates a sum point of the conventional transconductance amplifier with a continuous reset

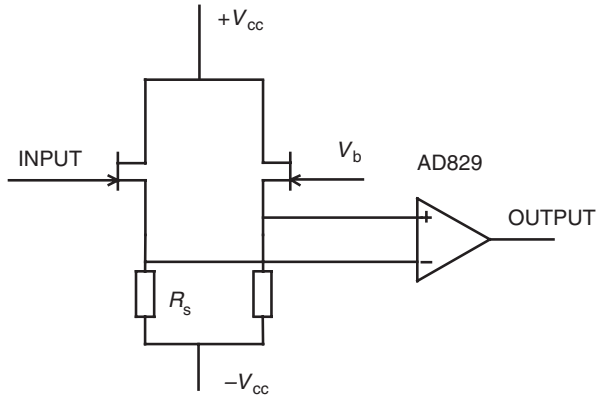


Figure 2.12 Ultrafast amplifier based on the differential follower stage. The front end provides only high impedance. The noise voltage of the following operational amplifier adds in quadrature to the noise of JFET without any attenuation. This type of the preamplifier has the shortest settling time

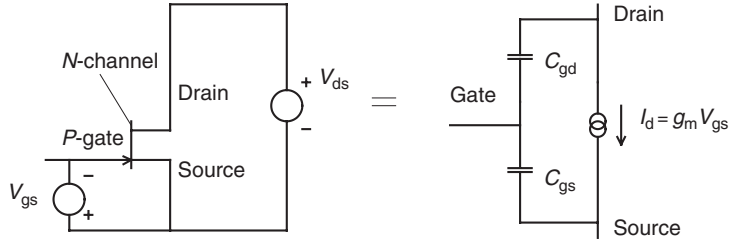


Figure 2.13 An equivalent circuit diagram of the N-channel JFET. It is desirable to operate the PN junction of the transistor in the reverse-biased mode, thereby providing the high input gate-to-channel impedance

Initially, the profile is influenced by a potential drop V_{ds} along the channel caused by a current flow I_d until a point when V_{ds} reaches a pinch-off value V_p. At $V_{ds} > V_p$ the drain current is almost independent of V_{ds} and entirely governed by V_{gs}. As V_{gs} increases, the width of the depletion region becomes larger until at some cut-off value, $V_{gs(off)}$, it occupies the whole channel bringing the drain current to virtually zero.

The Shockley power law equation shows the relation between the bias and the drain current [11]

$$I_d = I_{d0}\left(1 - \frac{V_{gs}}{V_{gs(off)}}\right)^n + I_{l|V_{gs}>V_{gs(off)}|} \qquad (2.2.2)$$

where I_{d0} is the drain current at $V_{gs} = 0$. I_l denotes the leakage current of the JFET. I_l is a function of V_{gd} and V_{gs}, as shown in Figure 2.14. Its major mechanism results from the thermal ionisation of carriers within the junction. Therefore, it is very beneficial to cool down the transistor because I_l halves in the magnitude for each 10 °C decrease in the junction temperature. An optimum operating temperature of JFETs is approximately 150 K. Below this point, the transconductance gains the downturn trend due to the freeze-out of carriers.

There is another component of I_l, which we mentioned before. It originates from carriers generated within the drain–gate region by the impact ionisation effect. The impact ionisation rate is a linear function of I_d and depends exponentially on V_{gd}. By adding small V_{gd} pulses, it is possible to create a very fast controlled discharging mechanism for the feedback capacitor [5].

One needs to remember here that the impact ionisation is also a function of the carrier mobility. Since the mobility has a negative temperature coefficient, the impact ionisation gate current appears at lower values of V_{ds} as temperature increases. Thus, the effective discharging circuit must include a thermometer that monitors the temperature of the cold plate on which the transistor is mounted.

The transconductance of JFET can be derived from the Shockley equation in the following form

$$g_m = \frac{dI_d}{dV_{gs}} = \frac{I_{d0}}{V_{gs(off)}}\left(1 - \frac{V_{gs}}{V_{gs(off)}}\right)^{n-1} \tag{2.2.3}$$

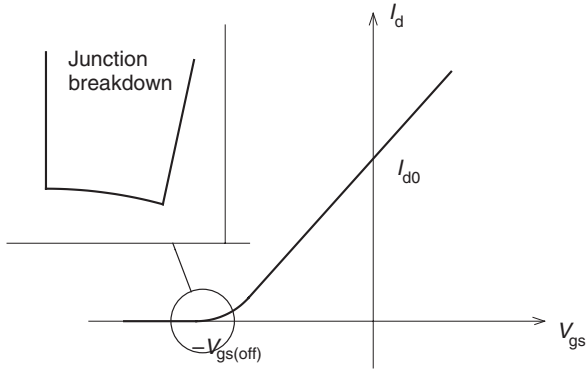


Figure 2.14 The drain current of the N-channel JFET as a function of the gate voltage. The insert illustrates an increase in the leakage current, as a result of the impact ionisation between the gate and the channel. This phenomenon is utilised to discharge the feedback capacitance in the charge sensitive amplifiers

The exponent n depends on the length of the channel. It ranges between 1.5 and 2. At an operating point $V_{gs} = 0$ and $I_d = I_{d0}$, where all best JFETs tend to work, the analytical expression for the transconductance can be presented in a simplified form, as follows

$$g_{m0} = -n\frac{I_{d0}}{V_{gs(off)}} \tag{2.2.4}$$

The output impedance R_{out} is another important parameter of the JFET. It helps to evaluate the gain loss due to a divider formed by R_{out} and the input impedance of the following stage as well as a connecting cable (if present). The output impedance can be found from the following expression

$$R_{out} = \frac{dV_{ds}}{dI_d} = \frac{V_{ds}\left(1 + \frac{V_{ds}}{2V_{gs(off)}}\right)}{I_d} \tag{2.2.5}$$

A typical value of R_{out} is approximately 10 Ω.

In order to achieve the maximum signal-to-noise ratio, the capacitance of the detector must be matched to the input capacitance of the read-out preamplifier. This condition is quite difficult to meet in the case of large-capacitance STJ detectors. One way to tackle the problem involves using several long channel JFETs connected in parallel [12].

The capacitance per unit area of the PN junction depends on the thickness of a depleted volume. This volume acts essentially as a dielectric dividing all three electrodes: gate–drain, gate–source. The input capacitance is given by

$$C_{gs} \approx \left(\frac{K}{0.6 + V_{gs}}\right)^{\frac{1}{n}} \tag{2.2.6}$$

where K is a constant defined by a material and an JFET design and an exponent n lies in the range of 2–3. The minimisation of C_{gs} is of paramount importance for low capacitive semiconductor detectors. The input capacitance of the JFET is less critical for STJ devices.

All major static and dynamic parameters of the JFET are mutually dependent. This is illustrated by the system of equations given below [13]

$$I_{ds} \approx \frac{1}{2}\mu\frac{C_{gs}}{L^2}V_{gs(off)}\left(1 - \frac{V_{gs}}{V_{gs(off)}}\right)^2 \tag{2.2.7}$$

$$g_m = \mu \frac{C_{gs}}{L^2} V_{gs(off)} \left(1 - \frac{V_{gs}}{V_{gs(off)}}\right) = \left(2\mu \frac{C_{gs}}{L^2}\right)^{\frac{1}{2}} I_d^{\frac{1}{2}} \qquad (2.2.8)$$

$$\frac{g_m}{C_{gs}} \approx \frac{\mu}{L^2} V_{gs(off)} \left(1 - \frac{V_{gs}}{V_{gs(off)}}\right) = \frac{1}{t_{el}} \qquad (2.2.9)$$

where μ is the electron mobility, L is the length of the channel and t_{el} is the electron transit time through the channel. The transit time defines the ultimate intrinsic gain-bandwidth product that can be achieved with a JFET design. It is given by

$$f_m = \frac{\mu}{L^2} V_{gs(off)} \left(1 - \frac{V_{gs}}{V_{gs(off)}}\right) \qquad (2.2.10)$$

For a JFET with $V_{gs(off)} = 1.5\,V$, f_m can be as large as 4 GHz. In real devices, however, it is somewhat lower due to the finite output impedance and parasitic capacitances.

2.2.2 Noise Characteristics of JFET

For future noise analysis of detector systems, it is convenient to represent the JFET as an ideal noise-free device with external noise sources, as shown in Figure 2.15. In this figure, di_1 is the Johnson current noise generated by the channel noise-resistance R_n. A value of the noise resistance was derived in [14] in the following form

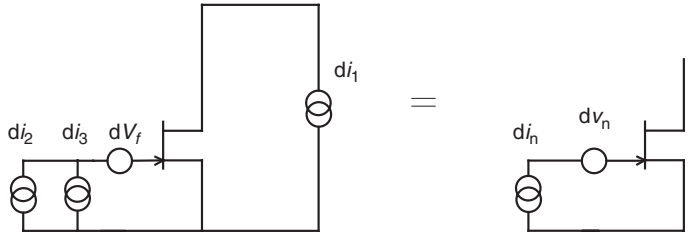


Figure 2.15 An equivalent noise diagram of the JFET. The voltage noise of the transistor originates from the Johnson drain current noise generated by the channel noise resistance. The equation indicates that dv_n is inversely proportional to the transconductance of the JFET. In this respect, the parallel connection of N transistors benefits from the noise reduction by a factor of $\frac{1}{\sqrt{N}}$

$$R_n = \frac{2}{3g_m} \tag{2.2.11}$$

Therefore, the current noise is given by

$$di_1^2 = \frac{4k_bT}{2\pi R_n}d\omega = \frac{4}{3\pi}k_bTg_m d\omega \tag{2.2.12}$$

Referred to the input this noise source appears as an input voltage noise

$$dv_1^2 = \frac{4}{3\pi g_m}k_bT d\omega \tag{2.2.13}$$

The voltage noise dv_n also includes a $1/f$ noise component

$$dv_f^2 = \frac{A_f}{\omega}d\omega \tag{2.2.14}$$

where $A_f = \frac{4}{3\pi}\frac{k_bT}{g_m}\omega_c$ with ω_c representing the corner frequency at which the spectral density of the $1/f$ noise is equal to $\sqrt{2}$ the spectral density of the white voltage noise.

The $1/f$ noise originates from trapping carriers in the Shockley–Read–Hall (SRH) generation–recombination centres in the junction depletion region. The centres are formed by impurities and crystal defects. They tend to emit alternatively a hole and an electron thus fluctuating between neutral and charged states. Fluctuating charges in the junction are indistinguishable from the true input signal applied to the gate and, as a result, have to be represented by the voltage noise at the input of the preamplifier.

The total input voltage noise dv_n (Figure 2.15) is given by

$$dv_n^2 = \frac{4}{3\pi}\frac{k_bT}{g_m}\left(1 + \frac{\omega_c}{\omega}\right)d\omega \tag{2.2.15}$$

The equation indicates that dv_n is inversely proportional to the transconductance of the JFET. In this respect, the parallel connection of N transistors benefits from the noise reduction by a factor of $\frac{1}{\sqrt{N}}$.

The total current noise includes two components

$$di_n^2 = di_2^2 + di_3^2 \tag{2.2.16}$$

These are the shot noise due to the leakage current I_l, which we discussed in the previous section

$$\mathrm{d}i_2^2 = \frac{1}{2\pi} 2I_l e \mathrm{d}\omega \tag{2.2.17}$$

and the current noise induced from the channel through the gate–source capacitance:

$$\mathrm{d}i_3^2 = \frac{1}{2\pi} \frac{k_\mathrm{b} T \omega^2 C_\mathrm{gs}^2}{g_\mathrm{m}} \mathrm{d}\omega \tag{2.2.18}$$

Reference specification materials often quote a noise factor instead of the current and voltage noise. A noise factor is defined as

$$F = 1 + \frac{\text{noise-power-at-the-input-of-JFET}}{\text{noise-power-of-source-resistor-}R_\mathrm{g}} \tag{2.2.19}$$

Figure 2.16 illustrates the expression (2.2.19). According to it, the noise factor can be represented in the following form

$$F = 1 + \frac{\mathrm{d}v_\mathrm{n}^2 + \mathrm{d}i_\mathrm{n}^2 R_\mathrm{g}^2}{4k_\mathrm{b} T R_\mathrm{g} B} \tag{2.2.20}$$

where B denotes the bandwidth of the system.

The noise figure is a logarithmic representation of the noise factor in dB.

$$NF = 10 \log F = 10 \left(1 + \frac{\mathrm{d}v_\mathrm{n}^2 + \mathrm{d}i_\mathrm{n}^2 R_\mathrm{g}^2}{4k_\mathrm{b} T R_\mathrm{g} B} \right) \mathrm{dB} \tag{2.2.21}$$

There is a general selection criterion of the front-end JFET, which is used specifically for low noise read-out of quasiparticle detectors. It was

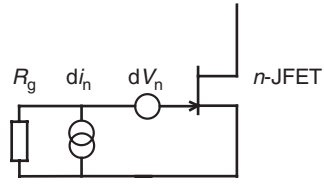


Figure 2.16 An equivalent noise diagram of the JFET that helps to calculate the noise factor widely quoted in technical specifications and the specialised literature

proposed originally for ionisation semiconductor detectors in [15]. Best suitable transistors must have the lowest factor K_s defined as

$$K_s = C_T\sqrt{\frac{8}{3}\frac{k_b T}{g_m}} = \sqrt{\frac{8}{3}\frac{k_b T}{g_m}}C'_d\left(m^{\frac{1}{2}} + m^{-\frac{1}{2}}\right) \tag{2.2.22}$$

where C_T is the total capacitance at the preamplifier input, $C'_d = C_T - C_{gs}$. $m = \frac{C'_d}{C_i}$ is the mismatch factor between the detector capacitance and the input capacitance of the JFET in whatever stage configuration it is used. According to equation (2.2.22), the lowest possible K_s would require using a transistor with the highest transition frequency

$$f_T = \frac{g_m}{2\pi C_{gs}} \tag{2.2.23}$$

with closely matched detector–JFET capacitances, that is $m \to 1$, provided that equation (2.2.15) holds. The last condition is not necessarily met in practical short channel JFETs.

2.2.3 The JFET Cascode Stage

The cascode represents an amplifier stage consisting of a common source JFET immediately followed by a common base bipolar transistor. Cascodes match well with a current-biased STJ. It has the particular advantage when an input-JFET operates at a lower temperature (optimally $\sim$150 K) with some distance separating it from following stages. A common base bipolar transistor has the low input emitter-base impedance. It practically eliminates the pole formed by the Miller effect $R_I C_{gd}(1 + g_m R_I)$ and reduces a loss in the amplification factor introduced by a parasitic capacitance of a cable connecting both parts of the stage.

The voltage across the source–drain terminals is set by biasing of the bipolar transistor, $V_{ds} \approx V_0 - 0.5$, whereas R_c adjusts the current flowing through the JFET. The value of R_c should be more than several hundred Ohms so that its presence did not affect the amplification factor.

Cascode circuits can be classified into two types depending on the polarity of the bipolar transistor they incorporate (see Figure 2.10 (a) with NPN type transistor, (b) with PNP type transistor). NPN bipolar

transistors are more advanced compared to their PNP counterparts in terms of the current gain β and the gain-bandwidth product f_T. These parameters play a crucial part in defining the noise and the bandwidth of the stage, as we will see later in this section. PNP transistors, however, are more convenient in the implementation. They allow operation at a smaller voltage range of the symmetric power supply $\pm V_{cc}$. With the quasizeroed output offset, the cascode can also be DC coupled to a following conventional operational amplifier, thereby simplifying the overall circuit design.

If the load impedance $|Z_l| = \left|R_l \frac{1}{1+j\omega R_l C}\right| \gg r_e$ (here r_e is the Shockley emitter resistance), the mean square current noise at the collector output generated by the bipolar transistor is given by [16]

$$i_{nc}^2 = \left(\frac{2\beta^2 r_e e I_e}{(\beta+1)^2 Z_l} + \frac{2el_e}{\beta}\right)\Delta f \tag{2.2.24}$$

On the other hand, the collector current noise originated from the JFET is as follows

$$i_f^2 = k_b T_c R_n g_m^2 = \frac{8}{3} k_b T_c g_m \tag{2.2.25}$$

where R_n is the noise resistance of the JFET channel. The noise contribution of the bipolar transistor becomes negligible when the condition $i_{nc}^2 \ll i_f^2$ is met. The condition can also be re-written in a more convenient form, as follows [16]

$$\frac{2}{Z_l^2 I_e} + \frac{3200 I_e}{\beta} \ll 116 g_m \tag{2.2.26}$$

where I_e is the current flowing through the emitter of the bipolar transistor. In the equation (2.2.26), we took into account that $r_e = \frac{0.025}{I_e}$ and $\beta \gg 1$.

The gain-bandwidth product of the JFETcascode stage is given by

$$\mathrm{GBW} = \frac{g_m}{2\pi(C_{in} + C_{out})} \tag{2.2.27}$$

with capacitances $C_{in} = C_{gs} + 2C_{dg}$ and $C_{out} = C_{dg}$. Expression (2.2.27) assumes $f_T \gg$ GBW.

2.2.4 Differential circuits based on JFET

A differential amplifier provides an output proportional to the difference between signals delivered to its input terminals. In this case, the non-inverting channel can be used to indirectly apply a voltage bias to the STJ detector through the feedback resistor, whereas the inverting input terminal accommodates a sum point of the conventional transconductance amplifier with a continuous reset.

Figures 2.11 and 2.12 show differential stages based on a common source pair of JFETs and two source followers respectively. Source followers are designed just to provide high input impedance. Their voltage gain given by

$$G_{SF} = \frac{g_m R_s}{1 + g_m R_s} \tag{2.2.28}$$

is less than the unity. Therefore, the following stage contributes all its voltage to the preamplifier input, as follows

$$dv_n^2 = dv_{nFET}^2 + \frac{dv_{n2}^2}{G_{SF}^2} R_s^2 di_{n2}^2 \tag{2.2.29}$$

where dv_{nFET} and dv_{n2} represent the voltage noise of the JFET and of the following stage respectively, and di_{n2} is the current noise of the second stage.

The output impedance of the source follower is known to be quite low

$$R_{out} = \frac{R_s}{1 + g_m R_s} \tag{2.2.30}$$

This enables one to match the differential stage and a fast low-noise bipolar operational amplifiers with the major pole shifted to a very high frequency. There is a number of fast low-noise operational amplifiers commercially available. For instance, a 120 MHz amplifier manufactured by Analog Device, AD829, contributes as low as $2\,\frac{nV}{\sqrt{Hz}}$ voltage noise referred to the input of the JFET stage. The voltage noise can be halved with 110 MHz AD797.

Common drain FET stages do not have the Miller effect. Their effective input and output capacitances are given by

$$C_{in}^* = C_{gs}(1 - G_{SF}) \tag{2.2.31}$$

$$C_{out}^* = C_{gs} + C_{sd} \tag{2.2.32}$$

Therefore, the principal pole

$$f_{\text{out}} \approx \frac{1}{2\pi C_{\text{out}} R_{\text{out}}} \tag{2.2.33}$$

can be well above 100 MHz. This circuit is normally utilised in preamplifiers required to provide extremely short settling time.

An amplification factor of the differential stage based on a pair of common source JFETs is given by

$$G_{\text{DS}} \approx 2 g_{\text{m}} R_{\text{l}} \tag{2.2.34}$$

for differential signal and

$$G_{\text{CM}} \approx \frac{R_{\text{l}}}{2R_{\text{I}}} \tag{2.2.35}$$

for a common mode input signal. Here R_{l} is the load resistor (Figure 2.11) and R_{I} is the impedance of the current source connected to the joined JFET sources.

The biggest concern when developing a preamplifier comprising either of the differential stages associates with their offset and particularly with the temperature drift. The temperature drift is defined as a change in the input voltage V_{gs} required to maintain a constant drain current I_{d}. Its mechanism originates from the reduction of the depleted volume in the channel, as temperature increases. This gives rise to the channel conductivity, the pinch-off voltage, V_{p}, and, eventually, the drain current I_{d}. The resulting effect of this mechanism on the symmetric output voltage can be significantly reduced in a well-matched pair of JFETs. Still commercially available general-purpose devices do not meet the stringent requirement with respect to the voltage bias stability placed by the STJ. In fact, the latter should not exceed 0.1 μV when dealing with measurement of the low energy single optical photons [17].

A large charge output of the STJ and a comparatively modest count rate make some room for the FET designer to further improve the temperature stability. For instance, the pinch-off voltage and the mobility have opposite temperature coefficients. The coefficients can be matched at a particular channel length and thickness so that I_{d} becomes temperature independent in the small signal limit. Moreover, JFET can be designed in such a way that the critical thickness is achieved at $V_{\text{gs}} = 0$. In this case, the cut-off voltage is expected to be $V_{\text{gs(off)}} = -0.63$ V.

Alternatively, the biasing circuitry of a JFET can be adjusted in order to reach the zero temperature drift operating point given by

$$I_{dz} = I_{d0}\left(\frac{0.63\ V}{V_{gs(off)}}\right)^2 \tag{2.2.36}$$

Operating the JFET in other operating points will result in a drift

$$\Delta V_{gs} \approx 2.2\,\text{mV}/^{\circ}\text{C}\left[1 - \left(\frac{I_d}{I_{dz}}\right)^{\frac{1}{2}}\right] \tag{2.2.37}$$

The frequency response of a differential stage based on JFETs with joined sources is defined by its input and output capacitances

$$C_{in} = \left(1 - \frac{g_m R_l}{1 + g_m R_l}\right) C_{gs} + (1 + g_m R_l) C_{gd} \tag{2.2.38}$$

$$C_{out} = C_{gd} \tag{2.2.39}$$

The major pole positions are given by

$$f_{01} = \frac{R_l + R_{in}}{2\pi R_l R_{in} C_{out}} \tag{2.2.40}$$

where R_{in} is the input impedance of the following stage, and

$$f_{02} = \frac{R_l + R_{in}}{2\pi (C_{in} + C_p) R_l R_{in}} \tag{2.2.41}$$

where C_p is an input capacitance of the following stage.

Figure 2.17 shows the noise model of the differential amplifier. An equivalent input noise is the sum in quadrature of fluctuations generated at both halves of the transistor pair.

$$dv_{ndif}^2 = dv_{n1}^2 + dv_{n2}^2 + di_{n1}^2 R_1^2 + di_{n2}^2 R_2^2 \tag{2.2.42}$$

The voltage bias is normally applied to the non-inverting input via a small value resistor R_2. In this case, the component $di_{n2}R_2$ can be eliminated from (2.2.42), but dv_{n2} must still be kept because its noise mechanism originates from the current fluctuations generated by the noise resistance of the JFET channel.

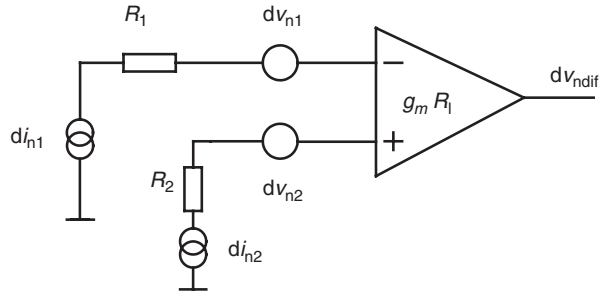


Figure 2.17 A noise model of the differential amplifier. An equivalent input noise is the sum in quadrature of fluctuations generated at both halves of the transistor pair

2.3 SQUID CURRENT AMPLIFIER

2.3.1 DC SQUID as a Superconductor Parametric Amplifier

The DC SQUID can be regarded as a self-pumping type of the parametric amplifier. Similarly to all parametric amplifiers, it also incorporates a non-linear component represented by the Josephson junctions and a resonator formed by the inductance of the superconductor loop L_s and the capacitance of the tunnel junction C. Its equivalent circuit diagram is shown schematically in Figure 2.18.

The tunnel junctions (denoted JJ in the Figure 2.18) operate in a cyclic mode being an AC current generator during one-half of the cycle and a

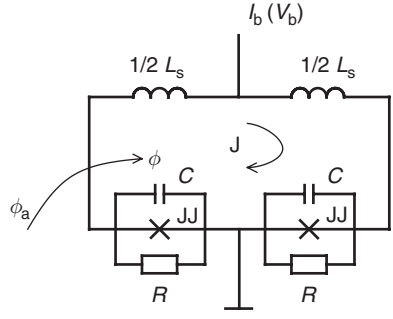


Figure 2.18 An equivalent circuit diagram of the DC SQUID. The device comprises superconducting inductive resonator, L_s, a couple of Josephson junctions, JJs and shunt resistors R to suppress the hysteresis in the junction's I–V curve. The SQUID is DC biased either with a current I_b or a voltage, V_b and responds to an external magnetic flux coupled to the inductance L_s

detector of oscillations during the other half. In doing so, they utilise both:

– the stationary Josephson effect for a supercurrent flowing through the junction given by

$$I = I_c \sin(\varphi) \tag{2.3.1}$$

– and the non-stationary Josephson effect for a voltage across the junction

$$V = \frac{\hbar}{2e}\frac{d\varphi}{dt} \tag{2.3.2}$$

where I_c is the critical current, φ is the quantum mechanical phase difference of the order parameter. The phase difference is directly coupled to the magnetic flux threading the SQUID loop as

$$\varphi = 2\pi\frac{\Phi}{\Phi_0} \tag{2.3.3}$$

with $\Phi_0 = 2.07 \times 10^{-15}$ Wb called the magnetic flux quantum.

An applied magnetic flux Φ_a primarily changes the state of the each junction φ. It affects the amplitude of the circulating current J and with that the average frequency of oscillations $\frac{d\varphi}{dt}$. According to equation (2.3.2), this must modulate the mean value of the voltage across the junctions. The block diagram, shown in Figure 2.19, summarises the principle of the SQUID parametric amplifier.

Due to the $\sin\varphi$ term in the stationary Josephson effect, the voltage response of the SQUID is periodic with respect to the magnetic flux. The

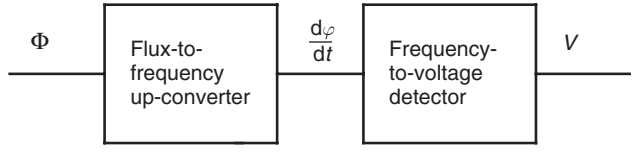


Figure 2.19 A block diagram explaining the principle of the DC SQUID as a parametric amplifier. The non-stationary Josephson effect acts as a pumping oscillator to up-convert the input magnetic flux Φ. L, C, R resonator sets up plasma frequency oscillations of the loop current J and forms a selective receiver. The mixed plasma frequency signal is then demodulated by non-linearity of the Josephson junctions. The DC SQUID is a self-pumping parametric amplifier

periodicity is defined with a very high precision via the ratio of two fundamental atomic constants $\Phi_0 = \frac{h}{2e}$ (h is the planck constant, e is the electron charge, Φ~2.07×10^{-15} Wb), called the flux quantum. This feature upgrades the SQUID from a low noise amplifier to a level of a self-calibrated measurement device. The flux quantum periodicity helps to evaluate the intrinsic energy resolution of the preamplifier explicitly without a need in the inter-comparison.

A system of two coupled motion equations describes the microscopics of the SQUID [18]:

$$\begin{aligned} \beta_c \frac{d^2 v}{dt^2} + \frac{dv}{dt} + \frac{\partial U}{\partial v} &= L_v(t) \\ \beta_c \frac{d\varphi}{dt^2} + \frac{d\varphi}{dt} + \frac{\partial U}{\partial \varphi} &= L_\varphi(t) \end{aligned} \tag{2.3.4}$$

with the potential energy of the force field governing the motion given by

$$U(v,\varphi) = \frac{(\varphi - \varphi_a)^2}{\beta_l} - jv - \cos v\varphi \tag{2.3.5}$$

Other dimensionless parameters in (2.3.4) are defined as follows

– the McCumber parameter

$$\beta_c = \frac{2\pi R^2 C I_c}{\Phi_0} \tag{2.3.6}$$

– the reduced SQUID inductance

$$\beta_l = \frac{2\pi L_s I_c}{\Phi_0} \tag{2.3.7}$$

– the Josephson current

$$i = \frac{I}{2I_c}, \text{ and}$$

– an applied external magnetic flux in the normalised form

$$\varphi_a = 2\pi \frac{\Phi_a}{\Phi_0}$$

The Langevin functions L_v and L_φ represent the random fluctuations. Their autocorrelation functions are given by

$$\langle L_v(t)L_v(t+\tau)\rangle\langle L_\varphi(t)L_\varphi(t+\tau)\rangle = \Gamma\delta(t) \tag{2.3.8}$$

where

$$\Gamma = \frac{2\pi k_b T}{I_c \Phi_0} \tag{2.3.9}$$

is the thermal excitation energy expressed through the Josephson coupling energy $E_J = \frac{I_c\Phi_0}{2\pi}$ and $\delta(t)$ is the Dirac delta function.

Turning to the mechanical analogy, the coupled equations (2.3.4) describe essentially the motion of a point mass in a two-dimensional force field of the periodic potential $U(v, \varphi)$. Their full analytical solution does not exist. There are a number of papers, however, presenting partial numerical solutions. All these solutions are in a very good agreement with the generic prediction made by Manley and Rowe [19]. The energy resolution of any parametric amplifier, including the SQUID, is given by

$$\varepsilon = \gamma \frac{k_b T}{\omega_p} \tag{2.3.10}$$

where ω_p is the pumping up-conversion frequency (in the SQUID case, the plasma frequency of the Josephson oscillations) and γ denotes the proportionality factor. The minimum proportionality factor essentially indicates how much thermal power k_bT is spent on the excitations of oscillations at frequencies other than ω_p. The energy losses are always present in parametric amplifiers. The oscillation linewidth broadens around ω_p by the nonlinear parametric elements. They also create a number of harmonic resonant modes in the resonator. The proportionality factor can further be increased if no measures are taken to eliminate the down-conversion of mixed parasitic modes into the measurement frequency range of the preamplifier.

Due to the large specific capacitance, the tunnel junctions normally employed in SQUIDs exhibit a hysteresis in their current-to-voltage characteristics (Figure 1.10). The hysteresis also absorbs the energy of oscillations proportional to its area. This would further deteriorate the energy resolution or even destroy the SQUID functionality. In order to avoid this the Josephson junction is connected in parallel with a shunt resistor, R. One chooses such a value of the shunt resistor that the modified I–V curve becomes non-hysteric. The upper value of R fully suppressing the hysteresis comes from the resistively shunted junction model (RSJ). According to it, the McCumber parameter β_c should not exceed a value of 0.7.

The SQUID can be operated in two modes. The first one is called the overdamped mode with limiting conditions $\beta_c \ll 0.7$ and $R \ll \frac{1}{\omega_p C}$. In this case, the capacitance does not influence the functionality of the SQUID. The relaxation–oscillation type of the resonator formed by R, L_s components with the plasma frequency resonance $\omega_p = \frac{R}{L_s}$ entirely defines the energy resolution as

$$\varepsilon_s = \gamma \frac{k_b T L_s}{R} \tag{2.3.11}$$

This mode of operation has proved to produce the best intrinsic energy resolution. It is not quite practical, however, due to a very low parametric amplification, which does not allow the SQUID to be coupled directly to the following semiconductor amplification stages.

The parametric amplification is given by

$$\frac{\partial V}{\mathrm{d}\Phi} \cong \frac{R_d}{\sqrt{2} L_s} \tag{2.3.12}$$

where R_d is the dynamic resistance of the SQUID. In the overdamped mode of operation, $R_d \approx \frac{R}{\sqrt{2}}$. With a typical value of $R = 1\,\Omega$ and $L = 100\,\mathrm{pH}$, the SQUID exhibits a parametric amplification

$$\eta = \frac{\partial V}{\partial \Phi} \approx \frac{R}{2L} \approx 10 \frac{\mu V}{\Phi_0}$$

Matching such a SQUID directly to a state-of-art preamplifier with a spectral density of the voltage noise $\sqrt{S_v} \approx 0.6 \frac{\mathrm{nV}}{\sqrt{\mathrm{Hz}}}$ results in an energy resolution $\varepsilon \approx \frac{S_v}{2L\eta^2} \approx 7.7 \times 10^{-29} \frac{\mathrm{J}}{\mathrm{Hz}}$, whereas an intrinsic energy resolution can be three orders of magnitude lower than that already at a temperature of 4.2 K.

Practical SQUIDs intended for fast applications operate in the critically damped mode with the McCumber parameter close to $\beta_c = 0.7$. In this case, the SQUID inductance L_s and the tunnel junction capacitance C form the resonator circuit with the plasma frequency resonance $\omega_p = \frac{1}{\sqrt{L_s C}}$. Following the equation (2.3.10), the energy resolution, therefore, can be written in the following form

$$\varepsilon = \gamma k_b T \sqrt{L_s C} \tag{2.3.13}$$

It is obvious that the LC-type of the plasma frequency resonance has potentially much higher quality factor compared to that that can be achieved with the relaxation–oscillation mode of operation. This has a positive impact on the dynamic resistance R_d as well as the parametric amplification η increasing them by a factor of 10 and higher [20].

A generic noise model of the SQUID applicable for the both modes of operation was developed in [21]. The model evaluates the voltage noise of the current-biased SQUID as

$$S_v = \frac{24k_bT}{R}\left[R_d^2 + 4\left(R_d - \frac{R}{2}\right)^2\right] \tag{2.3.14}$$

The flux noise can be found by substituting S_v derived with the equation (2.3.14) into

$$S_\phi = \frac{S_v}{\eta^2} \tag{2.3.15}$$

The intrinsic energy resolution of the SQUID is given by

$$\varepsilon = \frac{S_\Phi}{2L_s} \approx \frac{24k_bT\left[R_d^2 + 4\left(R_d - \frac{R}{2}\right)\right]^2}{RR_d^2}L_s \tag{2.3.16}$$

Analysis of equations (2.3.12) and (2.3.16) shows that the both most important parameters of the SQUID: the parametric amplification and the energy resolution benefit from the reduction of the SQUID inductance, L_s. In other words, the improved energy resolution can be preserved at the output of a read out circuit through the increase in the parametric amplification factor, η.

In fact the parametric amplification of SQUIDs with the minimised inductance L_s is large enough for implementation of the directly coupled semiconductor read-out electronics. Yet such a SQUID will require an intermediary transformer to match L_s to the inductance of the primary input coil. This configuration is discussed in the next section.

2.3.2 SQUID with an Intermediary Transformer

The concept of a SQUID with an intermediary transformer is shown in Figure 2.20. The transformer is formed by inductances L_2, L_3. Provided

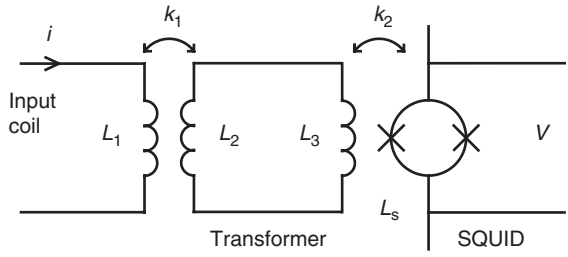


Figure 2.20 A schematic representation of the SQUID with an intermediary transformer. The intermediary transformer is formed by inductances L_2, L_3. At $L_2 \gg L_3$, the configuration practically does not introduce a loss in the magnetic flux and at the same time allows matching an arbitrarily large input inductance L_1 to a SQUID inductance L_s as small as 10 pH. The transfer function of a superconductor transformer does not attenuate down to 0 Hz

that coils L_2, L_3 are superconducting and coupling factors $k_1 = k_2 = 1$, the transformer couples the input magnetic flux $\Phi_a = iL_1$ to the SQUID with an attenuation factor given by

$$\zeta_a = \frac{L_2}{L_2 + L_3} \tag{2.3.17}$$

At $L_2 \gg L_3$, the configuration practically does not introduce a loss in the magnetic flux and at the same time allows matching an arbitrarily large input inductance L_1 to a SQUID inductance L_s as small as 10 pH.

A concept presented in Figure 2.20 was originally described in [22]. Although the potential superiority of the design has been recognised from the start, it was not in demand for some time. This can be explained by the fact that the management of transmission line resonances in the system of planar-coupled superconducting coils exposed to high-frequency Josephson oscillations presents quite a serious problem.

Figure 2.21 illustrates an equivalent circuit diagram of a double transformer SQUID without a bias, in which a virtual capacitance C_t is introduced to simulate a major $\lambda/4$ transmission line resonance of L_3. The calculations of the impedance across the a–b terminals show that the matched products L_3C_t and L_sC give rise to a single maximum resonance curve for the whole circuit. The coupled resonator shifts the plasma frequency from an intrinsic value $\omega_p/2\pi$ to a new value that can be evaluated with

$$f_{pt} \approx 2k_2 \frac{\omega_p}{2\pi} \tag{2.3.18}$$

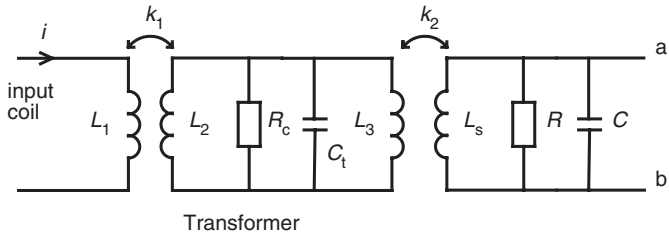


Figure 2.21 An equivalent circuit diagram of a double transformer SQUID without a bias circuitry. A virtual capacitance C_t is introduced to simulate a major $\lambda/4$ transmission line resonance between L_3 and the SQUID washer. The calculations of the impedance across the a–b terminals show that the matched products L_3C_t and L_sC give rise to a single maximum resonance curve for the whole circuit. In practice, there is always some mismatch between the products L_sC and L_3C_t. In this case, the role of the damping resistor R_c becomes crucial in bringing the plasma frequency resonance and the transmission line resonance together

This modified plasma frequency resonance essentially defines a maximum swing of the periodic flux-to-voltage characteristics given by

$$\Delta V \approx \frac{f_{pt}}{J} \tag{2.3.19}$$

where $J = \frac{\hbar}{2e} \approx 484 \frac{\text{MHz}}{\mu\text{V}}$ is the fundamental Josephson frequency-to-voltage ratio. An example of the flux-to-voltage characteristics is shown schematically in Figure 2.22.

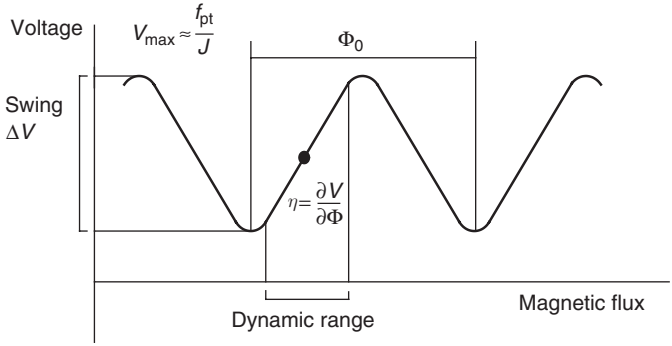


Figure 2.22 The periodic voltage-to-magnetic flux characteristics of the DC SQUID. The periodicity interval is equal to the flux quantum. In order to get a linear response, the operating point is kept within quasilinear dynamic range on a single branch of the characteristic. $\eta = \frac{\partial V}{\partial \Phi}$ represents the parametric amplification of the device

Thus, the double transformer configuration with the plasma frequency resonance matched to the major transmission line resonance delivers three most important benefits:

1. elimination of the adverse effects of the major transmission line resonance itself;
2. an increase in the parametric amplification η; and
3. a freedom to select any ratio between the inductance of the input coil and the SQUID inductance without affecting the energy resolution of the device.

A product $0.9I_cR$ limits the maximum value of the voltage swing V_m. In the optimally designed SQUID, values of I_c and R are derived from the McCumber parameter $\beta_c = 0.7$ and the dimensionless SQUID inductance $l = \frac{2L_sI_c}{\Phi_0} \approx 1$.

In practice, there is always some mismatch between the products L_sC and L_3C_t. In this case, the role of the damping resistor R_c becomes crucial in bringing the plasma frequency resonance and the transmission line resonance together. A reduction in R_c increases both the frequency f_{pt} and the quality factor of oscillations, which further enhances the parametric amplification. At some stage, however, a further increase in η starts reducing the quasilinear dynamic range of the SQUID. This is due to the fact that the magnitude of the down-converted Josephson oscillations cannot exceed the maximum voltage swing $V_m = 0.9I_cR$. Therfore, the V–Φ characteristics takes increasingly trapezoidal shape illustrated in Figure 2.23. Remarkably, R_c can be used for the fine tuning

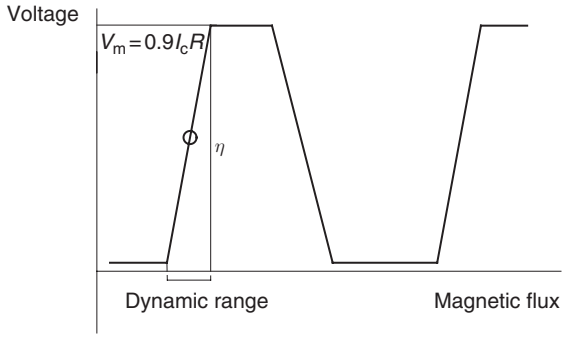


Figure 2.23 The V–Φ characteristics of the DC SQUID. The maximum swing of the characteristics is limited by a value $0.9\ I_cR$, where I_c is the critical current of the Josephson junction and R is the shunt resistor. A further increase in the parametric amplification reduces the linear dynamic range of the sensor

of the parametric amplification, η, as the hysteresis does not appear on the V–Φ curve at any value of it down to zero Ohm.

The damping resistor generates a thermal current noise, which adds an extra flux noise to the intrinsic flux noise of the SQUID, S_Φ. In order to reduce the contribution a minimum value of the damping resistor should be chosen in the range defined as

$$R_c > \frac{S_\Phi}{4k_b T\left(\frac{L_2 L_3}{L_2+L_3}\right)} \tag{2.3.20}$$

where S_Φ is the spectral density of the flux noise of the SQUID without a damping resistor across the coils L_2, L_3.

Figures 2.24 and 2.25 show recorded I–V and V–Φ characteristics of a tightly coupled double transformer SQUID. No traces of parasitic resonance structures and distortions are present in the whole frequency range of the Josephson oscillations from 20 GHz (40 μV) to 67 GHz (140 μV).

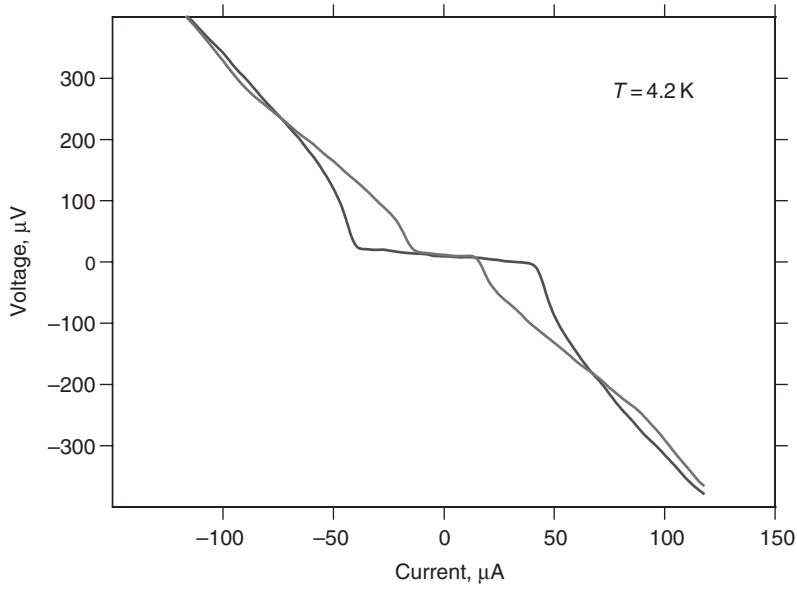


Figure 2.24 An experimental I–V curve of a SQUID with an intermediary transformer. It is very similar to an I–V curve of an autonomous SQUID without an input coil. This means that the system does not have any parasitic resonances

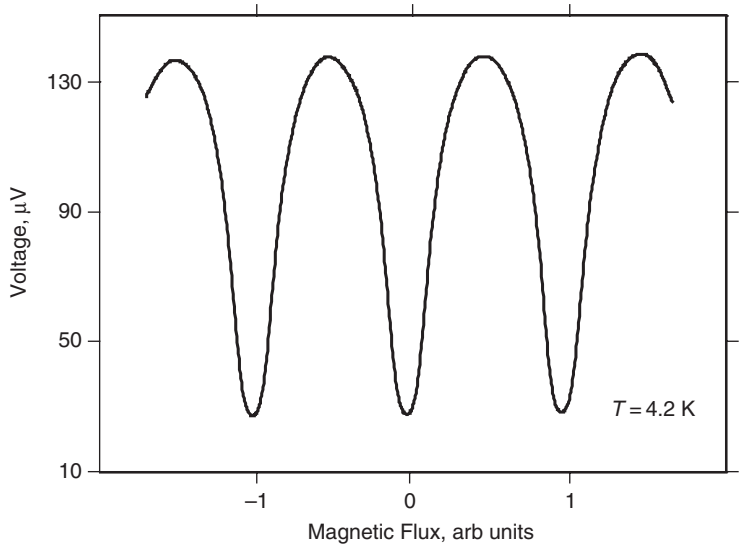


Figure 2.25 A recorded voltage-to-flux characteristics of a tightly coupled double transformer SQUID. No traces of parasitic resonance structures and distortions are present in the whole frequency range of the Josephson plasma oscillations from 20 GHz (40 μV) to 67 GHz (140 μV)

2.3.3 The Coupled Energy Resolution of a Double Transformer SQUID

The coupled energy resolution of any SQUID with an input coil attached is given by

$$\varepsilon_c = \frac{S_\Phi}{2k^2 L_s} \tag{2.3.21}$$

where k is the coupling factor between the input inductance L_1 and the SQUID. In the case of the double transformer configuration the overall coupling factor can be written as

$$k \approx k_1 k_2 \frac{L_2}{L_2 + L_3} \tag{2.3.22}$$

where the coupling factors k_1 and k_2 are explained in Figure 2.20.

Analytical calculations of coupling factors based on the device geometry are quite tedious. Here we will describe another approach to

tackle the issue experimentally. According to [23], the coupling factor can be defined as

$$k^2 = \frac{\varepsilon}{\varepsilon_c} \tag{2.3.23}$$

where ε is the intrinsic energy resolution of the SQUID evaluated with the equation (2.3.16) and ε_c is the coupled energy resolution, which can be written in the following form

$$\varepsilon_c = \frac{L_{1\text{eff}} S_i}{2} \tag{2.3.24}$$

In this expression, $L_{1\text{eff}}$ represents an effective input inductance L_1 and S_i is the spectral density of the input current noise. If a condition $L_2 \gg L_3$ is met, the intermediary transformer practically does not affect the low frequency SQUID inductance L_s. In this case, the spectral density of the current noise can be determined as follows

$$S_i \cong \frac{S_\Phi}{M^2} \tag{2.3.25}$$

where the mutual inductance between the input coil L_1 and the SQUID is given by

$$M = \frac{I_0}{\Phi_0} \tag{2.3.26}$$

with I_0 being the current flowing through the input inductance L_1 that produces in the SQUID a magnetic flux equal to Φ_0. Combining equations (2.3.24) and (2.3.25), we derive the sought coupling factor in the form

$$k = \sqrt{\frac{M^2}{L_{1\text{eff}} L_s}} \tag{2.3.27}$$

The effective value of the input inductance can be measured with the help of a resistor R_s attached across the input terminals of L_1. A value of R_s should be such that its thermal current noise is well above the current noise of the SQUID S_i. The effective input inductance is given by

$$L_{1\text{eff}} \approx \frac{R_s}{2\pi f_{3\text{dB}}} \tag{2.3.28}$$

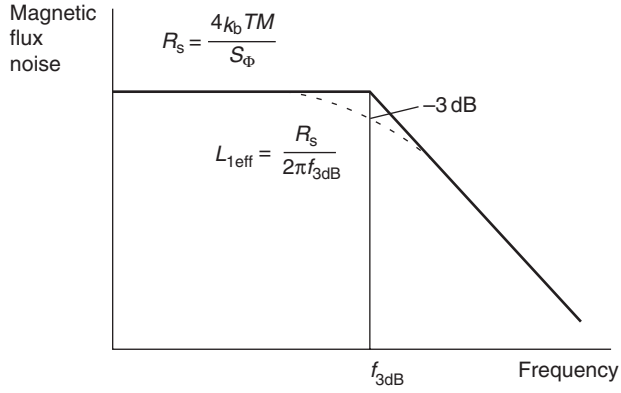


Figure 2.26 A spectral density of a SQUID amplifier with a small value resistor attached to its input terminal. The value of the resistor is chosen such that its Johnson noise prevails against the intrinsic current noise of the SQUID itself. The 3-dB roll-off frequency of the plot helps to determine an effective value of the input inductance of the amplifier. If a value of the noise resistor is not known beforehand, it can be recovered from the spectral density itself

with f_{3dB} being the 3-dB roll-off frequency of the spectral density of the flux noise, shown in Figure 2.26. The value of R_s itself can also be recovered from the same graph as

$$R_s = \frac{4k_bTM}{S_\Phi} \tag{2.3.29}$$

An example of a SQUID with an intermediary transformer was presented in [24]. Its main parameters are summarised in Table 2.1. *I–V* and *V–*Φ curves are shown in Figures 2.24 and 2.25 respectively. Figure 2.27 presents measured coupled energy resolution as a function of frequency.

Table 2.1 A summary of main parameters of a SQUID with an intermediary transformer

Shunt resistor, R	5.5 Ω
Critical current, I_c	21 μA
SQUID inductance, L_s	30 pH
Main parameter, l	0.61
McCumber parameter, β_c	0.7
Mutual inductance, M	2.5 nH
Coupling factor, k	0.77
Input inductance, L_{1eff}	367 nH
Inductance, L_2	1 nH
Inductance, L_3	0.1 nH
Parametric amplification, η	0.6 mV/Φ_0

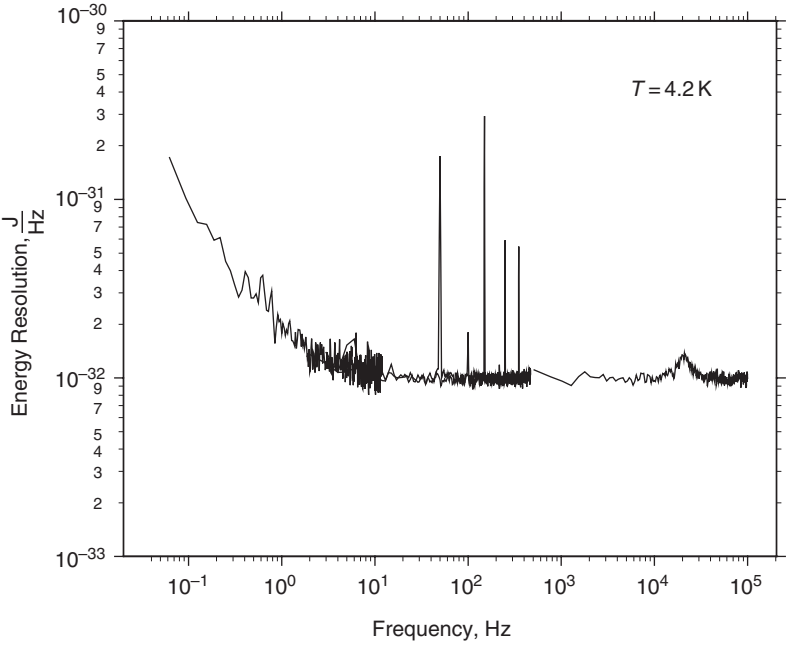


Figure 2.27 The coupled energy resolution of the DC SQUID with an intermediary transformer measured at a temperature of 4.2 K. Peaks correspond to mains frequency at 50 Hz and its harmonics. For comparison, the energy resolution of the best semiconductor transistor is as large as $10^{-26}\ \frac{\mathrm{J}}{\mathrm{Hz}}$

In the white noise region it corresponds to a value of $\varepsilon_c = 10^{-32}\ \frac{\mathrm{J}}{\mathrm{Hz}}$ (16 h) at a temperature of 4.2 K.

The spectral density of the current noise can be found from combining expressions (2.3.21) and (2.3.25) in the following form

$$S_i = 2k^2\varepsilon_c \frac{L_s}{M^2} \tag{2.3.30}$$

Data summarised in Table 2.1 yield a value of the current noise of approximately $0.25\frac{\mathrm{pA}}{\sqrt{\mathrm{Hz}}}$.

2.4 SQUID READ-OUT ELECTRONICS

The intrinsic energy resolution of SQUIDs is more than six orders of magnitude better than that of the best semiconductor transistors. It is, in fact, the most sensitive experimentation tool capable to detect the subtlest natural phenomena. Therefore, a particular care must be taken

at the design stage of a read-out electronics in order to preserve its resolution. The right choice and layout of components, impedance matching, grounding issues, cabling, screening, EMC filtering, all these and other essential skills of an electronic engineer undergo ultimate testing. Latest properly designed SQUID amplifiers function as robustly as any other conventional semiconductor counterparts. This enabled them to become a part of the currently most widespread calorimetric detectors and gradually win over users working with STJs.

A distinctive feature of the SQUID amplifier from other conventional transistor amplifiers is the fact that its transducer characteristics has multiple steady states. This means that an operating point will jump from one to another steady state if the feedback loop cannot follow the input signal and keep a non-compensated flux within the quasilinear dynamic range of the SQUID (see Figure 2.22). The problem is particularly relevant for detector applications. Here we have to deal with pulse signals having short rise times. Therefore, in addition to the low noise performance the read-out electronics must provide an adequate slew rate, which constitute a pair of contradictory requirements.

Fortunately, SQUIDs themselves are intrinsically very fast devices. With the plasma frequency resonance in the range of 20–60 GHz, their bandwidth can extend above 1 GHz. At present, the slew rate of practical current amplifiers is limited by the read-out electronics. It is imperative that we understand specific features of the SQUID system and all associated with them outstanding problems in order to take the full advantage of these extremely useful devices.

In this section, we will present some important information on the feedback loop design of systems with multiple equilibrium states. Conceptually, it is conveniently divided into two parts. First, we will start with the discussion of the small-signal limit, in which operating point stays within the dynamic range of the SQUID. After that, we will proceed with the large-signal limit, where we will study the dynamics of jumps of the operating point between different steady states (failure modes).

2.4.1 Small-Signal Model (dynamics)

There are two basic configurations, in which a SQUID can operate:

1. with a voltage bias and the current read-out [25]. Figure 2.28 represents a schematic diagram of this technique.
2. with a current bias and the voltage read-out [26]. This arrangement is illustrated in Figure 2.29.

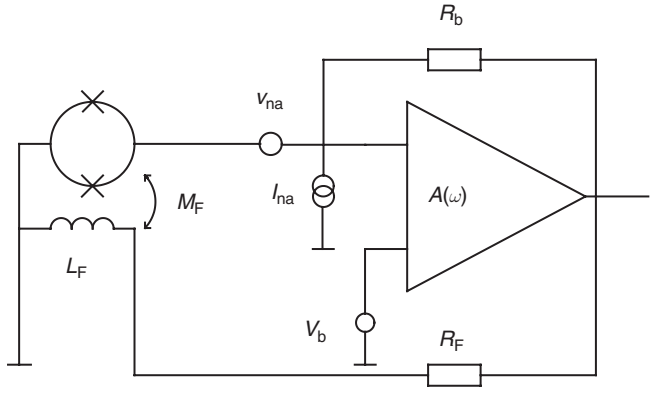


Figure 2.28 The basic configuration of the SQUID current amplifier with the voltage bias and the current read-out. The voltage bias is applied to the non-inverting terminal of the operational amplifier $A(\omega)$ and offsets the SQUID operating point via the feedback resistor R_b. Sources v_{na} and I_{na} represent the voltage and current noise of the read-out electronics. The amplifier works in the flux-locked mode with the loop formed by $A(\omega)$, R_F and L_F. M_F is the mutual inductance between the feedback coil and the SQUID

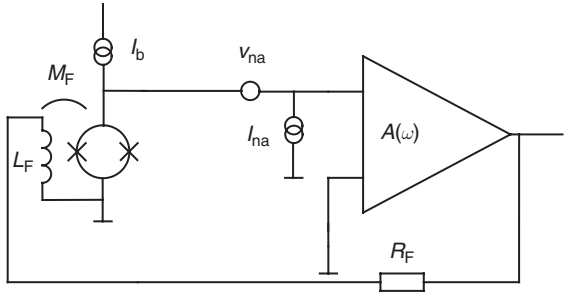


Figure 2.29 The basic configuration of the SQUID current amplifier with the current bias and the voltage readout. The current bias, I_b is applied directly to the SQUID and the inverting terminal of the operational amplifier $A(\omega)$ with carefully constructed frequency response. Sources v_{na} and I_{na} represent the voltage and current noise of the read-out electronics. The amplifier works in the flux-locked mode with the loop formed by $A(\omega)$, R_F and L_F. M_F is the mutual inductance between the feedback coil and the SQUID

In both the Figures, I_b and V_b denote the current and voltage bias respectively, and I_{na} and v_{na} are the current and voltage noise of the preamplifier respectively. $A(\omega)$ is the forward gain of the read-out circuit with a roll-off frequency slope adequate to provide the absolute stability of the feedback loop and a required slew rate. R_F and M_F are the

feedback resistor and the mutual inductance between the feedback coil L_F and the SQUID respectively.

Since the dynamic resistance of the SQUID is relatively low (normally it is in a range of 15–20 Ω), both the configurations produce similar performances in terms of the dynamics and the noise figures. In the following discussion, we will refer mostly to the SQUIDs with the current bias and the voltage read-out. Those, who are interested in the alternative configuration, can find its detailed analysis in [25].

An open loop gain of the circuit shown in Figure 2.29 is given by

$$G(\omega) = \eta \frac{M_F}{R_F} A(\omega) \tag{2.4.1}$$

A product of ηM_F represents an equivalent resistor R_{SQUID}, which can substitute the SQUID in the small-signal model. One needs to remember that such a substitution is only valid for analysing the signal dynamics of the system, not for noise calculations. The resistor can also be helpful in practice at the development stage and the circuit optimisation prior attaching it to a SQUID. This saves time, a cryogen and eliminates some uncertainties associated with cabling, distributed grounding. Figure 2.30 illustrates the arrangement.

If we assume a parametric amplification of the SQUID η to be constant, the SQUID current amplifier dynamics becomes essentially very

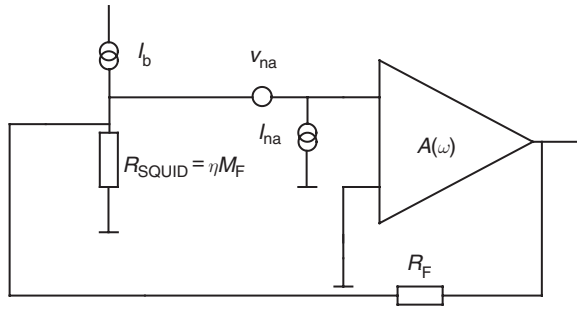


Figure 2.30 An equivalent circuit diagram of the SQUID current amplifier with the current bias and the voltage read-out. A product of the flux-to-voltage transfer function and the mutual inductance between the feedback coil and the SQUID ηM_F represents an equivalent resistor R_{SQUID}, which can substitute the SQUID in the small-signal model. One needs to remember that such a substitution is only valid for analysing the signal dynamics of the system, not for noise calculations. The resistor can also be helpful in practice at the development stage and the circuit optimisation prior attaching it to a SQUID. This saves time, a cryogen and eliminates some uncertainties associated with cabling, distributed grounding

similar to conventional linear amplifiers. In this case, all expressions given in the section 2.1.1 obviously remain relevant also to SQUID amplifiers used in the combination with STJ detectors. Some minor modification must be introduced into figures, like, for instance, for the settling time, which takes the following form

$$t_s^{\min} \approx \frac{R_F}{\eta M_F} \frac{1.2}{f_m} \tag{2.4.2}$$

For somewhat slower bolometric applications, including TES micro-calorimeters, the settling time does not need to be ultimately minimised. Instead, the system must provide the largest possible loop gain in the whole frequency range of operation. The stability of the base line as well as the reduction in the non-compensation error of the preamplifier, thereby, minimising its back reaction on the TES, establish top priorities in this type of detectors. These requirements are achieved by introducing a deep feedback, which borrows quite a lot from the open loop gain, and utilising an a-static integration stage that zeros the DC non-compensation error in the SQUID. Figure 2.31 exhibits a schematic diagram of a SQUID amplifier incorporating a single-pole integrator as an inertial element. Its equivalent block diagram with transfer functions expressed in the frequency domain using the Laplace transform is shown in Figure 2.32. In this figure, $\varphi_e = 2\pi \frac{\Phi_e}{\Phi_0}$, $\varphi_F = 2\pi \frac{\Phi_F}{\Phi_0}$ and $\varphi = 2\pi \frac{\Phi}{\Phi_0}$ represent normalised values of the external magnetic flux, the feedback magnetic flux and non-compensation error in the preamplifier respectively, and β is the transfer function of the voltage-to-flux feedback transducer.

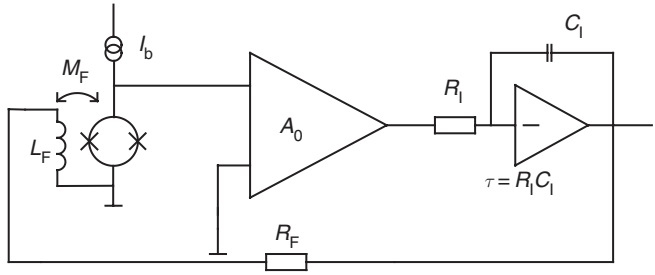


Figure 2.31 A schematic diagram of a SQUID current amplifier with $A(\omega)$ incorporating a single-pole integrator. The maximum slew rate of the system cannot exceed a value of $\pi f_m \Phi_0$, where f_m is the gain-bandwidth product of the loop

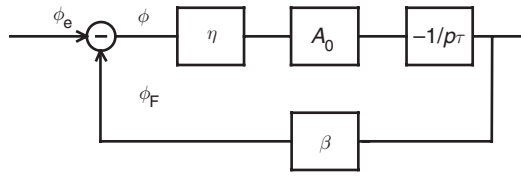


Figure 2.32 A functional block-diagram of the current amplifier with transfer functions expressed in the frequency domain using the Laplace transform. ϕ_e, ϕ_F are the normalised input and feedback signals, ϕ is non-compensated error, $A(\omega) = -A_0 \frac{1}{p\tau}$ is the transfer function of the direct amplification channel, β is the transfer function of the feedback. The SQUID and the feedback are assumed here to be non-inertial. This may not be true in very fast current amplifiers

Let us assume that numerical values of transfer functions of the loop are such that the gain-bandwidth product of the preamplifier amounts to $f_m = 10\,\text{MHz}$ with a closed loop gain

$$K_{loop} = \frac{R_F}{M_F \eta} = 100$$

where $R_F = 1\,\text{k}\Omega$ and $R_{SQUID} = M_F \eta = 10\,\Omega$.

The frequency response of this amplifier is shown in Figure 2.33. The 3-dB roll-off frequency is approximately 100 kHz.

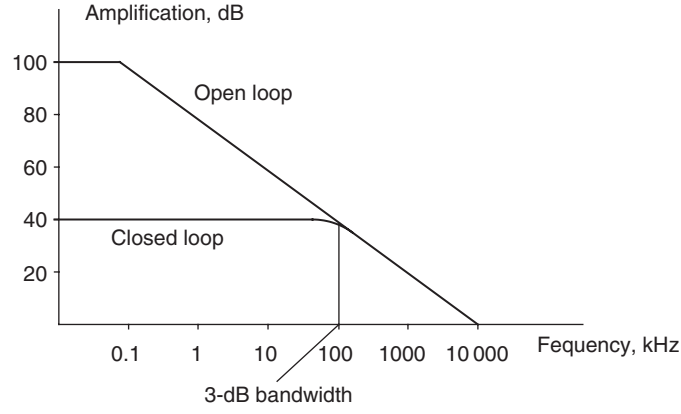


Figure 2.33 An example of the frequency response of an ideal single-pole SQUID current amplifier. A gain-bandwidth product is assumed to be approximately 10 MHz. The system is absolutely stable and produces a minimum settling time for the step-like input signal. The overshots or undershots may appear in real amplifiers due to the presence of parasitic inertialities only

The frequency response helps us to find a maximum input magnetic flux, which the system can tolerate, as a function of the frequency. In general, it is defined as

$$\Phi_{\max} = \Delta\Phi |G(\omega)| \tag{2.4.3}$$

where $\Delta\Phi$ was introduced in the previous section as the dynamic range of the SQUID. Figure 2.33 suggests that for a system with a single-pole integrator

$$|G(j\omega)| = \frac{\omega_{\mathrm{m}}}{\omega} \tag{2.4.4}$$

where $\omega_{\mathrm{m}} = 2\pi f_{\mathrm{m}}$ is the unity gain frequency of the preamplifier (gain-bandwidth product).

Substituting (2.4.4) into (2.4.3) and differentiating the result, we obtain another very important characteristics of the single-pole SQUID amplifier, which is called a slew rate:

$$\alpha = \omega_{\mathrm{m}} \Delta\Phi \tag{2.4.5}$$

The maximum slew rate cannot exceed a value of $\pi f_{\mathrm{m}} \Phi_0$.

Fast cryogenic detectors and, in particular, applications involving multiplexors require maximisation of the slew rate. Increasing the bandwidth has its limits when dealing with SQUIDs. Josephson junctions may pick up the high-frequency noise from the amplification loop. This distorts the I–V and V–Φ characteristics, smears the flux modulation and degrades the current resolution of the preamplifier as a whole. Therefore, before extending the unity gain frequency, it is advisable to maximise the loop amplification $G(\omega)$ within an existing bandwidth f_{m} by employing a multipole inertial stage.

Already a two-pole integrator provides a substantial improvement in the gain-bandwidth product and the slew rate at low frequencies. A circuit diagram and the functional block diagram of a SQUID amplifier with a two-pole integrator (to be more precise an integrator and a proportionally integrating circuit) are presented in Figs 2.34 and 2.35 respectively. The loop amplification for this configuration is given by

$$G(j\omega) = \frac{\eta M_{\mathrm{F}}}{R_{\mathrm{F}}} A_0 \frac{1}{j\omega\tau_1} \left(1 + \frac{1}{j\omega\tau_2}\right) \tag{2.4.6}$$

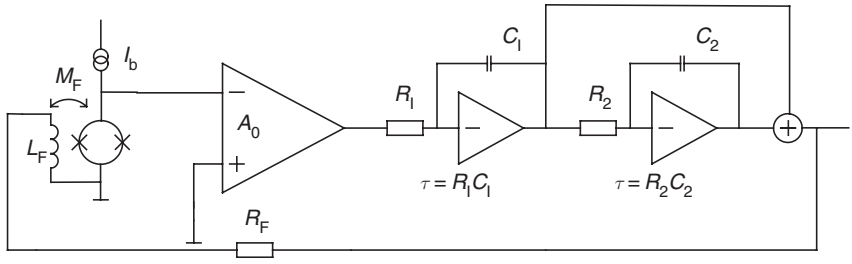


Figure 2.34 A schematic diagram of a SQUID current amplifier with $A(\omega)$ incorporating a two-pole integrator. The second pole is formed by the proportional integrating stage R_2C_2. The proportionality link makes the system absolutely stable, whereas the condition $\tau_2 = 4\tau_1$ minimises the settling time of the preamplifier for the step-like input signal

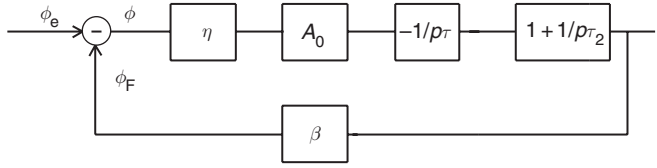


Figure 2.35 A functional block-diagram of the current amplifier with transfer functions expressed in the frequency domain using the Laplace transform. ϕ_e and ϕ_F are the normalised input and feedback signals, ϕ is non-compensated error, $A(\omega) = -A_0 \frac{1}{p\tau}(1 + \frac{1}{p\tau_2})$ is the transfer function of the direct amplification channel, β is the transfer function of the feedback

where $\tau_1 = R_1C_1$ and $\tau_2 = R_2C_2$ are time constants of the integrator and proportional-integrating circuit respectively.

In order to preserve the absolute stability of the system with the closed feedback the ratio between the time constants of the integrator and the proportional integrating stage must be

$$\tau_2 \geq 4\tau_1 \tag{2.4.7}$$

Figure 2.36 illustrates a critical case, when $\tau_2 = 4\tau_1$. In practice, the condition $\tau_2 > 4\tau_1$ must be met so that the system would have a $\frac{\pi}{6}$ margin in the total phase shift of the loop gain. This takes care of the inaccuracy of electronic components, the presence of parasitic inertialities and temperature effects including fluctuations.

The frequency dependence of the maximum input magnetic flux takes the following modified form

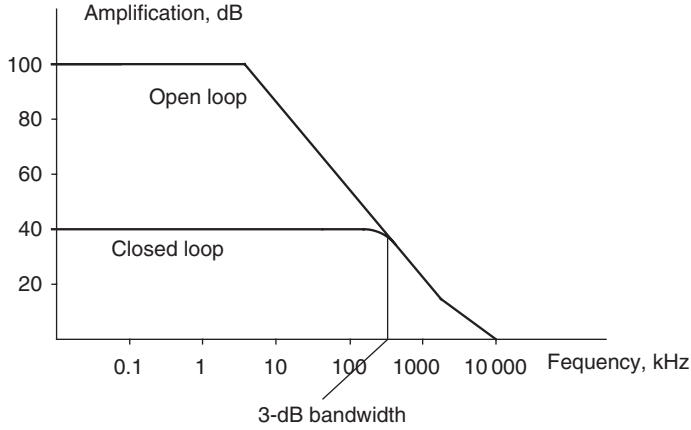


Figure 2.36 An example of the frequency response of an ideal two-pole SQUID current amplifier. A gain-bandwidth product is assumed to be approximately 10 MHz. The system is absolutely stable and produces a minimum settling time for the step-like input signal if a condition $\tau_2 = 4\tau_1$ is met. The overshots or undershots may appear in real amplifiers due to presence of parasitic inertialities

$$\Phi_{\max} = \Delta\Phi|G(j\omega)| = \Delta\Phi A_0 \frac{\sqrt{1+\omega^2\tau_2^2}}{\omega^2\tau_1\tau_2} \tag{2.4.8}$$

According to the equation (2.4.8), both the two pole and single-pole systems behave similarly at frequencies above $f_m/4$. Therefore, in this frequency range, the slew rate can be derived directly from the equation (2.4.5). When $f < f_m/4$, $\Phi_{\max}$ follows asymptotically a functional dependence given by

$$\Phi_{\max} \approx \frac{\omega_2\omega_m}{\omega^2}\Delta\Phi \tag{2.4.9}$$

The slew rate results from the differentiation of (2.4.9) in the following form

$$\alpha = \frac{\omega_1}{\omega}\omega_m\Delta\Phi \tag{2.4.10}$$

A further increase in the loop amplification $G(\omega)$ can be achieved by employing a more sophisticated n-pole inertial circuit with $n > 2$ [27]. Increasing a factor n, we gradually synthesise a frequency response approaching the Bode optimum. The Bode-type loop has an infinite number of poles and, therefore, cannot be realised with analogue components. For long time, it was of purely theoretical interest. The

digital signal processing, however, changes the perception of practicality. A digital filter is nothing but a series of mathematical operations on a digitised signal. In this respect, once a signal has been digitised, it can further be transformed into any function, which has some form of an explicit analytical or numerical description.

An idea to build a digital feedback loop in SQUID amplifiers has not delivered a competitive system so far. The spurious signals generated by digital components and the fast link between the analogue part of the circuit and a host computer affected adversely the overall energy resolution. An embedded design should help in the elimination of the cross-talk between digital and analogue grounds and preserve intrinsic performance of the amplifier. Having this future prospect in mind, in what follows, we will set up guidelines to designing a SQUID system with the maximised loop amplification. We will avoid here the reproduction of lengthy details of the Bode theory. They can be found in the original interpretation regarding linear amplifiers, in general, [28] and SQUID amplifiers, specifically [27]. Here we will only summarise practical recommendations and final expressions that help to synthesise the optimum system.

Figure 2.37 exhibits a Bode plot of an amplifier with desirable frequency dependence of the loop amplification. From the start, we will divide the whole frequency axis into four ranges of interest:

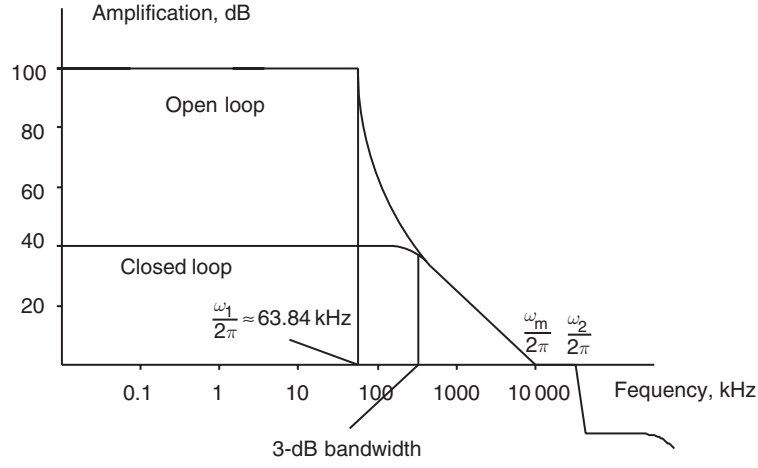


Figure 2.37 A frequency response of the SQUID current amplifier synthesised following Bode recommendations. A unity gain frequency was chosen 10 MHz to allow the direct comparison with single- and double-pole amplifiers. At a frequency of 100 kHz, the loop amplification of the Bode system is approximately 63 dB, as opposite to 40 dB and 55 dB provided by the single- and two-pole integrators

1. $0 < \omega < \omega_1$, from zero to ω_1 called the cut-off frequency;
2. $\omega_1 < \omega < \omega_m$, the interval of the fastest roll-off;
3. $\omega_m < \omega < \omega_2$, the correction interval; and
4. $\omega > \omega_2$, the high-frequency interval.

The second interval is crucial for maximisation of both: $G(\omega)$ and a slew rate, α. A roll-off slope of the loop gain

$$\delta L(\omega) = \frac{\mathrm{d}L(\omega)}{\mathrm{d}\omega} \tag{2.4.11}$$

should be as large as possible with one limiting condition – that the phase shift introduced by the interval stage at a frequency ω_m

$$\phi(\omega_m) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\partial L(\nu)}{\partial \nu} \ln \coth \left|\frac{\nu}{2}\right| \mathrm{d}\nu \tag{2.4.12}$$

should not exceed $-\pi$. In the equation (2.4.11), the loop gain is expressed in the logarithm form as $L(\omega) = 20\ \mathrm{Log}[G(\omega)]$.

In order to minimise the influence of the $\delta L(\omega)$ on $\phi(\omega_m)$ and preserve a steep roll-off of $\delta L(\omega_1 < \omega < \omega_m)$ Bode had found that the loop amplification must be zeroed in the correction interval, that is

$$\delta L(\omega_m < \omega < \omega_2) = 0 \tag{2.4.13}$$

A frequency response $\delta L(\omega_m < \omega < \omega_2) > 0$ is more beneficial from the point of view of the dynamics. But, in this case, the feedback throws an extra high-frequency noise back into the SQUID. If this noise affects the V–Φ characteristics δL should be reduced even below zero.

In the fourth high-frequency interval, the loop should provide the maximum possible filtration of high-frequency noise components.

Frequency characteristics of the amplifier with a feedback in all four ranges are mutually dependent. For instance, the roll-off slope at high frequencies, $\omega > \omega_m$, does influence the phase shift at $\omega < \omega_1$. To take the mutual dependence into consideration, we will introduce cross-correlated correction factors. In doing that, we represent the loop gain as a sum of two components

$$L(\omega) = L_1(\omega) + L_2(\omega) \tag{2.4.14}$$

where

$$L_1(\omega) = L(\omega) \quad \text{at} \quad \omega < \omega_m \tag{2.4.15}$$

$$L_1(\omega) = 0 \quad \text{at} \quad \omega > \omega_m$$

and

$$L_2(\omega) = L(\omega) \quad \text{at} \quad \omega > \omega_m$$

$$L_2(\omega) = 0 \quad \text{at} \quad \omega < \omega_m \tag{2.4.16}$$

In a similar way, we divide the phase shift into two parts:

$$\phi(\omega) = j[L_1(j\omega) - L_1(\omega)] \tag{2.4.17}$$

$$\phi(\omega) = j[L_2(j\omega) - L_2(\omega)] \tag{2.4.18}$$

Following the discussion in [29], we can obtain the phase shift below the cut-off frequency in the following form

$$\phi_1(\omega < \omega_1) = \frac{2\omega}{\pi}\sqrt{1 - \left(\frac{\omega}{\omega_1}\right)^2}\int_{\omega_1}^{\infty} \frac{\phi(\omega')}{\sqrt{\left(\frac{\omega'}{\omega_1}\right)^2 - 1[(\omega')^2 - \omega^2]}}\mathrm{d}\omega' \tag{2.4.19}$$

provided that functions $L(\omega < \omega_1)$ and $\phi(\omega > \omega_1)$ are frequency independent. The roll-off of the loop amplification in the second interval is given by

$$\Delta L_1(\omega) = L_1(\omega_1) - L_1(\omega) = \frac{17.4\omega}{\pi}\sqrt{\left(\frac{\omega}{\omega_1}\right)^2 - 1}$$

$$\int_{\omega_1}^{\infty} \frac{\phi(\omega')}{\sqrt{\left(\frac{\omega'}{\omega}\right)^2 - 1[(\omega')^2 - \omega^2]}}\mathrm{d}\omega' \tag{2.4.20}$$

Now, let us introduce a normalised pole factor

$$n = \frac{2\phi(\omega)}{\pi} = \frac{\delta L(\omega)}{20} \tag{2.4.21}$$

Substituting it into (2.4.19) and (2.4.20) and taking integrals, we obtain $\phi(\omega < \omega_1)$ and $\Delta L_1(\omega > \omega_1)$ in a simplified form

$$\phi(\omega < \omega_1) \approx n \arcsin\left(\frac{\omega}{\omega_1}\right) \tag{2.4.22}$$

$$\Delta L_1(\omega > \omega_1) \approx -8.7n \ln\left(\frac{\omega}{\omega_1} + \sqrt{\left(\frac{\omega}{\omega_1}\right)^2 - 1}\right) \tag{2.4.23}$$

From (2.4.22) and (2.4.23), the maximum cut-off frequency $\frac{\omega_1}{2\pi} \approx 63.84\,\text{kHz}$ with $f_\text{m} = 10\,\text{MHz}$ and $\Delta L_1(\omega_\text{m}) = -100\,\text{dB}$.

In order to flatten the open loop gain in the correction interval $[\omega_\text{m} < \omega < \omega_2]$, a system should incorporate a high-frequency filter with the semi-infinite characteristic amplification. Its transfer function is expected to rise at a rate of approximately 40 dB/dec starting from the unity gain frequency $\omega = \omega_\text{m}$. The phase shift produced by this stage in the whole frequency range can be written in a simplified form as follows

$$\phi_2(\omega) = \frac{2}{\pi} n \left[\frac{\omega}{\omega_\text{m}} + \frac{1}{9}\left(\frac{\omega}{\omega_\text{m}}\right)^3 + \frac{1}{25}\left(\frac{\omega}{\omega_\text{m}}\right)^5 + \cdots\right] \tag{2.4.24}$$

The width of the correction interval $[\omega_\text{m} < \omega < \omega_2]$ has effect on the maximum roll-off at frequencies $\omega > \omega_2$. The phase shift produced by $\delta L(\omega > \omega_2)$ can be found from the expression, which is similar to (2.4.22) only with the opposite sign and ω_2 instead of ω_1. For instance, at $\omega_2 = 2\omega_\text{m}$, the system remains stable with the drop in the loop amplification $\Delta L = 25\,\text{dB}$ and a roll-off of approximately 70 dB/dec.

The calculation of the frequency response will be complete if we take into account the influence of $L_2(\omega > \omega_\text{m)}$ on the phase shift $\phi_2(\omega < \omega_\text{m})$. It is given by [28]

$$\phi_2(\omega) = \frac{2}{8.7}\omega \int_{\omega_2}^{\infty} \frac{L_2(\omega)}{(\omega')^2 - \omega^2}\,\mathrm{d}\omega' \tag{2.4.25}$$

The fully synthesised frequency response of the Bode amplifier is shown in Figure 2.37. A unity gain frequency was chosen 10 MHz to allow the direct comparison with single- and double-pole amplifiers. At a frequency of 100 kHz, the loop amplification of the Bode system is approximately 63 dB, as opposite to 40 and 55 dB provided by the single- and two-pole integrators.

2.5 SQUID AMPLIFIER IN THE SMALL-SIGNAL LIMIT (NOISE)

In this section, we shall discuss noise properties of a single-stage SQUID current amplifier. The single stage configuration comprises a SQUID directly coupled to a low noise semiconductor operational preamplifier. Figure 2.38 illustrates an equivalent noise model of such a configuration. The SQUID itself is represented by the dynamic resistance, R_d, and a source of the intrinsic voltage noise, $S_{\nu s} = S_\Phi \eta^2$ attached in series with it. A following preamplifier includes the voltage and current noise sources, S_{va}, S_{ia}, connected to input terminals. The total voltage noise across the SQUID is given by

$$S_{\nu\Sigma} = S_\Phi \eta^2 + S_{va} + S_{ia} R_d^2 \tag{2.5.1}$$

where S_Φ is the intrinsic flux noise of the SQUID.

Thus, the voltage and current noise of the read-out preamplifier increase the total flux noise, $S_{\Phi\Sigma}$, referred to the input of the SQUID. $S_{\Phi\Sigma}$ can be found from the equation (2.5.1) divided by the transfer function of the SQUID as follows

$$S_{\Phi\Sigma} = S_\Phi + \frac{1}{\eta^2}(S_{va} + S_{ia} R_d^2) \tag{2.5.2}$$

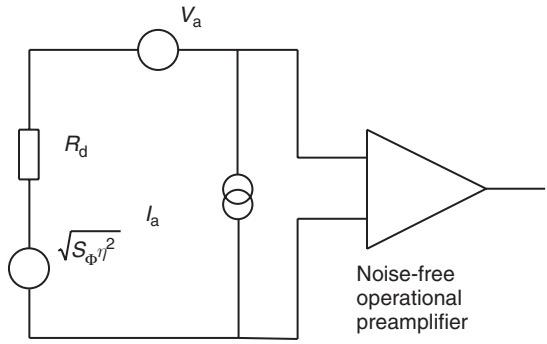


Figure 2.38 An equivalent circuit diagram of a single-stage SQUID directly connected to a differential amplifier. The SQUID itself is represented by the dynamic resistance, R_d, and a source of the intrinsic voltage noise, $S_{vs} = S_\Phi \eta^2$ attached in series with it. A following preamplifier includes the voltage and current noise sources, S_{va}, S_{ia}, connected to input terminals

If a value of the dynamic resistance is unknown, the equation (2.5.2) can be re-written in the following form

$$S_{\Phi\Sigma} = S_{\Phi} + \frac{S_{va}}{\eta^2} + 2S_{ia}L_s \tag{2.5.3}$$

One has to note that all three expressions assume the input impedance of the preamplifier to be much larger compared to the dynamic resistance of the SQUID R_d.

The total current noise of the SQUID current amplifier is derived from the equation (2.5.3) as

$$S_{i\Sigma} = \frac{S_{\Phi\Sigma}}{M^2} \tag{2.5.4}$$

with M being the mutual inductance between the input coil and the SQUID.

An example of the detailed practical circuit diagram of a single-stage SQUID current amplifier is shown in Figure 2.39. The preamplifier can be

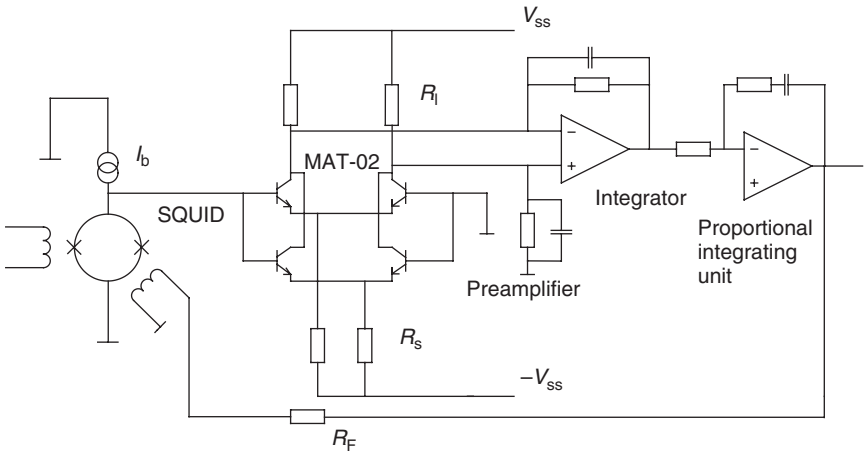


Figure 2.39 An example of the detailed practical circuit diagram of a single-stage SQUID current amplifier. The preamplifier incorporates an additional input differential stage built with an ultra low noise matched transistor pair MAT-02 from Analog Devices. Its voltage and current noise are as low as $0.8\frac{nV}{\sqrt{Hz}}$ and $1\frac{pA}{\sqrt{Hz}}$ respectively. The latter design is widely used in SQUID systems as it adds more flexibility in tailoring the performance of the preamplifier in terms of noise and the gain-bandwidth product. The circuit includes a two pole inertial stage, which is different from one shown in Figure 2.34, but has the same transfer function

based entirely on a commercial low noise operational amplifier, such as AD797 manufactured by Analog Device. Alternatively, it may incorporate an additional input differential stage built out of ultra low noise discrete-matched transistor pairs. The most popular pair of matched transistors is MAT-02 from Analog Devices. Its voltage and current noise are as low as $0.8\,\frac{\mathrm{nV}}{\sqrt{\mathrm{Hz}}}$ and $1\,\frac{\mathrm{pA}}{\sqrt{\mathrm{Hz}}}$ respectively. The latter design is more widely used in SQUID systems as it adds more flexibility in tailoring the performance of the preamplifier in terms of noise and the gain-bandwidth product.

According to [18], the expression (2.5.3) for the total flux noise can be written in terms of measurable parameters as follows

$$S_{\Phi\Sigma} \approx \frac{2k_{\mathrm{b}}TL_{\mathrm{s}}^2}{R}\left\{1+4\frac{R}{R_{\mathrm{d}}}\left[1+\frac{T_{\mathrm{a}}}{2T}\left(\frac{R_{\mathrm{opt}}}{R_{\mathrm{d}}}+\frac{R_{\mathrm{d}}}{R_{\mathrm{opt}}}\right)\right]\right\} \tag{2.5.5}$$

where T and $\mathrm{T_a}$ represent a temperature of the SQUID shunt resistor, R, and the differential stage of the preamplifier respectively and $R_{\mathrm{opt}} = \sqrt{\frac{S_{\mathrm{va}}}{S_{\mathrm{ia}}}}$ is the optimal source impedance for the preamplifier. The noise temperature is defined as $T_{\mathrm{a}} = \frac{(S_{\mathrm{va}}S_{\mathrm{ia}})^{0.5}}{2k_{\mathrm{b}}}$. Expression (2.5.5) sets an optimum matching condition between the SQUID and the preamplifier given by

$$R_{\mathrm{d}} = R_{\mathrm{opt}} \tag{2.5.6}$$

The differential gain of the differential stage can be found from the following equation

$$G_{\mathrm{dm}} \approx \frac{-2R_{\mathrm{l}}}{r_{\mathrm{e}}+\frac{R_{\mathrm{s}}}{\beta}} \tag{2.5.7}$$

where R_{l}, R_{s} and r_{e} denote the load resistance, the resistance attached to the transistor emitter and actual emitter resistance of the transistor respectively.

The common mode amplification is given by

$$G_{\mathrm{cm}} \approx \frac{-R_{\mathrm{l}}}{r_{\mathrm{e}}+\frac{R_{\mathrm{s}}}{\beta}+2R_{\mathrm{s}}} \tag{2.5.8}$$

Both expressions (2.5.7) and (2.5.8) assume that load resistors are perfectly matched. In order to increase the common mode rejection ratio, these resistors should be carefully selected.

The noise model of the differential stage is shown in Figure 2.40. The total voltage noise measured at the input terminal of the stage is given by

$$\mathrm{d}v_{\mathrm{n}}^2 = \mathrm{d}v_{\mathrm{n1}}^2 + \mathrm{d}v_{\mathrm{n2}}^2 + \mathrm{d}i_{\mathrm{n1}}^2 R_1^2 + \mathrm{d}i_{i2}^2 R_2^2 + \frac{2\mathrm{d}v_3^2}{G_{\mathrm{dm}}^2} + \frac{\mathrm{d}v_4^2 + \mathrm{d}v_5^2 + \mathrm{d}v_6^2}{G_{\mathrm{cm}}^2} \tag{2.5.9}$$

where $\mathrm{d}v_{\mathrm{n1}}$, $\mathrm{d}v_{\mathrm{n2}}$, $\mathrm{d}i_{\mathrm{n1}}$ and $\mathrm{d}i_{\mathrm{n2}}$ represent voltage and current noise of bipolar transistor pair; R_1 and R_2 are impedances of signal sources V_{signal} and V_{bias}; $\mathrm{d}v_3$ and $\mathrm{d}v_4$ are the noises generated by R_{l} and R_{s} respectively (Figure 2.39); and, finally, $\mathrm{d}v_5$ and $\mathrm{d}v_6$ are responsible for voltage fluctuations on the power supply lines.

Single channel systems normally employ a conventional four-wire technique to read out the SQUID, as illustrated in Figure 2.39. It is a simple reliable configuration which provides very good common mode rejection ratio without a need of any know how in the electronic design. In multichannel or/and ULT (UltraLow Temperature) systems, however, the three-wire technique becomes much more attractive. In these applications, every spare or omitted lead is a premium improving the reliability or/and reducing the thermoload on the ULT stage. D. Drung

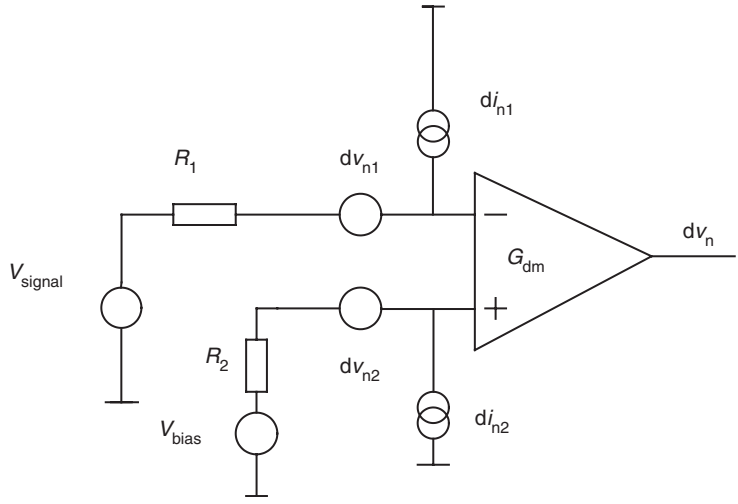


Figure 2.40 A noise model of the differential stage. $\mathrm{d}v_{\mathrm{n1}}$, $\mathrm{d}v_{\mathrm{n2}}$, $\mathrm{d}i_{\mathrm{n1}}$ and $\mathrm{d}i_{\mathrm{n2}}$ represent voltage and current noise of bipolar transistor pair, R_1 and R_2 are impedances of signal sources V_{signal} and V_{bias}, $\mathrm{d}v_3$ and $\mathrm{d}v_4$ is the noise generated by R_{l} and R_{s} respectively (Figure 2.39) and, finally, $\mathrm{d}v_5$ and $\mathrm{d}v_6$ are responsible for voltage fluctuations on the power supply lines

brought the three-wire technique to a level of perfection where it performs almost as well as the four-wire connection. Interested readers should refer to [30] for further details.

Example 2.1

Let us assume that a parametric amplification of the SQUID is $\eta = 0.6\,\mathrm{mV}/\Phi_0$, a SQUID inductance $L_s = 30\,\mathrm{pH}$. A spectral density of the voltage noise of the differential stage based on a couple of matched transistor pairs MAT-02 connected in parallel $\sqrt{\frac{S_{va(MAT-02)}}{2}} \approx 0.66\,\frac{nV}{\sqrt{Hz}}$ (compare it with the measurement data presented in Figure 2.41). From data sheet, we find a spectral density of the current noise equal to $2\sqrt{S_{ia(MAT-02)}} \approx 2\,\frac{pA}{\sqrt{Hz}}$. Thus, a dynamic resistance of the SQUID and an optimum resistance of the two-pair differential stage have the following values

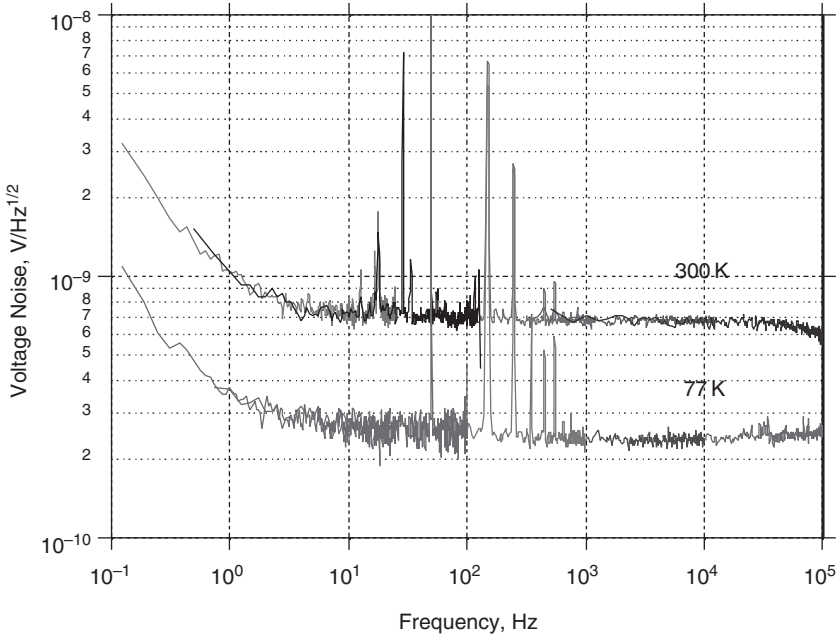


Figure 2.41 A spectral density of the voltage noise of a differential stage based on two parallel pair of MAT-02 (Analog Devices). The spectrum was measured at 300 and 77 K

$$R_{\mathrm{d}} \approx \frac{\eta}{\sqrt{2}L_{\mathrm{s}}} \approx 10\,\Omega \quad \text{and} \quad R_{\mathrm{opt}} \approx \sqrt{\frac{S_{\mathrm{va}}}{S_{\mathrm{ia}}}} \approx 330\,\Omega \text{ respectively.}$$

The noise temperature of the preamplifier with a SQUID placed at $T = 4.2\,\mathrm{K}$

$$T_{\mathrm{a}} = \frac{(S_{\mathrm{va}}S_{\mathrm{ia}})^{0.5}}{4k_{\mathrm{b}}} \approx 43\,K$$

In the case of the voltage amplification, that is $R_{\mathrm{opt}} \gg R_{\mathrm{d}}$, the expression (2.5.5) can be written down in a simplified form, as follows

$$S_{\Phi\Sigma} = \frac{2k_{\mathrm{b}}TL_{\mathrm{s}}^2}{R}\left[1 + \frac{R}{R_{\mathrm{d}}}\left(4 + 2\frac{R_{\mathrm{opt}}}{R_{\mathrm{d}}}\frac{T_{\mathrm{a}}}{T}\right)\right] \tag{2.5.10}$$

Using the equation (2.5.10), we derive a total spectral density of the flux noise equal to approximately $\sqrt{S_{\Phi\Sigma}} \approx 8 \times 10^{-7}\,\frac{\Phi_0}{\sqrt{\mathrm{Hz}}}$. This is in a good agreement with measurement results presented in Figure 2.42.

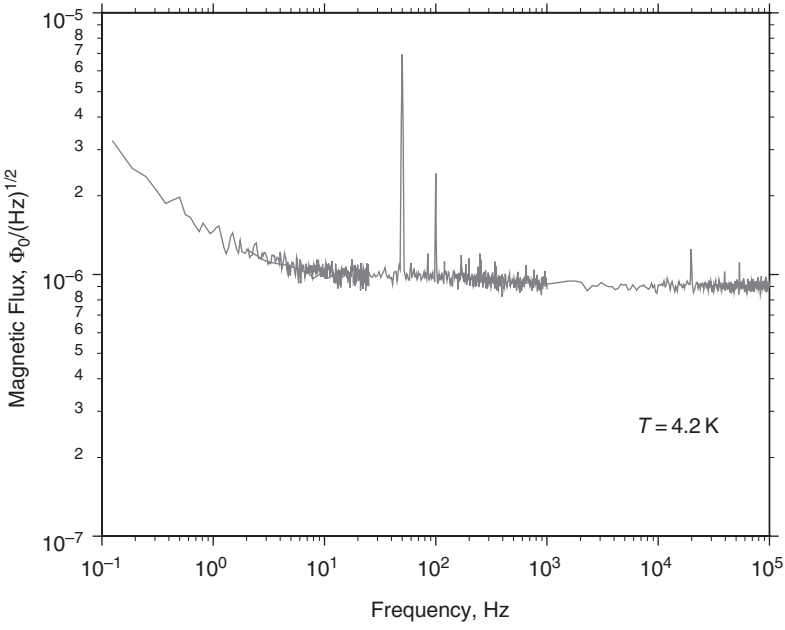


Figure 2.42 A spectral density of the flux noise of a single stage SQUID attached to a differential stage based on two parallel pair of MAT-02 (Analog Devices). The spectrum was measured for the differential stage operating at a room temperature

The noise contribution of the preamplifier can further be reduced to satisfy the optimum matching condition $R_d = R_{opt}$. In order to achieve that the input differential stage should include a larger number of transistor pairs coupled in parallel. The effective total voltage noise and the current noise of a differential stage incorporating N parallel bipolar transistor pairs are given by

$$dv_{na\Sigma} = \frac{dv_{na}}{\sqrt{N}} \tag{2.5.11}$$

$$di_{na\Sigma} = di_{na}\sqrt{N} \tag{2.5.12}$$

respectively. This leads to an optimum resistance given by

$$R_{opt\Sigma} = \frac{dv_{na\Sigma}}{di_{na\Sigma}} = \frac{R_{opt}}{N} \tag{2.5.13}$$

In practice, the optimum matching condition (2.5.7) is rarely met. A differential stage based on multiple transistor pairs in parallel affects adversely the dynamics of the SQUID current amplifier. First, it causes a signal loss due to lower input impedance R_I, which makes stronger the dividing effect between the input impedance and the dynamic resistance $\frac{Z_i}{Z_i+R_d}$. The divider is obviously frequency dependent and may form a new major pole involving the SQUID-dynamic resistance and the input capacitance of the differential stage. The input capacitance is dominated by the Miller effect. It increases in proportion with a number of transistor pairs N. The creation of a new major pole will incur some modifications in the design of inertial stages of the system loop to partially compensate for an additional roll-off of frequency characteristics. The partiality comes from a fact that the SQUID-dynamic resistance changes with moving operating point, which causes the instability of the pole position. Thus, one has to expect a some reduction of the slew rate and an increase in the settling time of the step response.

The low input impedance of multiple-pair differential stages also loads the SQUID reducing its parametric amplification, η. Such stages throw large current noise, which, starting from a certain number N, becomes a dominating factor in the signal-to-noise ratio of the system.

A more efficient way to reduce the noise temperature of the preamplifier and to preserve the fast frequency response at the same time is to operate the differential stage at a lower physical temperature. All liquid-free cryocoolers have an intermediary cooling stage with a temperature around 70 K and a cooling power of several tens of watts [31]. The spectral density

of the voltage noise drops in proportion with temperature. Figure 2.41 demonstrates the voltage noise of the same differential stage incorporating two transistor pairs in parallel at approximately 77 K. A spectral density of the voltage noise reduces to a value of $0.25\frac{\text{nV}}{\sqrt{\text{Hz}}}$, which leads to a noise temperature of the SQUID amplifier as low as 18 K.

2.6 SQUID CURRENT AMPLIFIER IN THE LARGE-SIGNAL LIMIT (DYNAMICS)

The large signal dynamics of a SQUID amplifier takes into account the periodicity of the flux-to-voltage transfer function. This is particularly important in conjunction with detectors producing current pulses with sharp front edges and relatively long decay time. If a slew rate of the preamplifier is insufficient to follow the input signal, a non-compensation error changes the positioning of the operating point on the V–Φ curve. In fact, the non-compensation error in excess of a value of $\frac{\Phi_0}{4}$ drives the operating point onto a branch of V–Φ curve with the opposite slope, thereby switching the polarity of the feedback to the positive sign. The positive feedback accelerates the operating point even further until the system has found another steady state or its output reached a power rail voltage.

The transition onto a new stable branch of V–Φ curve produces an offset voltage at the system output. The offset is quite large in real terms and proportional to $n\Phi_0$, where n is the number of transitions. The phenomenon is often called a failure in operation. It causes a particular concern in applications, where the output is DC coupled to a pulse processor unit. It either puts the signal out of the range of an ADC in the major processing channel or upsets a system monitoring the base temperature variations like in the case of TES calorimetric detectors with external electro-thermal feedback.

The main goal of the system analysis in the large signal limit is to define a measurable or calculable indicator that characterises the failure immunity of SQUID systems. Such an indicator can be, for example, a failure rate occurred over certain period time at predefined conditions. Predefined conditions are important as the failure rate depends on a type of application (shapes of the input signals), internal fluctuations circulating in the loop and external electromagnetic spikes superimposed on the signal pulses. In any case, a general criterion of the system usability is that the failure rate must be insignificantly small when compared to a rate of incoming events.

2.6.1 SQUID Amplifier with a Single-Pole Integrator

The following analysis is based on a block-diagram shown in Figure 2.43. It consists of a SQUID-preamplifier unit with the joined transfer function $F(\varphi)$, an integrator with a Laplace transfer function $K(p) = \frac{1}{p\tau}$ and a non-inertial linear feedback circuit denoted as β.

The flux-to-voltage transfer function of the SQUID-preamplifier unit can be written in the general form as a series

$$F(\varphi) = \sum_{n=1}^{\infty} U_{m,n} \sin\left(2\pi n \frac{\Phi}{\Phi_0}\right) = \sum_{n=1}^{\infty} U_{m,n} \sin(n\varphi) \tag{2.6.1}$$

where $\varphi = \varphi_e - \varphi_\beta$ is the normalised non-compensation magnetic flux error, φ_e and φ_β represent the input and feedback signals respectively, and $U_{m,n}$ is the peak-to-peak voltage modulation depth of the $F(\varphi)$ characteristics. In order to derive results in a simple analytical form, we assume that $n = 1$. This may slightly alter final numerical values of the analysis, but will not affect the generality of made conclusions. Expression (2.6.1) can now be written in a simplified form as

$$F(\varphi) \approx U_m \sin\varphi \tag{2.6.2}$$

The motion equation of the SQUID amplifier is given by [32]

$$\frac{d\varphi}{dt} + \omega_m \sin[\varphi(t)] = \frac{d\varphi_e}{dt} \tag{2.6.3}$$

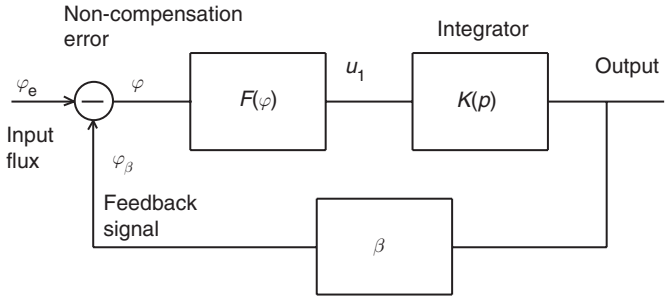


Figure 2.43 A block-diagram of a SQUID current amplifier in the flux-locked mode of operation. It consists of a SQUID-preamplifier unit with the joined transfer function $F(\varphi)$, an integrator with a Laplace transfer function $K(p) = \frac{1}{p\tau}$ and a non-inertial linear feedback circuit denoted as β

In the equation (2.6.3),

$$\frac{\omega_m}{2\pi} = \frac{\beta}{2\pi\tau} U_m \tag{2.6.4}$$

denotes a characteristic maximum frequency of the effective bandwidth of the SQUID amplifier with the feedback.

Equation (2.6.3) has an analytical solution. Making a substitution of

$$y = tg\left[\frac{\varphi(t)}{2} - \frac{\pi}{4}\right] \tag{2.6.5}$$

into (2.6.3) we derive a modified Riccarti differential equation in the following form

$$\dot{y} + \alpha y^2 + \varpi = 0 \tag{2.6.6}$$

where $\alpha = \frac{1}{2}[\omega_m + \frac{d\varphi_e(t)}{dt}]$ and $\bar{\omega} = \frac{1}{2}[\omega_m - \frac{d\varphi_e}{dt}]$.

In the range $0 < \frac{d\varphi_e(t)}{dt} < \omega_m$, the system is capable to follow the input signal φ_e. The non-compensation error derived from the equations (2.6.3) to (2.6.6) has the form

$$\varphi(t) = 2\text{arc}tg \frac{tg\left[\frac{\varphi(0)}{2} - \frac{\pi}{4}\right]\sqrt{\alpha\varpi} - \varpi th\sqrt{\alpha\varpi}t}{\sqrt{\alpha\varpi} - tg\left[\frac{\varphi(0)}{2} - \frac{\pi}{4}\right]\alpha th\sqrt{\alpha\varpi}t} + \frac{\pi}{4} \pm 2n\pi \tag{2.6.7}$$

Within the range under discussion, an input signal gradient $\frac{d\varphi_e(t)}{dt} = \text{const}$ settles the uncompensated error φ to a constant value φ_0 that can be evaluated from the equation (2.6.7). Figure 2.44 (curve 1) gives a qualitative illustration of the case.

On the phase plane, shown in Figure 2.45, the φ_0 constitutes a focal point of an "attraction" interval extending from $[-\pi - \arcsin(\frac{1}{\omega_m}\frac{d\varphi_e}{dt})]$ to $[\pi - \arcsin(\frac{1}{\omega_m}\frac{d\varphi_e}{dt})]$. The attraction interval means that the operating point within this interval has a chance to return to the focal position φ_0. If a pulse signal is applied on top of the existing $\frac{d\varphi}{dt}$ gradient offset and the non-compensated error exceeds the limits of the attraction interval, the system will make a definite transition into a new equilibrium state.

In the limit of $\frac{d\varphi_e}{dt} > \omega_m\dot{\varphi}(\varphi)$ curve on the phase plane (Figure 2.45) no longer crosses the φ axis. This means that the system does not have a

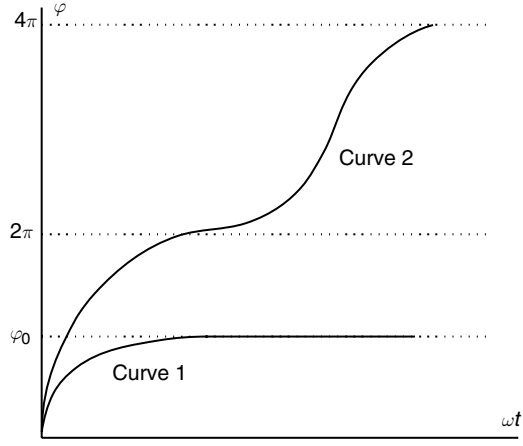


Figure 2.44 The non-compensation error in a SQUID current amplifier as a function normalised time. Curve 1 represents a case for the input signal changing in the range $0 < \frac{d\varphi_e(t)}{dt} < \omega_m$. An input signal gradient $\frac{d\varphi_e(t)}{dt} = \text{const}$ settles the uncompensated error φ to a constant value φ_0. Curve 2 illustrates a case of $\frac{d\varphi_e}{dt} > \omega_m$. The operating point continuously moves from one steady state to another. The non-compensation error is mounting until the system has reached a power saturation limit

steady-state mode of operation. The solution of equation (2.6.3) at these operating conditions takes the form

$$\varphi(t) = 2\text{arctg}\frac{tg\left[\frac{\varphi(0)}{2} - \frac{\pi}{4}\right]\sqrt{\alpha\varpi} - \varpi tg\sqrt{-\alpha\varpi}t}{\sqrt{-\alpha\varpi} - tg\left[\frac{\varphi(0)}{2} - \frac{\pi}{4}\right]\alpha tg\sqrt{-\alpha\varpi}t} + \frac{\pi}{4} \pm 2n\pi \qquad (2.6.8)$$

The curve 2 in Figure 2.44 illustrates this solution. The operating point continuously moves from one steady state to another. The non-compensation error is mounting until the system has reached a power saturation limit.

According to equations (2.6.7) and (2.6.8), it takes some time for a system to proceed to a new steady state. This finite transition time can be used to prevent failures in operation. Systems incorporating such a prevention circuitry are often called "adaptive systems". There are two ways to keep an operating point in the same steady state:

1. to generate a compensation signal [33]; and
2. to increase the maximum frequency ω_m for a pulse duration, which may cause a potential failure. One way to do so is using fast unity gain stages in the feed back loop, as proposed in [34].

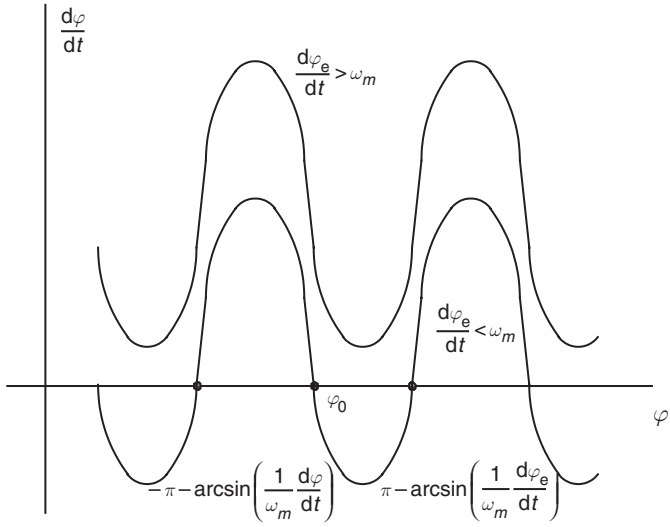


Figure 2.45 The phase plane of the SQUID current amplifier operating in the flux-locked mode. The lower curve corresponds to a case when $0 < \frac{d\varphi_e(t)}{dt} < \omega_m$. A non-compensation flux φ_0 constitutes a focal point of an "attraction" interval extending from $[-\pi - \arcsin(\frac{1}{\omega_m}\frac{d\varphi_e}{dt})]$ to $[\pi - \arcsin(\frac{1}{\omega_m}\frac{d\varphi_e}{dt})]$. The attraction interval means that the operating point within this interval has a chance to return to the focal position φ_0 from anywhere within the interval. If a pulse signal is applied on top of the defined gradient and the non-compensation error exceeds the limits defined by the attraction interval, the system will make a definite transition into a new equilibrium state. In the limit of $\frac{d\varphi_e}{dt} > \omega_m$, $\dot{\varphi}(\varphi)$ characteristics on the phase plane (the upper curve) no longer crosses the φ axis. This means that the system does not have a steady-state mode of operation

Although the theory allows for the operating point to vary in a range of $\pm\pi$, in practice, however, the failure-free operation requires a non-compensation error φ to be kept well within $\pm\frac{\pi}{4}$.

2.6.2 SQUID Current Amplifier in the Large Signal Limit (Noise)

It is obvious that the noise circulating in the amplifier loop increases an attempt frequency for the operating point to jump into another steady state. Therefore, this section is designed to quantify the phenomena and introduce noise corrections into expressions given in the previous section.

Figure 2.46 illustrates a SQUID amplifier incorporating the noise source $n(t)$. The spectral density of $n(t)$ is a quadrature sum of spectral densities of all noise sources present in the loop and caused by all possible

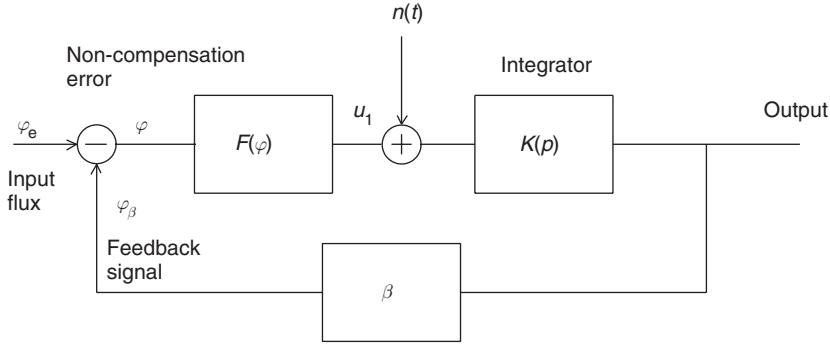


Figure 2.46 A block-diagram of a SQUID current amplifier in the flux-locked mode of operation incorporating the noise source $n(t)$. It consists of a SQUID-preamplifier unit with the joined transfer function $F(\varphi)$, an integrator with a Laplace transfer function $K(p) = \frac{1}{p\tau}$ and a non-inertial linear feedback circuit denoted as β. A spectral density of $n(t)$ is a quadrature sum of spectral densities of all noise sources present in the loop and caused by all possible mechanisms. Placing the noise source between the output of the preamplifier and the input of integrator is convenient as in real systems, the spectral characteristics of the total circulating noise would be measured at this particular point

mechanisms. Placing the noise source between the output of the preamplifier and the input of the integrator is convenient as in real systems, the spectral characteristics of the total circulating noise would be measured at this particular point.

The noise source $n(t)$ upgrades the motion equation (2.6.3) to a form given by

$$\frac{\mathrm{d}\varphi_\mathrm{e}}{\mathrm{d}t} + \omega_m \sin\varphi = \frac{\mathrm{d}\varphi_\mathrm{e}}{\mathrm{d}t} + \frac{\beta}{\tau} n(t) \tag{2.6.9}$$

Assuming the SQUID and the preamplifier non-inertial, which is generally true in real terms compared to the time constant of the integrator τ, $n(t)$ can be regarded as a white noise with a zero mean value and a spectral density S_v. The mean time duration from a moment, when the noise is applied to the system, to a moment, when the first failure in operation occurs, is defined by [35]

$$t_\mathrm{m} = \int_0^\infty \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \psi(\varphi, t)\mathrm{d}t\mathrm{d}\varphi \tag{2.6.10}$$

where $\psi(\varphi, t)$ represents the probability density of the fact that the non-compensation error has not reached $\pm\frac{\pi}{2}$ boundaries at the moment t. The

function $\psi(\varphi t)$ can be derived from the Fokker–Planck equation given by

$$\frac{\partial \psi(\varphi, t)}{\partial t} = \frac{\partial}{\partial \varphi}[\omega_{\mathrm{m}} \sin \varphi \cdot \psi(\varphi, t)] + \frac{\beta^2 S_{\mathrm{v}}}{4\tau^2} \frac{\partial^2 \psi(\varphi, t)}{\partial \varphi^2} \tag{2.6.11}$$

with the initial condition

$$\psi(\varphi, 0) = S_{\mathrm{v}} \delta[\varphi - \varphi(0)] \quad \text{at} \quad |\varphi| < \frac{\pi}{2} \tag{2.6.12}$$

and the normalised requirement defined as

$$\psi(\varphi, t) = 0 \quad \text{at} \quad |\varphi| > \frac{\pi}{2} \tag{2.6.13}$$

In the equation (2.6.12), $\delta(\varphi)$ is the delta function.

An equation similar to the equation (2.6.11) was treated by Viterbi in [29] regarding the dynamics of synchronous communication systems. Following his approach, the sought failure rate in a SQUID current amplifier with a single-pole integrator can be estimated using the following expression

$$N_{\mathrm{f}} = \frac{1}{t_{\mathrm{m}}} \approx \frac{\omega_{\mathrm{m}}}{\pi} \mathrm{e}^{\frac{\nu_{\mathrm{m}}^2}{S_{\mathrm{v}} \omega_{\mathrm{m}}}} \tag{2.6.14}$$

This equation holds for $\frac{\mathrm{d}\varphi_{\mathrm{e}}}{\mathrm{d}t} = 0$ and $\varphi(0) = 0$. In a general case, when the input signal gradient is present, one has to use expressions derived in the previous section (2.6.3)–(2.6.8), with a modified maximum frequency [36] given by

$$\omega'_{\mathrm{m}} = \omega_{\mathrm{m}} \sin\left[\frac{\pi}{2} - \frac{\sigma_{\varphi}(t)}{2} - \bar{\varphi}(t)\right] \tag{2.6.15}$$

with

$$\sigma_{\varphi}(t) = \sqrt{\int_0^{\infty} (\varphi - \bar{\varphi})^2 \psi(\varphi, t) \mathrm{d}\varphi} \tag{2.6.16}$$

responsible for the rms value of the non-compensation normalised magnetic flux error and $\bar{\varphi}$ is the mean value of it.

A similar exercise can be done for a two-pole system including an additional proportional-integrating stage. Interested readers should refer to the analysis presented in [37]. Here, we will only notice that fast SQUID current amplifiers tend to be built with a single major pole. This is explained by the fact that the feedback system must accommodate a number of parasitic poles associated, for instance, with long cables and/or an ultra low noise preamplifier. Some margin is required also to be immune against the noise circulating in the loop and temperature variations. These parasitic factors can consume a substantial part of the phase and amplification margins, which otherwise would be used to increase the loop amplification.

2.7 SQUID CURRENT AMPLIFIER AT ULTRALOW TEMPERATURE

In the section 2.3, we evaluated an energy resolution of the SQUID amplifier at a temperature of liquid helium, which amounts to approximately 4.2 K at normal laboratory conditions. All cryogenic detectors, however, are known to operate in the mK temperature range. For instance, tantalum tunnel junctions require cooling down to 250 mK, whereas transition edge sensor microcalorimeters show their best performance at a base temperature of 100 mK and lower.

Placing the SQUID in the close proximity with a primary sensor at the ULT stage is very beneficial for detector systems from two points of view:

1. reduction in the parasitic inductance of connecting wires and associated with them the microphonics and electromagnetic pick-up; and
2. further improvement of the intrinsic energy resolution of the SQUID, which, according to equation(2.3.10), should scale with temperature.

It is obvious that the basic principle of the ULT SQUID remains unchanged compared to its 4.2 K counterpart provided that the material of shunts stays resistive. Still, a SQUID designed for 4.2 K may not function at all or will not show expected improvements at ULT until some necessary specific alterations have been introduced into the design as well as the read-out technique. These specific features of the ULT operation will be discussed in this section.

The ULT SQUID requires a read-out preamplifier with a reduced voltage and current noise. For instance, a spectral density of the intrinsic flux noise is expected to be approximately $0.1\frac{\mu\Phi_0}{\sqrt{\text{Hz}}}$. With a parametric amplification $\eta = 0.5\frac{\text{mV}}{\Phi_0}$, this results in a spectral density of the voltage noise across the SQUID of approximately $50\frac{\text{pV}}{\sqrt{\text{Hz}}}$. This is much lower compared to the best practical semiconductor amplifiers. For instance, a differential stage based on a number of parallel MAT-02 demonstrated a voltage noise as low as $100\frac{\text{pV}}{\sqrt{\text{Hz}}}$ at a temperature close to 77 K. There is no obvious prospect to further reduce this figure. In order to overcome the noise mismatch and to preserve the intrinsic energy resolution is was proposed to utilise a second SQUID current amplifier as a preamplifier for the primary SQUID. Such configuration is often called the double stage SQUID current amplifier. Figure 2.47 demonstrates its schematic diagram. If a value of the shunt resistor, R_{sh}, is much lower than the dynamic resistance of the SQUID1, the latter operates in the voltage-biased mode. In this case, an applied magnetic flux modulates a current flowing through the input inductance of the SQUID2, L_{i2}. The current is further amplified by a second SQUID stage and fed back to the SQUID1 via mutual inductance M_{F}.

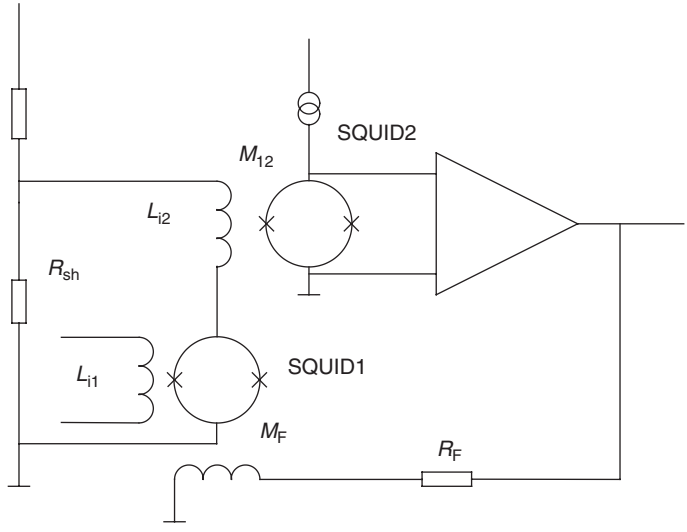


Figure 2.47 A double-stage SQUID current amplifier. If a value of the shunt resistor, R_{sh}, is much lower than the dynamic resistance of the SQUID1, the latter operates in the voltage-biased mode. In this case, an applied magnetic flux modulates a current flowing through the input inductance of the SQUID2, L_{i2}. The current is further amplified by a second SQUID stage, semiconductor unit and fed back to the SQUID1 via mutual inductance M_{F}

Speaking about parametric amplification for the input stage, we now have in mind a magnetic flux-to-current transfer function

$$\eta_{\mathrm{I}} = \frac{\partial I}{\partial \Phi} \tag{2.7.1}$$

as opposite to the flux-to-voltage parametric amplification

$$\eta_{\mathrm{V}} = \frac{\partial V}{\partial \Phi} \tag{2.7.2}$$

which we introduced earlier for a SQUID operating in the current-biased mode.

So far, a thorough analytical model of the voltage-biased SQUID has not been developed. In practice, it is more convenient to characterise the SQUID in the current-biased mode of operation, for which all necessary models exist, and then convert all major parameters into a form suitable for the voltage-biased mode. Rough estimations can be done based on an assumption that the SQUID dynamic resistance does not depend on a type of the bias. This, however, may lead to a serious error. Whereas a high-frequency current circulating in the SQUID loop itself may not be affected significantly by the bias circuitry, the output current signal certainly is. The matter is that a signal current of several μA couples inductively back to the SQUID loop. As a result, this extra negative or positive feedback alters the shape of its Φ–*I* characteristics and a parametric amplification, η_{I}. An example of the changing shape is shown in Figure 2.48. Figure 2.48(a, b) represent a *V*–Φ and *I*–Φ characteristics of the same SQUID measured with the current and voltage bias respectively. An additional slope of the *I*–Φ curve can alter the parametric amplification by as much as 50%. In order to correct for this and other unaccounted errors, one needs to know a real cross-correlated dynamic resistance of the SQUID. The cross-correlated dynamic resistance is defined as follows [38]

$$R_{\mathrm{dCV}} = \frac{\eta_{\mathrm{V}}}{\eta_{\mathrm{I}}} \tag{2.7.3}$$

With the cross-correlated dynamic resistance, we can predict accurately all characteristics of a voltage-biased SQUID using well-developed models for the current-biased SQUID. For instance, the total spectral density of flux noise of the double stage SQUID amplifier is given by [39]

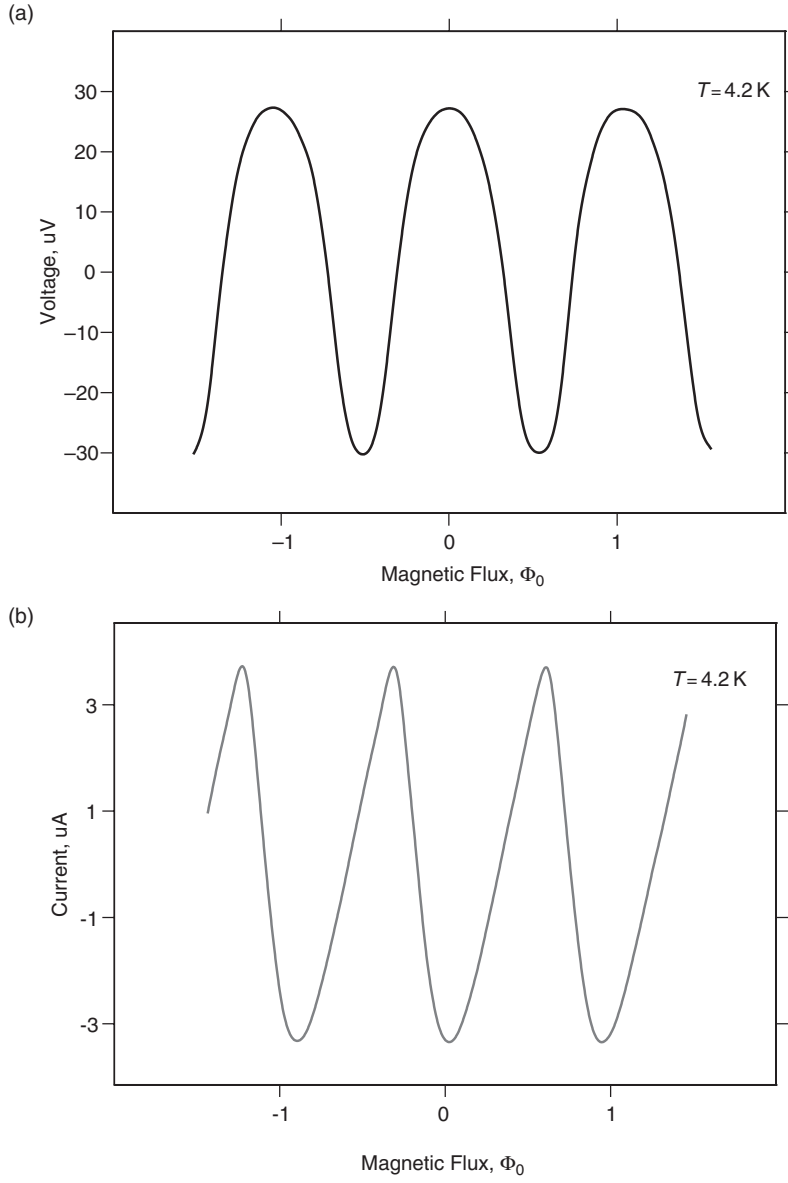


Figure 2.48 A V–Φ characteristics and an I–Φ characteristics of the same SQUID measured with the current and voltage bias respectively. A signal current of several μA in the voltage-biased configuration couples inductively back to the SQUID loop. As a result, this extra negative or positive feedback alters the shape of its Φ–I characteristics and a parametric amplification, η_I. An additional slope of the I–Φ curve can alter the parametric amplification by as much as 50%. In order to correct for this and other unaccounted errors, one needs to know a real cross-correlated dynamic resistance of the SQUID

$$S_{\Phi\Sigma} = \frac{1}{\eta_{I1}^2}\left[\frac{S_{V1}}{(R_{dCV1} + R_{sh})^2} + \frac{S_{\Phi 2}}{M_{12}^2}\right] \tag{2.7.4}$$

where, according to [18], a spectral density of the voltage noise of the front SQUID is defined as

$$S_{V1} = \frac{24k_b T_1}{R_1}\left[R_{d1}^2 + \left(R_{d1} - \frac{R_1}{2}\right)^2\right] \tag{2.7.5}$$

and a spectral density of the flux noise of the second SQUID can be found from the following equation

$$S_{\Phi 2} = \frac{2k_b T_2 L_{s2}^2}{R_2}\left[1 + \frac{R_2}{R_{d2}}\left(4 + 2\frac{R_{opt}}{R_{d2}}\frac{T_a}{T_2}\right)\right] \tag{2.7.6}$$

In the equations (2.7.4)–(2.7.6), R_{d1} and R_{d2} represent the dynamic resistance of the SQUID1 and the SQUID2 in the current-biased mode respectively, η_{V1} and η_{I1} are the parametric amplification of the SQUID1 in the current-biased mode and in the voltage-biased mode respectively, R_1 and R_2 are shunt resistors incorporated into SQUID1 and SQUID2 with temperatures T_1 and T_2 respectively, R_{opt} and T_a denote the optimum source resistance and the noise temperature of the semiconductor preamplifier following the SQUID2 respectively, R_{sh} is the shunt resistance of the voltage bias circuitry, M_{12} is the mutual inductance between the SQUID2 and its own input coil, and L_{s2} is the inductance of the SQUID2.

A coupled resolution of the whole double-stage amplifier can be estimated as

$$\varepsilon_{c\Sigma} = \frac{S_{\Phi\Sigma}}{2k_1 L_{s1}} \tag{2.7.7}$$

with k_1 being the coupling factor between the SQUID1 and its input coil, whereas L_{s1}, is the inductance of the SQUID1.

A spectral density of the current noise is given by

$$S_{I\Sigma} = \frac{S_{\Phi\Sigma}}{M_1^2} \tag{2.7.8}$$

where M_1 is the mutual inductance between the SQUID1 and its input coil.

A plot of the coupled energy resolution of a SQUID with parameters summarised in Table 2.1 is shown in Figure 2.27. It was measured in a double SQUID configuration at a temperature of 4.2 K. With an effective value of the input inductance $L_{1\text{eff}} = 360\,\text{nH}$ the energy resolution results in a spectral density of the current noise as low as $0.25\,\frac{\text{pA}}{\sqrt{\text{Hz}}}$. Such a current resolution is already sufficient for a majority of X-ray spectrometers based on superconductor detectors. For optical measurements in the infrared region, however, we need to look for the further improvement in the current resolution of approximately another order of magnitude. As we mentioned before, this can come only from the lowering of the electron temperature T_1 of the SQUID1 resistive shunts.

ULT SQUID chips normally operate in a high vacuum chamber being attached to a polished surface of the cold finger. Resistive shunts can dissipate their $I_c^2 R$ power only through the electron–phonon exchange mechanism between the shunt and a chip substrate and further between the substrate and the cold finger. Below 1 K, the average temperature of the electron system in shunt resistors becomes higher compared to a base temperature due to reduction in the electron–phonon scattering rate and the Kapitza resistance. In order to prevent the "overheating", SQUIDs, like any other electronic components at a room temperature, require attaching cooling radiators. The radiator represents a fin made of a normal metal film connected to one side of a shunt [40]. An area and a volume of the fin must be chosen much larger compared to the area and the volume of the shunt. Therefore, electrons have better chance to scatter on other "cold" electrons or/and phonons. A cooling mechanism is provided by the interdiffusion of free electrons between the shunt and the fin. Whereas hot electrons diffuse out of the shunt into the cooling fin for the relaxation, relaxed electrons take their place back in the shunt, thus, reducing its electron temperature.

The thermal balance in thin metal films is based on the free-electron model in the dirty limit. The model takes into account existing impurities and the boundary scattering from the film surfaces. Four basic phenomena define the temperature of the electron system in a thin film deposited on a dielectric substrate. These are:

1. inelastic electron–electron collisions;
2. inelastic electron–phonon scattering;
3. elastic scattering on impurities and surfaces of the thin-film; and
4. phonon–phonon interaction between the film and the substrate.

Once a power $P = \frac{V^2}{R}$ is applied to a shunt resistor, its each electron acquires energy

$$\Xi = eEI_{\mathrm{d}} \tag{2.7.9}$$

Here e is the electron charge, $E = V/L$, L the length of the shunt, and I_{d} is the diffusion length of the shunt material defined as

$$I_{\mathrm{d}} = \upsilon_{\mathrm{F}} \sqrt{\frac{1}{3} r_{\mathrm{imp}} r_{\mathrm{e-ph}}} \tag{2.7.10}$$

In the expression (2.7.10), υ_{F} denotes the Fermi velocity, r_{imp} is the scattering rate on impurities and $r_{\mathrm{e-ph}}$ is the electron–phonon scattering rate.

With no radiators in place, the energy of "hot" electron in the shunt, partially averaged by the electron–electron collisions (often called the thermalisation process), is transferred to phonons in the film. This then is carried away from the shunt by the phonon–phonon interaction mechanism between the film and the substrate. As an example, the electron–electron and electron–phonon scattering rates for a 50-nm thick copper–gold film are given by [41–43]

$$\Gamma_{\mathrm{e-e}} = (1.34 \times 10^8 R_{\mathrm{sq}} T_{\mathrm{e}}) + 1.7 \times 10^7 T_{\mathrm{e}}^2 \tag{2.7.11}$$

$$\Gamma_{\mathrm{e-ph}} = 5 \frac{W T_{\mathrm{e}}^3}{\gamma} \tag{2.7.12}$$

respectively. In these equations, R_{sq} is the resistance of the film per square, $\gamma = 71.5 \, \frac{\mathrm{J}}{\mathrm{m}^3\mathrm{K}^2}$ is the linear term of the specific heat of Au, $\mathrm{W} = 2.4 \times 10^9 \, \mathrm{Wm}^{-3}\mathrm{K}^{-5}$ is a factor found by Wellstood *et al.* for Cu–Au in the paper [40]. Figure 2.49 illustrates equations (2.7.11) and (2.7.12). For a typical resistor shunt with $L = 10$-μm long and 100 nm thick, a point of $T_{\mathrm{cr}} = 600$ mK represents a cross-over temperature, below which the electron–phonon mechanism is too slow to remove all heat without additional cooling fins.

The temperature of the electron system in a thin film is given by [40]

$$T_{\mathrm{e}} = \left(\frac{P}{W\Omega} + T_{\mathrm{p}}^5 \right)^{\frac{1}{5}} \tag{2.7.13}$$

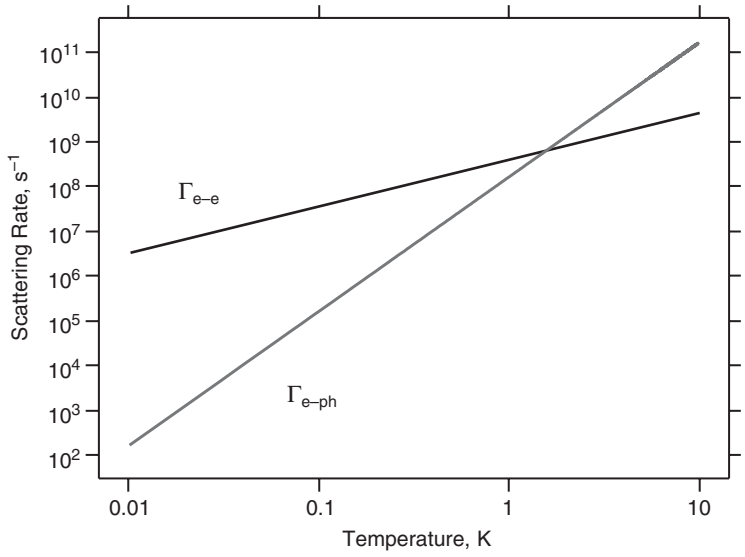


Figure 2.49 The electron–electron and electron–phonon scattering rates for a 50 nm thick copper–gold film

where Ω is, in our case, the volume of the shunt resistor including fins and T_p is the phonon temperature of the film defined as [44]

$$T_p = \left(\frac{4R^k_{f-s} P T_s^3}{A_f} + T_s^4 \right)^{\frac{1}{4}} \tag{2.7.14}$$

In expression (2.7.14), R^k_{f-s} denotes the Kapitza resistance, A_f is the area of the shunt and the fin attached directly to the substrate, and T_s is the temperature of the substrate.

Example 2.2

A SQUID with a modulation depth of the V–Φ characteristics of 150 μV peak-to-peak and $R = 10\,\Omega$ dissipates a power $P = 2$ nW. Figure 2.50 demonstrates a plot of an electron temperature, T_e, as a function of the base temperature, T_s, that was calculated using equations (2.7.13) and (2.7.14). In the calculations, the fins had an area of $A_f = 1\,\text{mm}^2$ and 150-nm thick. A factor R^k_{f-s} was chosen to be $2 \times 10^{-3}\ \frac{\text{Km}^2}{\text{W}}$ [44]. One can conclude from $T_e(T_s)$ curve that fins with selected dimensions are efficient in the absorption and conduction to the base all the dissipation

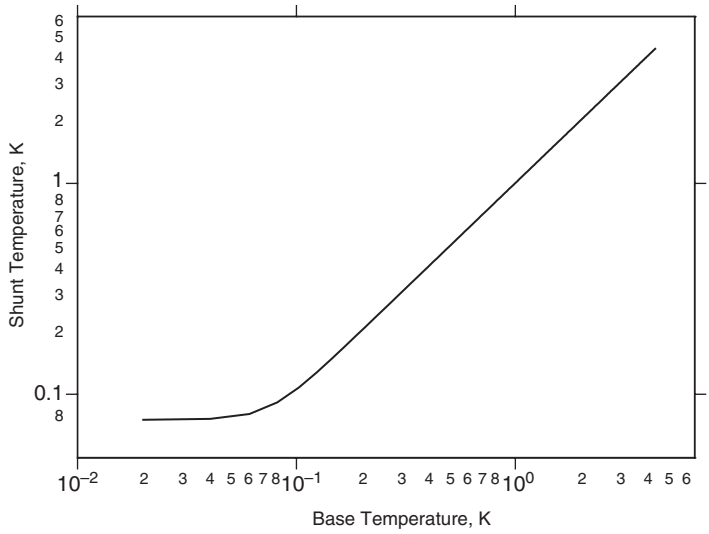


Figure 2.50 A plot of an electron temperature, T_e, as a function of the base temperature, T_s for dissipating power of $P = 2\,\text{nW}$. In the calculations, fins had an area of $A_f = 1\,\text{mm}^2$ and 150-nm thick. $T_e(T_s)$ shows that fins with selected dimensions are efficient in the absorption and conduction to the base all the dissipation power down to 100 mK. If a lower temperature is required the area A_f must be further increased

power down to 100 mK. If a lower temperature is required the area A_f must be increased. An operating temperature of the majority of practical TES detectors is approximately 100 mK. Therefore, an ULT SQUID with a couple of $1\,\text{mm}^2$ thin-film cooling radiators would be a good choice for a general purpose device.

Now we can evaluate a current resolution of a double-stage SQUID amplifier. Here, we assume that SQUID1 operates at a temperature of 100 mK, whereas SQUID2 is placed at 4.2 K. In order to reduce the cross-talk between SQUIDs, a second-stage SQUID2 was chosen with a reduced input inductance of 58 nH. Figure 2.51 presents a spectral density of the total flux noise and components contributing into it. At $T = 100\,\text{mK}$, it amounts to $\sqrt{S_{\Phi\Sigma}} \approx 8.5 \times 10^{-8}\,\frac{\Phi_0}{\sqrt{\text{Hz}}}$. This corresponds to a coupled energy resolution as low as $1.3\,h$, where $h = 6.626 \times 10^{-34}$ Js is the Planck constant. Below 220 mK, the noise is still dominated by a cooled semiconductor preamplifier based on bipolar transistors with $R_{opt} = 250\,\Omega$ and $T_a = 18\,\text{K}$. The current noise of such SQUID amplifier is expected to be $70\,\frac{\text{fA}}{\sqrt{\text{Hz}}}$ with an effective value of the input inductance of 367 nH.

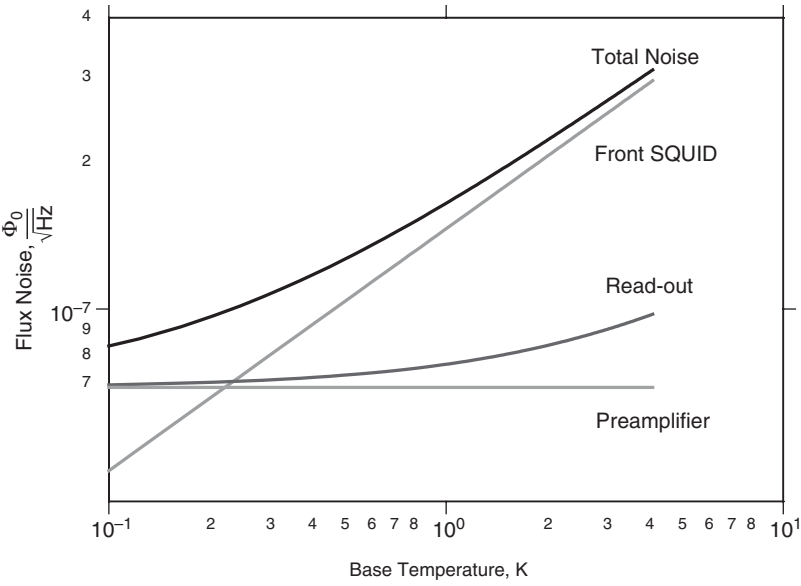


Figure 2.51 A spectral density of the flux noise of the double stage SQUID current amplifier. The total noise is a quadrature sum of three components: noise of the front SQUID, the second-stage SQUID and semiconductor preamplifier. Below 220 mK, the noise is still dominated by the semiconductor preamplifier based on bipolar transistors with $R_{\text{opt}} = 250\,\Omega$ and $T_{\text{a}} = 18\,\text{K}$. The current noise of such SQUID amplifier is expected to be $70\,\frac{\text{fA}}{\sqrt{\text{Hz}}}$ with an effective value of the input inductance of 367 nH

2.8 SQUID VOLTAGE AMPLIFIER

2.8.1 Basic Principles

The SQUID voltage amplifier is useful in the combination with a transition edge sensor (TES) microcalorimeter operating in the current-bias mode. In this mode of operation, the transfer function of the TES does not depend on the output current caused by an incoming event. This improves the microcalorimeter performance in terms of the linearity and the dynamic range.

There are two ways to build a SQUID voltage amplifier. A first option is based on employing a system with the flux-nulling feedback. Figure 2.52 illustrates the concept. In this figure, E_{s} represents a voltage source to be measured with an output impedance R_{s}. R_{ref} is the reference resistance, the value of which is well defined at a temperature of operation. The SQUID system works essentially as a linear current meter. The device is self-calibrated using the periodicity of the V–Φ curve (section 2.3). Once

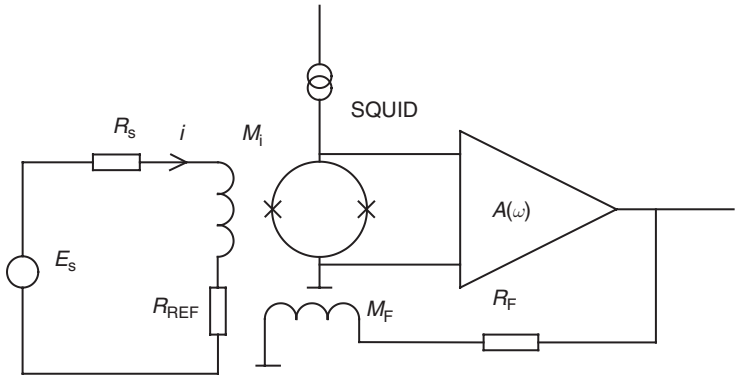


Figure 2.52 A circuit diagram of the SQUID voltage amplifier with the flux-nulling feedback. E_s represents a voltage source to be measured with an output impedance R_s. R_{ref} is the reference resistance, the value of which is well defined at a temperature of operation. Voltage measurements require a condition $R_{REF} \gg R_s$ to be met which is not desirable as large R_{REF} reduces the sensitivity of the voltage amplifier

an input current I was measured, we can derive the sought value of E_s from the following simple equation

$$E_s = (R_s + R_{ref})i \tag{2.8.1}$$

If the true voltage measurement condition given by

$$R_{REF} \gg R_s \tag{2.8.2}$$

is met the input current I becomes a direct measure of E_s.

In practice, however, the condition set by (2.8.2) is not desirable. A large R_{REF} limits the input current I and together with it reduces the sensitivity of the voltage amplifier. To avoid the drawback one can employ another feedback technique incorporating a current-nulling feedback. Figures 2.53 and 2.54 show a schematic circuit diagram and a functional block diagram of the technique, respectively. In this amplifier, the voltage E_s is compensated by the feedback current that is fed directly into the input circuit. The input impedance of the amplifier is given by

$$Z_i(j\omega) \approx R_{REF} + j\omega L_{1eff} + (R_T + j\omega L_{1eff})G_{loop}(j\omega) \tag{2.8.3}$$

where

$$G_{loop}(j\omega) = \Gamma(j\omega)\eta K_1(j\omega)K_2(j\omega)\beta \tag{2.8.4}$$

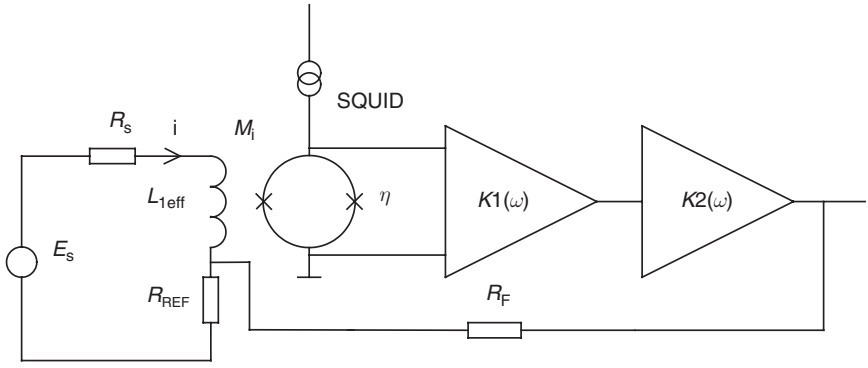


Figure 2.53 A circuit diagram of the SQUID voltage amplifier with the current-nulling feedback. The feedback loop increases the input impedance of the SQUID voltage amplifier proportionally to the loop amplification, G_{loop}. In this configuration, a value of R_{REF} can be reduced to a level when its noise contribution becomes negligibly small compared to the thermal noise generated by signal source, R_s

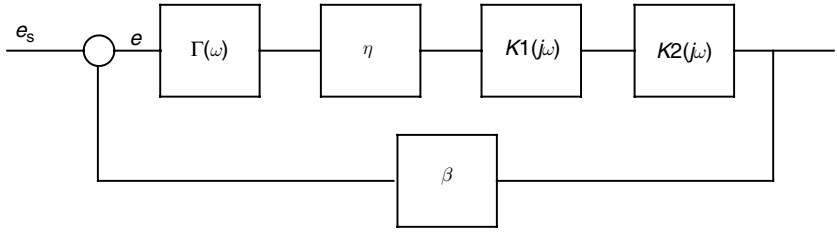


Figure 2.54 A functional block diagram of the SQUID voltage amplifier with the current-nulling feedback. $\Gamma(j\omega)$ is the transfer function of the input circuit, η is the parametric amplification of the SQUID, $K_1(j\omega)$, $K_2(j\omega)$ are transfer functions of inertial circuitry, β is the feedback factor

is the loop gain of the voltage amplifier,

$$\Gamma(j\omega) = \frac{M_i}{R_T + j\omega L_{1eff}} \tag{2.8.5}$$

is the voltage-to-current transfer function of the input circuit,

$$R_T = R_s + R_{REF} \tag{2.8.6}$$

is the total resistance of the input circuit,

$$K_1(j\omega) = K_0\left(1 + \frac{1}{j\omega\tau_1}\right) \tag{2.8.7}$$

is the transfer function of the proportional integrated circuit,

$$K_2(j\omega) = \frac{1}{j\omega\tau_2} \tag{2.8.8}$$

is the transfer function of a single-pole integrator,

$$\beta = \frac{R_{\mathrm{REF}}}{R_{\mathrm{REF}} + R_{\mathrm{F}}} \tag{2.8.9}$$

is the transfer function of the feedback circuit. In equation (2.8.9), we assumed $R_{\mathrm{REF}} \ll R_{\mathrm{s}} \cdot R_{\mathrm{F}}$ is the feedback resistor and M_{I} is the mutual inductance between the SQUID and its input coil.

According to equation (2.8.3), the feedback loop increases the input impedance of the SQUID voltage amplifier proportionally to the loop amplification, G_{loop}. In this case, a value of R_{REF} can be reduced to a level when its noise contribution becomes negligibly small compared to thermal noise generated by sinusoidal signal source, R_{s}.

A slew rate of the voltage amplifier is defined as

$$\alpha = E_{\mathrm{s\,max}}\omega \tag{2.8.10}$$

where $E_{\mathrm{s\,max}}$ is the maximum amplitude of the input voltage that the system can tolerate, that is system functions in the linear mode without jumps of the operating point into neighbouring steady states. The maximum amplitude of the input sinusoidal signal is given by the product

$$E_{\mathrm{s\,max}} = \Delta E|G(j\omega)| \tag{2.8.11}$$

where ΔE is the dynamic range of the preamplifier. The dynamic range is entirely governed by the flux dynamic range of the SQUID as it can be seen from the following equation

$$\Delta E = \frac{R_{\mathrm{T}}}{M_i}\Delta\Phi \tag{2.8.12}$$

According to equations (2.8.3) and (2.8.10), both the input impedance and the slew rate benefit of the maximisation of the loop amplification, which in its expanded form is given by

$$G_{\mathrm{loop}}(j\omega) = \frac{M_i}{R_{\mathrm{T}}(1 + j\omega\tau_i)} \frac{\eta K_0}{1 + j\omega\tau_{\mathrm{a}}} \left(1 + \frac{1}{j\omega\tau_1}\right) \frac{1}{j\omega\tau_2}\beta \tag{2.8.13}$$

where $\tau_i = \frac{L_{\text{ieff}}}{R_{\text{T}}}$ represents the time constant of the input circuit, $\frac{K_0}{1+j\omega\tau_{\text{a}}}$ is the intrinsic frequency response of the preamplifier reading out the information from the SQUID.

Assuming that the SQUID voltage amplifier provides a linear response, that is a SQUID parametric amplification $\eta = \text{const}$, we can apply the Bode theorem to explore stability of the amplifier. Following [28], a system incorporating minimum phase components is absolutely stable if the a phase shift at the unity gain frequency ω_{m}

$$\phi(\omega_{\text{m}}) = \frac{1}{\pi}\int_0^{\infty} \frac{dG_{\text{loop}}(\omega)}{d\omega} \ln \coth \left|\frac{\omega}{2}\right| \text{d}\omega \tag{2.8.14}$$

is less than $|\pi$-stability margin$|$. Substituting (2.8.13) into (2.8.14) and taking the integral, we derive three limiting condition imposed on the system. These are

1. $\tau_i = \frac{L_{\text{ieff}}}{R_{\text{T}}} \ll \frac{1}{\omega_{\text{m}}}$ (2.8.15)

2. $\tau_{\text{a}} \ll \frac{1}{\omega_{\text{m}}}$ (2.8.16)

3. $\tau_1 > \frac{4}{\omega_{\text{m}}}$ (2.8.17)

One needs to mention that ordinary correction and phase compensation methods, developed for conventional semiconductor amplifiers helping to ensure their absolute stability, are not always applicable for SQUID voltage amplifier. In the latter, the output impedance of the signal source, R_{s}, directly affects the loop amplification through the time constant and the transfer function of the input circuit at the same time. Therefore, a perfectly corrected system for one value of R_{s} may become unstable when R_{s} has changed. An example of a signal source with a variable R_{s} is a transition edge sensor. Its impedance can change from zero to several Ohms. This requires the SQUID voltage amplifier to maintain its absolute stability in the whole range of source impedance R_{s} without extra correction stages.

The minimum loop amplification in the working frequency range can be derived from the following condition

$$R_{\text{s max}} \ll |Z_i(j\omega)| \tag{2.8.18}$$

whereas the minimum input inductance is given by

$$L_i^{\min} = \frac{R_T S_\Phi}{4k_c^2 L_s k_b T} \tag{2.8.19}$$

where k_c is the coupling factor between the SQUID and the input coil, k_b is the Boltzmann constant, L_s is the SQUID inductance.

Equation (2.8.19) assumes that the noise contribution from a signal source output impedance and the SQUID amplifier are equal. In practice, it is desirable to have $L_{ieff} \gg L_I^{\min}$ to minimise the latter, unless requirement to provide a slew rate as large as possible takes priority over the noise minimisation.

The voltage amplification factor of the SQUID voltage amplifier is given by

$$K_{fb} \approx \frac{1}{\beta} \frac{G_{loop}(j\omega)}{1 + G_{loop}(j\omega)} \tag{2.8.20}$$

In a frequency range, where a condition $|G_{loop}(j\omega)| \gg 1$ is met, expression (2.8.20) can be written in the simplified form

$$K_{fb} \approx 1 + \frac{R_F}{R_{REF}} \tag{2.8.21}$$

2.8.2 Noise of SQUID Voltage Amplifier

A small-signal equivalent circuit diagram of the SQUID voltage amplifier is shown in Figure 2.55. In this diagram, e_n represents a voltage noise with a spectral density given by

$$S_e \cong 4k_b T R_{REF} + \frac{R_T^2}{M^2} S_\Phi \tag{2.8.22}$$

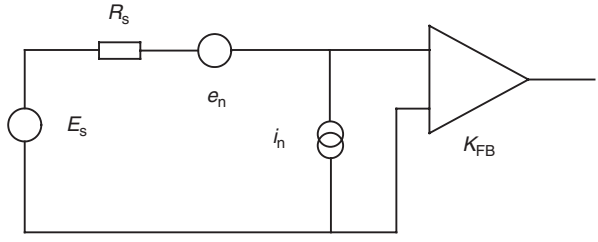


Figure 2.55 A small-signal equivalent circuit diagram of the SQUID voltage amplifier

There is also some current noise i_n induced by the fluctuation current circulating in the SQUID. Denoting S_e^{SQ} as a spectral density of the voltage noise induced across the input inductance L_{ieff}, a spectral density of the current noise in the input circuit takes the following form

$$S_i = \frac{S_e^{SQ}}{|Z_i|^2} \tag{2.8.23}$$

At a temperature of 4.2 K the current noise can safely be ignored.

Sometimes, the noise property of SQUID voltage amplifiers is quoted as a noise temperature, T_n. The noise temperature is defined as

$$T_n = \frac{S_e + S_i R_s^2}{4k_b R_s} \approx \frac{S_e}{4k_b R_s} \tag{2.8.24}$$

Example 2.3

In this example, we will evaluate a spectral density of the voltage noise and the noise temperature of a SQUID voltage amplifier operating in conjunction with a TES detector. All necessary parameters of the SQUID are summarised in Table 2.1. We assume that the TES has a nominal resistance at a working point $R_s = 1\,\Omega$. The SQUID itself is placed at a temperature of 4.2 K, whereas the TES and the reference resistor $R_{REF} = 0.1\,\Omega$ are mounted on the flat surface of a cold finger with a base temperature of $T_{cf} = 50$ mK. The bias power elevates temperature of the sensor to a value of $T_{TES} = 100$ mK.

Substituting the expression (2.3.21) into (2.8.22), a spectral density of the voltage noise of the amplifier is given by

$$\sqrt{S_e} = \sqrt{4k_b T_2 R_{REF} + \frac{(R_s + R_{REF})^2}{M^2}\varepsilon 2k_c^2 L_s} \tag{2.8.25}$$

If a double stage SQUID amplifier is employed the voltage noise amounts to

$$\sqrt{S_e} = \sqrt{\begin{array}{c} 4 \times 1.38 \times 10^{-23} \times 0.05 \times 0.1 + \dfrac{(1+0.1)^2}{(2.5 \times 10^{-9})^2} \\ \times 10^{-32} \times 2(0.77)^2 \times 30 \times 10^{-12} \end{array}} \approx 0.6\,\frac{\text{pV}}{\sqrt{\text{Hz}}}$$

The figure increases to a value of approximately $1\,\frac{\mathrm{pV}}{\sqrt{\mathrm{Hz}}}$ with a single stage SQUID amplifier whose coupled energy resolution is equal to $\varepsilon = 1.25 \times 10^{-31}\,\frac{\mathrm{J}}{\mathrm{Hz}}$.

2.8.3 The Large Signal Dynamics of the SQUID Voltage Amplifier

The presence of the SQUID makes the voltage amplifier also a system with multiple steady states. Similar to the case with the current amplifier, an operating point on the V–Φ characteristics can jump from a dedicated quasilinear interval onto neighbour branches. There are two mechanisms involved that cause these jumps. These are:

1. the derivative of the input signal exceeds the slew rate so that the feedback cannot follow the input signal; and
2. if the noise-bandwidth product referred to the SQUID input becomes comparable with the dynamic range $\Delta\Phi$.

The second condition implies that large noise circulating in the amplifier loop sets the upper limit on its bandwidth.

The major components of the noise in the SQUID voltage amplifier include a flux noise of the SQUID itself and a current noise of the input circuit generated by the reference resistor as well as by a signal source. Both of them are joined into one noise source e_{n} attached to the output of the SQUID preamplifier, where it would be characterised in a real system (Figure 2.56). In the figure, E_{s} is the voltage input signal, E_{F} is the feedback voltage and e is the voltage non-compensation error.

The equation of motion for a system including a single-pole integrator $K(j\omega) = \frac{1}{j\omega\tau_2}$ only is given by

$$\frac{\mathrm{d}e(t)}{\mathrm{d}t} - \omega_{\mathrm{m}} e(t) = \frac{\beta}{\tau_2} e_n(t) + \frac{\mathrm{d}e_{\mathrm{s}}(t)}{\mathrm{d}t} \tag{2.8.26}$$

with the bandwidth defined as

$$\omega_{\mathrm{m}} = \frac{M_i \eta K_0 \beta}{R_T \tau_2} \tag{2.8.27}$$

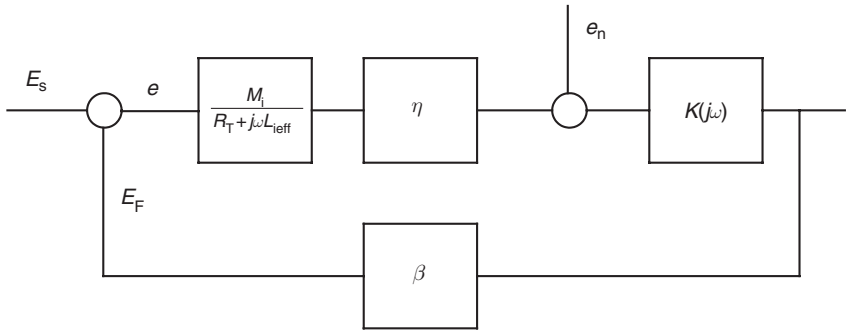


Figure 2.56 A functional block diagram of the SQUID voltage amplifier with the current-nulling feedback with incorporating the noise source $e_n(t)$. It consists of a SQUID-preamplifier unit with the joined transfer function η, an integrator with a Laplace transfer function $K(p) = \frac{1}{p\tau}$ and a non-inertial linear feedback circuit denoted as β. A spectral density of $e_n(t)$ is a quadrature sum of spectral densities of all noise sources present in the loop and caused by all possible mechanisms. Placing the noise source between the output of the preamplifier and the input of the integrator is convenient as in real systems, the spectral characteristics of the total circulating noise would be measured at this particular point

Provided that a time constant of the input circuit $\tau_i < \frac{1}{\omega_m}$, e_n can be treated as a Markowian process with a zero mean value and a spectral density given by

$$S_{en} \approx \eta^2 \left(S_\Phi + \frac{4k_b T M_i^2}{R_T} \right) \tag{2.8.28}$$

Perhaps the most visual parameter characterising the influence of a random process on the system dynamics is an average time t_{av} from a moment when the process starts until a moment when an operating point hits one of the boundaries of the dynamic range $[-e_b, e_b]$. According to [35], the average time can be found utilising the following expression

$$t_{av} = \int_0^\infty \int_{-e_b}^{e_b} w(e,t) \mathrm{d}t \mathrm{d}e \tag{2.8.29}$$

where e_b is a value of the input voltage at which the operating point of the SQUID reaches the end of the dynamic range $\Delta\Phi$ when the loop is open, that is

$$|e_b| = \frac{\Delta\Phi}{2M_i} R_T \tag{2.8.30}$$

The distribution function of the spectral density of the random process $e(t)$ can be derived from a solution of the Fokker–Planck equation written for (2.8.26) in the form

$$\frac{\partial w(e,t)}{\partial t} = \frac{\partial}{\partial e}[\omega_{\mathrm{m}} w(e,t)] + \frac{S_{\mathrm{en}}\beta}{\tau_2}\frac{\partial^2 w(e,t)}{\partial e^2} \tag{2.8.31}$$

with an initial condition

$$w(e,t) = S_{\mathrm{en}}\delta[e - e(0)] \quad \text{at} \quad |e| < e_{\mathrm{b}} \tag{2.8.32}$$

and the normalising condition

$$w(e,t) = 0 \quad \text{at} \quad |e| \geq e_{\mathrm{b}} \tag{2.8.33}$$

Here δ is the Dirac function and $e(0)$ is a value of $e(t)$ at $t = 0$. According to [35], the average rate of transitions of the working point beyond the dynamic range is given by

$$N = \frac{1}{t_{\mathrm{av}}} = \frac{\omega_{\mathrm{m}}}{\pi}\exp\left(-\frac{V_{\mathrm{m}}^2}{S_{\mathrm{en}}\omega_{\mathrm{m}}}\right) \tag{2.8.34}$$

where $|V_{\mathrm{m}}| = \eta^2\frac{\Delta\Phi}{2}$ is the voltage swing of the V–Φ characteristics extending from the middle to the end of the dynamic range ΔV.

REFERENCES

[1] E. Fairstein and J. Hahn. *Nucleonics* **23**, 56 (1965).
[2] K. Kurokado. *X-ray Spectrometry* **29**, 137 (2000).
[3] R.J. Schoelkopf, P. Wahlgren, A.A. Kozhevnikov *et al. Science* **280**, 1238 (1998).
[4] G. Bertuccio and P. Rehak, D. Xi. *Nuclear Instruments and Methods in Physics Research* **A326**, 71 (1993).
[5] F. Fiorini and P. Lechner. *IEEE Trans. Nucl. Sci.* **46**, 761 (1999).
[6] F. Goulding, D. Landis and N. Madden. *IEEE Trans. Nucl. Sci.* **30**, 301 (1983).
[7] T. Nashashibi and G. White. *IEEE Trans. Nucl. Sci.* **37**, 452 (1990).
[8] See, for instance, http:/ /www.xia.com.
[9] Application Note of Analog Device, Inc., AN-359, 13 (1970).
[10] J. Millman and A. Grabel, *Microelectronics*. New York: McGraw-Hill, 1987, p. 579.
[11] W. Shockley. *Proc. IRE* **40**, 1365 (1952).
[12] M. Kishimoto, M. Ukibe, M. Katagiri *et al. Nucl. Instrum. Meth. In Phys. Res.* **A370**, 126 (1996).

[13] V. Radeka. *Ann. Rev. Nucl. Part. Sci.* **38**, 217 (1988).
[14] A. Van der Ziel. *Proc. IRE* **50**, 1808 (1962).
[15] G. Bertuccio, A. Pulia and G. De Geronimo. *Nucl. Instrum. Meth. In Phys. Res.* **A380**, 301 (1996).
[16] T. Blaloch. *IEEE Trans. Nucl. Sci.* **NS13**, 457 (1966).
[17] R. den Hartog, A. Kozorezov, J. Wigmore *et al. Phys. Rev.* **B66**, 094511/1–14 (2002).
[18] T. Ryhanen. PhD Thesis, Helsinki University, 1995 (unpublished).
[19] J.M. Manley and H.E. Rowe. *Proceedings of IRE* p. 904 (1956).
[20] K. Tsukada, J. Kawai, G. Uehara and H. Kado. *Rev. Sci. Instrum.* **66**, 2631 (1995).
[21] R. Cantor, T. Ryhanen, D. Drung, H. Koch and H. Seppa, *IEEE Trans. Magn.* **27**, 2927 (1991).
[22] B. Muhlfelder, W. Johnson and M. Cromar. *IEEE Trans. Magn.* **19**, 303 (1983).
[23] H. Seppa. *Private communication*. 2 May, 2002.
[24] V. Polushkin, D. Glowacka, R. Hart and J. Lumley. *Rev. Sci. Instrum.* **70**, 1713 (1999).
[25] M. Kiviranta and H. Seppa. *IEEE Trans. Appl. Supercond.* **5**, 2146 (1995).
[26] D. Drung, R. Cantor, M. Peters *et al. Appl. Phys. Lett.* **57**, 406 (1990).
[27] V. Polushkin. *Rus. Metrology* **5**, 48 (1984).
[28] H. Bode. Network Analysis and Feedback Amplifier Design. D. VanNostrand Inc., 1945.
[29] A. Viterbi. *Phase-locked loops and Their Applications*, IEEE, NY, 1978.
[30] D. Drung, R. Cantor, M. Peters *et al.*, *Appl. Phys. Lett.* **57**, 406, 1990.
[31] Cryomech Inc. Technical Note, http:/ /www.cryomech.com/Brochures/PT405% 2011-25–01%20Front-N-Back.pdf
[32] V. Polushkin. *Sov. Scientific Instrumentation* **6**, 68 (1984).
[33] A. Petrov, *Private communication* 1984.
[34] V. Polushkin. *Sov. Scientific Instrum.* **7**, 53 (1985).
[35] V. Tikhonov, M. Mironov. Markovian processes, Rus. Nauka, 1978.
[36] V. Polushkin. PhD Thesis. Kiev Technical University, 1985 (unpublished).
[37] V. Polushkin. *Sov. Scientific Instrumentation* **11**, 59 (1983).
[38] V. Polushkin. *Journal of Low Temperature Physics* **118**, 1 (2000).
[39] V. Polushkin, D. Goldie and J. Lumley. *Physica B*, **284–288**, 2119 (2000).
[40] F.C. Wellstood, C. Urbina and J. Clarke, *Appl. Phys. Lett.* **54**, 2599 (1989).
[41] E. Abrahams, P.W. Anderson, P.A. Lee and T.V. Ramakrishnan, *Phys. Rev.* **B24**, 6783 (1981).
[42] B. Altshuler, A.G. Aronov and D.E. Khmelnitsky, *J. Phys.* **C15**, 7367 (1982).
[43] D. Goldie, *Private communication* 2001.
[44] M.L. Roukes, M. Freeman, R. Germain *et al. Phys. Rev. Lett.* **55**, 422 (1985).

3

Energy Resolution (FWHM) of Superconducting Detectors

INTRODUCTION

The height or the area of a pulse signal at the output of a preamplifier carry the information about energy E that an incoming photon has deposited in the detector absorber. The signal is always smeared by a substantial amount of fluctuations, which, if not dealt with, prevent accurate measurements of E. There are at least two major sources of fluctuations. The first component, N_s, relates to inherent properties of a chosen primary detector, such as its type, design, material and so on. For instance, in particle detectors, N_s would include the uncertainty in the charge output produced by the statistics of the energy cascade, as it was discussed in Chapter 1. A second fluctuation component, N_e, has an electronic origin. It comprises noise generators present in the primary sensor, in the biasing circuitry and the amplification/processing channel due to the Nyquist thermal excitations, the leakage currents, dielectric losses, quantisation errors, cross-talks and so on. The minimisation of the spectral density of the electronic noise N_e is certainly desirable, but one has to bear in mind that N_e on its own does not reflect the quality of a detector system. What we are looking for is the reduction in the electronic noise relative to the signal amplitude or, in other words, the maximisation of the signal-to-noise ratio

$$\eta = \frac{P_s}{N_e B} \tag{3.1}$$

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin
 ISBN: 0-470-85759-5

which defines the ultimate accuracy of E measurements. In the equation (3.1), P_s is the signal power, N_e is the product of the spectral density of the electronic noise and B is the bandwidth of the electronic channel through which signal and the noise power propogate.

The fact that the signal-to-noise ratio includes the bandwidth of the nucleonic channel implies the necessity to seek a compromise between the count rate capability (pulse throughput) and the energy resolution of the detector systems. The lowest η will be observed at the output of the preamplification unit. Driven by stringent rise time requirements, preamplifiers often must have the largest possible gain-bandwidth product. In order to improve the accuracy of measurements, system designers apply a combined hardware–software optimum filtration, which narrows the bandwidth with the least possible effect on the signal pulse.

Apart from minimising B, filtration units have to shape pulse signals in such a way that enables us to perform measurements of the pulse maximum without introducing a substantial dynamic error. Figure 3.1 illustrates the statement. An infinite cusp-like pulse shape (Figure 3.1a) provides arguably the best signal-to-noise ratio [1]. Yet, in practice, optimised pulses tend to have a flat top (Figure 3.1b). The duration of the flat top must be in excess of the detector charge collection time in order to eliminate the ballistic deficit variations. This is particularly important for superconductor detectors with large absorbers whose signal rise time depends strongly on the interaction position. An additional constraint factor influencing the filtration is set up by the input count rate

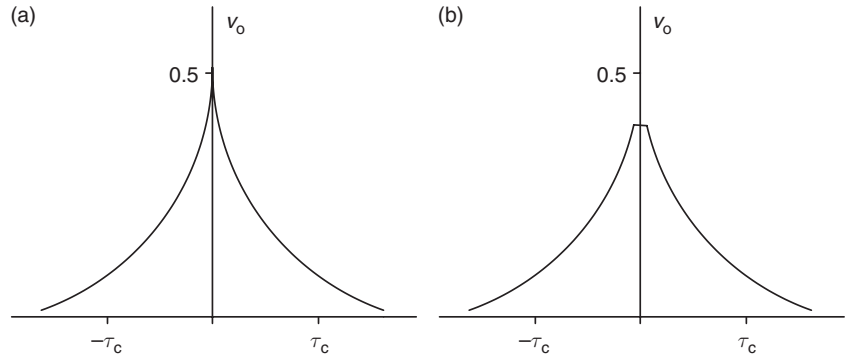


Figure 3.1 Infinite cusp-like pulse shapes with a sharp maximum (a) and a flat top (b). The duration of the flat top must be in excess of the detector charge collection time in order to eliminate the ballistic deficit variations. This is particularly important for superconductor detectors with large absorbers whose signal rise time depends strongly on the interaction position

and pile-up considerations. It requires the overall processing time spent per event to be reduced as the rate of incoming events increases.

The trade off between the throughput and the resolution of spectometers often results in a conclusion that the best signal-to-noise ratio is to be achieved with pulse shapes other than, say, semigaussian. Often these shapes cannot be provided with practical analogue filters of reasonable complexity. Hence, the modern analytical instrumentation tends to incorporate digital filtration units. The digital filter deals with the digitised output signal of the preamplifier. It is nothing but a mathematical algorithm processing data either in the frequency domain, based on the Fourier transform, or in the time domain using weighting functions. Digital filters are truly noise free. As long as the sampling rate is twice as large as the noise bandwidth, digital processing is capable to provide the best possible signal-to-noise ratio. Many optimal shapes like a trapezoid or a quasicusp of finite duration, being in the past of a purely theoretical interest, have become common choices in the modern advanced systems.

In this chapter, we will discuss the noise analysis and the optimum filtration in the nucleonic channels designed for STJ detectors and superconductor microcalorimeters. The first two sections cover linear invariant system analysed in the frequency domain, whereas the next two sections brief on the time-variant processor for which the energy resolution is derived in the time domain. Time-variant processors are rarely used in the modern commercial detector systems, but up to now they exhibit the best energy resolution compared to their time-invariant counterparts.

3.1 SIGNAL-TO-NOISE RATIO, EQUIVALENT NOISE CHARGE AND NOISE LINEWIDTH OF SPECTROMETERS: GENERAL FORMULATIONS

The dynamics of a linear system can adequately be described either in the time domain using the overall impulse response $h(t)$ or in the frequency domain with the help of the transfer function $H(\omega)$. Both the domains relate to each other via the Parseval theorem

$$\int_{-\infty}^{\infty} h^2(t)\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(j\omega)|^2 \mathrm{d}\omega \tag{3.1.1}$$

suggesting that the results should not depend on whether the analysis has been done fully or partially in either of them.

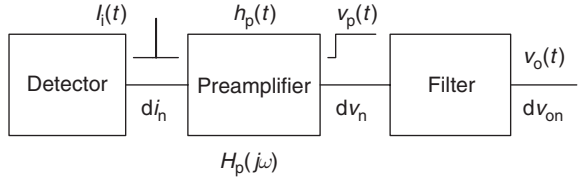


Figure 3.2 Schematic representation of the nucleonic channel of an X-ray spectrometer. A detector generates a short current pulse $I_i(t)$ in response to an incoming photon event. The transconductance preamplifier transforms the current pulse into a voltage signal, $v_p(t)$. This signal is superimposed on the voltage noise dv_n. Both the components pass through the pulse-shaping filter with a transfer function $H_f(j\omega)$ maximising the ratio between them

An example of the nucleonic channel of an X-ray spectrometer is shown schematically in Figure 3.2. A detector generates a short current pulse $I_i(t)$ in response to an incoming photon event. The transconductance preamplifier transforms the current pulse into a voltage signal, $v_p(t)$. This signal is superimposed on the voltage noise dv_n. Both the components pass through the pulse-shaping filter with a transfer function $H_f(j\omega)$ maximising the ratio between them.

The complete transfer function of the channel is given by

$$H(j\omega) = H_p(j\omega)H_f(j\omega) \tag{3.1.2}$$

which corresponds to the overall impulse response expressed through the convolution of individual impulse response functions $h_p(t)$ and $h_f(t)$ as

$$h(t) = h_p(t) * h_f(t) \tag{3.1.3}$$

Now we will make some general remarks on propagation of the signal and noise through the nucleonic channel.

3.1.1 Signal

The output voltage of the pulse-shaping filter in response to $I_i(t)$ can be calculated using the expression

$$v_o(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I_i(j\omega)H(j\omega)\exp(j\omega t)d\omega \tag{3.1.4}$$

in the frequency domain or

$$v_o(t) = \int_{-\infty}^{\infty} I_i(\tau)h(t-\tau)\mathrm{d}\tau \tag{3.1.5}$$

in the time domain. In order to evaluate the signal-to-noise ratio, we are primarily interested in the maximum value of the output voltage, which is defined as

$$v_o|_{max} = \mathrm{Re}\left[\frac{1}{\pi}\int_0^{\infty} I_i(j\omega)H(j\omega)\exp(j\omega t)\mathrm{d}\omega\right]|_{max} \tag{3.1.6}$$

3.1.2 Noise

The noise propagation can also be analysed in both the domains. The mean square voltage or the power at the output of the pulse-shaping circuit is given by

$$V_{on}^2 = \int_0^{\infty} S_i(\omega)|H(j\omega)|^2\mathrm{d}\omega \tag{3.1.7}$$

where $S_i(\omega)$ is the one-sided power spectral density of the current noise $\mathrm{d}i_n$ at the input of the preamplifier.

$$S_i(\omega) \equiv |\mathrm{DPSD}_i(\omega)| + |\mathrm{DPSD}_i(-\omega)| \quad 0 \leq \omega \leq \infty \tag{3.1.8}$$

where $\mathrm{DPSD}_i(\omega)$ is the double-sided power spectral density.

For spectral density of the real noise, $S_i(\omega) = 2|\mathrm{DPSD}_i(\omega)|$. Further, in this book, we will use the one-sided power spectral density.

In the time domain, the mean square voltage is derived from

$$V_{on}^2 = \int_{-\infty}^{\infty} \mathrm{d}\tau R_n(\tau) \int_{-\infty}^{\infty} \mathrm{d}t \cdot h(t) \cdot h(t+\tau) \tag{3.1.9}$$

with $R_n(\tau)$ being the autocorrelation function of the noise source.

Equations (3.1.6) and (3.1.9) allow us to write down the definition for the sought signal-to-noise ratio in the following form

$$\vartheta^2 = \frac{\upsilon_{\mathrm{o}}^2|_{\max}}{V_{\mathrm{on}}^2} \tag{3.1.10}$$

Practical spectrometers are normally characterised with the Equivalent Noise Charge (ENC) rather than the signal-to-noise ratio. ENC is a derivative of υ and interpreted as an amount of charge at the detector output signal, which leads to a value of the signal-to-noise ratio equal to 1.

Where a peak in the pulse height spectrum can be approximated by the Gaussian distribution function, the ENC directly relates to the energy resolution at the FWHM as follows

$$\mathrm{NLW(FWHM)} = 2.355\frac{\mathrm{ENC_e}}{e}\varepsilon \tag{3.1.11}$$

where an electron charge $e = 1.6 \times 10^{-19}$ C and ε is an average energy required to produce a quasiparticle in the detector absorber.

A total energy resolution of a detector system can be found as a quadrature sum of the noise linewidth and an intrinsic energy resolution of the primary detector in the form

$$\mathrm{FWHM}^2 = \mathrm{FWHM}_{\mathrm{i}}^2 + \mathrm{NLW}^2\mathrm{(FWHM)} \tag{3.1.12}$$

The equation (3.1.12) assumes that the intrinsic fluctuations caused by the energy cascade and the electronic noise are not correlated. This may not always be true for superconducting detectors.

A generic transfer function of any type of preamplifiers, which have been discussed in the previous chapter, is given by

$$H_{\mathrm{p}}(j\omega) = \frac{V_{\mathrm{p}}(j\omega)}{I_{\mathrm{i}}(j(j\omega))} \tag{3.1.13}$$

where $V_{\mathrm{p}}(j\omega)$ and $I_{\mathrm{i}}(j\omega)$ represent the Fourier transforms of the $\upsilon_{\mathrm{p}}(t)$ and $I_{\mathrm{i}}(t)$ respectively. The transfer function is commonly derived from the response of a circuit to the δ-function input. Taking this into account, the equation (3.1.13) expands to a form given by

$$H_{\mathrm{p}}(j\omega) = \frac{A(j\omega)}{j\omega(C_{\mathrm{i}} - C_{\mathrm{f}}) + \mathrm{Y_f}[1 - A(j\omega)]} \tag{3.1.14}$$

where C_i is the total capacitance across the input terminal of the preamplifier including parasitic capacitance, C_f is the feedback capacitance and Y_f is the transmittance of the feedback circuit. If we assume that the open loop gain is large, $A(\omega) \gg 1$, within the bandwidth of the pulse-shaping filter, the equation (3.1.14) can be represented in the simplified form as

$$H_p(j\omega) = \frac{1}{G_f + j\omega C_f} \tag{3.1.15}$$

In the preamplifiers, employing the pulse reset the conductivity $G_f = 0$. In the case of the continuous reset mode of operation, Y_f is given by

$$Y_f = \frac{1}{R_f} + j\omega C_f \tag{3.1.16}$$

with R_f being the discharge resistor in the feedback circuitry parallel to the C_f or resistance of the forward-biased p–n junction of the front-JFET. In the latter case, R_f becomes a function of the voltage across the primary detector. The impulse response of the continuous reset pre-amplifiers can be approximated by the exponent function as

$$h_p(t) = \frac{1}{C_f}\exp\left(-\frac{t}{\tau_f}\right) \quad t > 0 \tag{3.1.17}$$

where $\tau_f = R_f C_f$ and $G_f = \frac{1}{R_f}$.

The best signal-to-noise ratio is achieved in preamplifiers operating in the reset mode with $G_f = 0$. In this mode, the transfer function of an ideal preamplifier is given by $H_p(j\omega) = \frac{1}{j\omega C_f}$.

In the time domain, the output voltage v_p in response to incoming photons–phonons takes the shape of a staircase of non-decaying steps. In the case of employing an STJ detector, the amplitude of each individual step is equal to $V_{step} = \frac{Q}{C_f}$. Here Q is the charge generated by the tunnel junction in response to an incident photon–phonon.

The time domain analysis deals with real signals in the real time. Therefore, this technique is often preferred when a general theoretical analysis has to be done in the analytical form. Electronic engineers, however, normally perform practical calculations in the frequency domain, as the transfer functions are also needed at the circuit design stage to locate and optimise positions of poles and zeroes. The latter approach becomes particularly indispensable when we have to deal with

the non-stationary fluctuations such as $1/f$ or $1/f^2$ noise. Their time representation is approximate and may take a very complex form.

So far, we have discussed the continuous time systems, which operate with signals represented in the purely analogue form. It is very useful for gaining a general knowledge about detector systems and getting a feel for signal processing techniques. As we mentioned in the introduction to this chapter, modern spectrometers tend to incorporate mostly digital pulse shapers. Thus, the whole pulse propagation channel is divided into two parts: (1) analogue – the detector, the preamplifier and pre-filter; and (2) digital – the digital signal processor with an Analogue-to-Digital Converter at the heart of the system. The digital part of the spectrometer operates with sampled measurement data called discrete-time signals. The discrete-time signal is termed "a signal", independent variables of which take only discrete values defined at discrete times. In other words, in discrete-time signals the amplitude and time are represented as sequences of numbers. Digital signal processing deals with these sequences in order to filter out noise components as much as possible and then represent the sampled data in a convenient meaningful form.

The relationship between signal representations in the analogue and digital form is regulated by the sampling theorem. According to it, continuous signal $x(t)$, bandwidth limited to the Nyquist critical frequency f_c, will be completely determined in its discrete form if it is sampled at $2f_c$ rate or higher, that is,

$$x(t) = \Delta_{\text{samp}} \sum_{n=-\infty}^{\infty} x_n \frac{\sin[2\pi f_c(t - n\Delta_{\text{samp}})]}{\pi(t - n\Delta_{\text{samp}})} \tag{3.1.18}$$

where $\Delta_{\text{samp}} \leq \frac{1}{2f_c}$ is the sampling time interval.

The sampling theorem suggests that the digital processing unit preserves the entire information content of the preamplifier signal by sampling it at a rate equal to twice the maximum frequency of the preamplifier bandwidth. Sometimes preamplifiers may have a bandwidth exceeding the critical Nyquist frequency. In this case, the requirement of the sampling theorem must be met with a specially designed prefilter unit. Otherwise, the discrete sampling distorts the signal through the phenomena called "aliasing". The aliasing is the false translation of the high-frequency part of the signal spectrum outside f_c back into the frequency range below f_c.

Of other multiple topics of the information theory, we will discuss briefly only the frequency domain representation of discrete-time systems

and signals. It is useful in searching the optimum frequency response or impulse function of a digital filter using the computer simulation tools.

Figure 3.3 illustrates the propagation of the signal sequence $x(n)$ through the time-invariant circuit units with digitised impulse functions $h_1(n)$ and $h_2(n)$. The output sequence of the whole system is derived from the convolution of $x(n)$ with $h(n)$ as

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) \tag{3.1.19}$$

where $h(n) = h_1(n) * h_2(n)$. Suppose that $x(n)$, $h(n)$ and $y(n)$ are numerical sequences defined at the discrete-time moments $t_n = n\Delta_{\text{samp}}$ with $n = 0, 1, 2, \ldots, N-1$, we will try to find the digital Fourier transforms for $x(n)$ and $h(n)$ at discrete frequency values

$$f_k \equiv \frac{k}{N\Delta_{\text{samp}}} \tag{3.1.20}$$

with $k = -\frac{N}{2}, \ldots, \frac{N}{2}$. The extreme values of k in the expression (3.1.19) correspond exactly to the lower and upper limits of the Nyquist critical frequency range. The discrete Fourier transform for the impulse response function h is given by [2]

$$H(f_l) = \int_{-\infty}^{\infty} h(t)e^{2\pi j f_l t}\mathrm{d}t \approx \sum_{k=0}^{N-1} h_k e^{2\pi j f_l t_k}\Delta_{\text{samp}} = \Delta_{\text{samp}} \sum h_k e^{2\pi j k n/N} \tag{3.1.21}$$

The Fourier transform of the input signal sequence $x(n)$ can also be found from the equation (3.1.21) after substituting h by x.

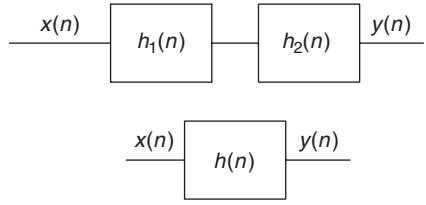


Figure 3.3 The propagation of the signal sequence $x(n)$ through the time-invariant circuit units with digitised impulse functions $h_1(n)$ and $h_2(n)$. The output sequence of the whole system is derived from the convolution of $x(n)$ with $h(n)$

The frequency components of an output sequence can be expressed by the product of $X_l H_l$ such that

$$Y(f_l) = X(f_l)H(f_l) \tag{3.1.22}$$

The inverse discrete Fourier transform can recover the sequence $y(n)$ exactly from Y_l

$$y_l = \frac{1}{N}\sum_{n=0}^{N-1} Y_n e^{-2\pi j \ln/N} \tag{3.1.23}$$

The direct and inverse Fourier transforms enable us to formulate the discrete form of the Parseval theorem in the following form

$$\sum_{n=0}^{N-1} |h_n|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |H_k|^2 \tag{3.1.24}$$

3.2 SIGNAL-TO-NOISE RATIO, ENC, ENERGY RESOLUTION AT FWHM OF TUNNEL JUNCTIONS

3.2.1 The Tunnel Junction Coupled to a JFET Transconductance Amplifier

In this section, we will quantify the signal-to-noise ratio, ϑ, ENC and FWHM of a tunnel junction operating in conjunction with a JFET transconductance preamplifier. For the time being, we will neglect effects of the pulse pile-up, the ballistic deficit, the base line instability and the external interference. We also will assume that the system works in the pulse-reset mode described earlier in section 2.1.

The output signal of the preamplifier looks like a staircase of non-decaying steps, as shown in Figure 3.4. A pulse processing unit subtracts the height of a previous step from the output signal so that it deals with a single step at a time with an amplitude given by

$$v_i = \frac{Q_i}{C_F} \tag{3.2.1}$$

where Q_i is the charge signal of the detector and C_F is the feedback capacitance of the charge-sensitive preamplifier.

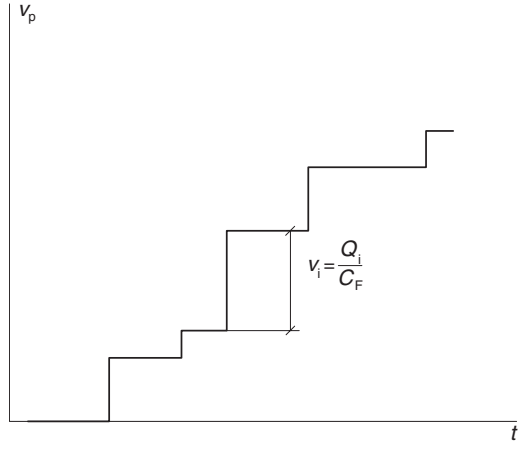


Figure 3.4 The output signal of the preamplifier operating in the pulse-reset mode. With no feedback resistor, the voltage takes a form of a staircase of non-decaying steps. A pulse processing unit subtracts the height of a previous step from the output signal so that it deals with a single step at a time. In the figure, Q_i is the charge signal of the detector, C_F is the feedback capacitance of the charge-sensitive preamplifier

An equivalent circuit diagram of the detector–preamplifier pair is depicted in Figure 3.5. In this figure, the STJ detector is represented explicitly by the signal current source, I_s, the dynamic resistance R_d and a generator of the shot noise current

$$di_2^2 = \frac{1}{2\pi} 2I_d e d\omega \tag{3.2.2}$$

with I_d being the leakage current of the STJ. The total capacitance, C_i, across the preamplifier input terminals includes the capacitance of the STJ, C_d, the input capacitance of the preamplifier, C_p, and the stray

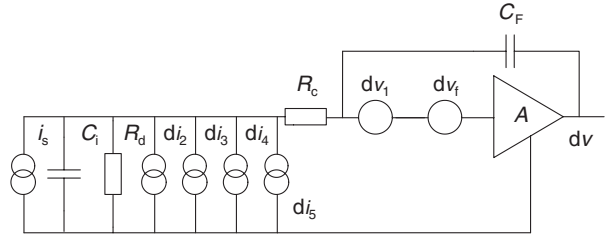


Figure 3.5 An equivalent circuit diagram of the detector–preamplifier pair. All components of the circuit are explained in text

capacitance contributed by a connecting cable and packaging, C_{stray}, in the following form

$$C_{\text{i}} = C_{\text{d}} + C_{\text{p}} + C_{\text{stray}} \tag{3.2.3}$$

In what follows, we will denote other current and voltage generators shown in the equivalent circuit diagram:

– the shot noise due to the leakage current, I_{g}, of the JFET preamplifier

$$\text{d}i_3^2 = \frac{1}{2\pi} 2I_{\text{g}} e \text{d}\omega \tag{3.2.4}$$

– the current noise induced by the channel fluctuation current through the gate-source capacitance, C_{gs}, of the input-JFET

$$\text{d}i_4^2 = \frac{1}{2\pi} \frac{k_{\text{b}} T \omega^2 G_{\text{gs}}^2}{g_{\text{m}}} \text{d}\omega \tag{3.2.5}$$

– the Johnson noise of the cable resistance between the STJ and preamplifier

$$\text{d}i_5^2 = \frac{1}{2\pi} 4k_{\text{b}} T R_{\text{c}} \omega^2 C_{\text{dst}} \text{d}\omega \tag{3.2.6}$$

where $C_{\text{dst}} = C_{\text{d}} + C_{\text{stray}}$;

– the input voltage noise of the JFET preamplifier

$$\text{d}v_1^2 = \frac{4}{3\pi g_{\text{m}}} k_{\text{b}} T \text{d}\omega + \text{d}v_2^2 \tag{3.2.7}$$

with $\text{d}v_2$ taking into account the white noise contributed by other components of the preamplifier in addition to the input-JFET. Some JFETs show the voltage noise with a spectral density in excess of the figure given by $\frac{4}{3\pi g_{\text{m}}} k_{\text{b}} T$. In this case, it should be taken from the transistor specification sheet or direct measurements of the transconductance;

– $1/f$ voltage noise component

$$\text{d}v_{\text{f}}^2 = \frac{A_{\text{f}}}{\omega} \text{d}\omega \tag{3.2.8}$$

where a factor A_f has been defined in equation (2.2.14).

In order to find the voltage noise at the output of the charge-sensitive preamplifier we will do some further transformations of the circuit, shown in Figure 3.5. First, we will group all current sources into one equivalent noise generator, di_p. In the literature, it is often called a parallel noise

$$di_p^2 = di_2^2 + di_3^2 + di_4^2 + di_5^2 \tag{3.2.9}$$

Second, a serious noise generator includes the white and $1/f$ voltage noise components

$$di_s^2 = (dv_1^2 + dv_f^2)\omega^2(C_i + C_F)^2 \tag{3.2.10}$$

These two noise generators form a simplified equivalent circuit diagram, which is presented in Figure 3.6. Now, the output voltage noise can easily be derived in the following form

$$dv_n^2 = \frac{A^2}{[C_i + C_F(1+A)]^2}(di_p^2 + di_s^2) \tag{3.2.11}$$

Substituting all expressions starting from (3.2.2) to (3.2.11) into the equation (3.1.7) we derive the sought mean square voltage noise at the output of the pulse-shaping filter as

$$V_{on}^2 = \int_0^\infty \left\{ \frac{1}{2\pi}\left[\frac{k_b T C_{gs}^2}{g_m C_F^2} + \frac{8}{3}\frac{k_b T}{g_m}\left(1 + \frac{C_i}{C_F}\right)^2 + 4k_b T_c \frac{C_{dst}^2}{C_F^2}\right] + \frac{2e(I_d + I_g)}{2\pi\omega^2 C_F^2} + \frac{A_f}{\omega}\left(1 + \frac{C_i}{C_F}\right)^2 \right\} |H_f(j\omega)|^2 d\omega \tag{3.2.12}$$

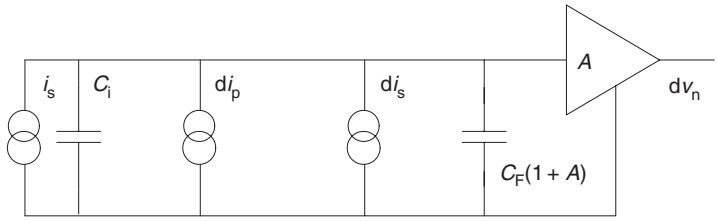


Figure 3.6 An equivalent circuit diagram of the detector–preamplifier pair in which all noise components are classified into two types: parallel and series noise

The noise components in expression (3.2.12) fall into three major categories. Taking this into account it can be re-written in a more readable form as

$$V_{\text{on}}^2 = a^2 I_{\text{a}} + b^2 I_{\text{b}} + c^2 I_{\text{c}} \tag{3.2.13}$$

If the induced noise $\mathrm{d}i_4$ and the Johnson noise of a cable resistance can be neglected, factors *a, b* and *c* represent spectral densities of the series, parallel and 1/*f* noise components, respectively, which are given by

$$a^2 \approx S_{vl} \tag{3.2.14}$$

$$b^2 = \frac{2e}{2\pi C_F^2}(I_{\text{d}} + I_{\text{g}}) \tag{3.2.15}$$

$$c^2 = A_{\text{f}}\left(1 + \frac{C_i}{C_{\text{F}}}\right)^2 \tag{3.2.16}$$

and I_{a}, I_{b} and I_{c} are respective noise indexes, which are entirely defined by the transfer function of the pulse-shaping filter as

$$I_{\text{a}} = \int_0^\infty \left|H_{\text{f}}(j\omega)\right|^2 \mathrm{d}\omega \tag{3.2.17}$$

$$I_{\text{b}} = \int_0^\infty \frac{\left|H(j\omega)\right|^2}{\omega^2}\mathrm{d}\omega \tag{3.2.18}$$

$$I_{\text{c}} = \int_0^\infty \frac{\left|H(j\omega)\right|^2}{\omega}\mathrm{d}\omega \tag{3.2.19}$$

Example 3.1

In this example, we will analyse the propagation of the signal and noise through the preamplifier and a CR–RC type pulse-shaping filter. Apart from the educational purpose, results of the analysis are valuable on their own due to the fact that the noise indexes of the symmetric CR–RC filter coincide with those of the optimum infinite cusp. Therefore,

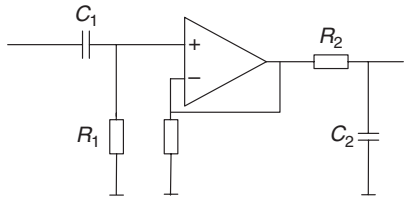


Figure 3.7 A schematic circuit diagram of the CR–RC pulse-shaping network

straight after the example, we will be able to derive the best signal-to-noise ratio, which can be achieved in an ideal spectrometer.

One possible realisation of the CR–RC filter is shown in Figure 3.7. Its transfer function is given by

$$H_{\mathrm{f}}(j\omega) = \frac{j\omega\tau_1}{(1+j\omega\tau_1)(1+j\omega\tau_2)} \tag{3.2.20}$$

where $\tau_1 = R_1C_1$ and $\tau_2 = R_2C_2$ are the time constants.

Signal propagation

The response of the filter to a voltage step with amplitude $V_{\mathrm{i}} = \frac{Q_{\mathrm{i}}}{C_{\mathrm{F}}}$ can be written in the following form

$$V_{\mathrm{o}}(j\omega) = \frac{Q_{\mathrm{i}}}{j\omega C_{\mathrm{F}}} \frac{j\omega\tau_1}{(1+j\omega\tau_1)(1+j\omega\tau_2)} \tag{3.2.21}$$

in the frequency domain. Taking the inverse Fourier transform of the equation (3.2.21), the same response can be represented in the time domain as follows

$$v_o(t) = \frac{Q_i}{C_{\mathrm{F}}} \frac{\tau_1}{\tau_1 - \tau_2} \left(e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}} \right) \quad \text{for } t \geq 0 \tag{3.2.22}$$

The function $v_0(t)$ has a singularity when the time constants of both the halves of the filter are equal, that is, $\tau = \tau_1 = \tau_2$. This case is particularly interesting as at $\tau = \tau_1 = \tau_2$ the overall network exhibits the best signal-to-noise ratio. Using the l'Hopital rule the equation (3.1.22) can be re-written in the form

$$v_o(t) = \frac{Q_{\mathrm{i}}}{C_{\mathrm{F}}} \frac{t}{\tau} e^{-\frac{t}{\tau}} \tag{3.2.23}$$

with the maximum value

$$v_o|_{\max} = \frac{Q_i}{C_F}\frac{1}{\exp(1)} \tag{3.2.24}$$

Using expressions (3.2.17) to (3.2.19), we derive noise indexes for CR–RC filter as

$$I_a = \int_0^\infty \left|H_f(j\omega)\right| d\omega = \frac{\pi}{4\tau} \tag{3.2.25}$$

$$I_b = \int_0^\infty \frac{\left|H_f(j\omega)\right|}{\omega^2} d\omega = \frac{\pi}{4}\tau \tag{3.2.26}$$

$$I_c = \int_0^\infty \frac{\left|H_f(j\omega)\right|}{\omega} d\omega = \frac{1}{2} \tag{3.2.27}$$

Voltage noise

Following equation (3.2.13), the mean square voltage noise at the output of the CR–RC network is given by

$$V_n^2 = \frac{\pi}{4\tau}a^2 + \frac{\pi\tau}{4}b^2 + \frac{1}{2}c^2 \tag{3.2.28}$$

Now we can write down the sought signal-to-noise ratio:

$$\vartheta^2 = \frac{v_o^2|_{\max}}{V_n^2} = \frac{Q_i^2}{\exp^2 C_F^2}\left(\frac{\pi}{4\tau}a^2 + \frac{\pi\tau}{4}b^2 + \frac{1}{2}c^2\right)^{-1} \tag{3.2.29}$$

the equivalent noise charge:

$$(\mathrm{ENC})^2 = \exp^2 C_F^2\left(\frac{\pi}{4\tau}a^2 + \frac{\pi\tau}{4}b^2 + \frac{1}{2}c^2\right) \tag{3.2.30}$$

and, finally, the electronic noise linewidth at FWHM:

$$\begin{aligned}\mathrm{NLW(FWHM)} &= 2.355\frac{\mathrm{ENC}}{e}\varepsilon \\ &= \frac{2.355\varepsilon}{2}\exp\cdot \mathrm{C_F}\sqrt{\frac{\pi}{4\tau}a^2+\frac{\pi\tau}{4}b^2+\frac{1}{2}c^2}\end{aligned} \tag{3.2.31}$$

According to [3], an ideal spectrometer should be incorporating a pulse-shaping network which would transform a step signal into an infinite cusp. Such a system cannot be realised in practice, but is often used as a reference for other types of filters. Whereas the noise indexes I_a, I_b and I_c of the infinite cusp are the same as those of CR–RC network, the maximum value of the voltage signal is larger. It is given by (see, for instance, [3], p. 135)

$$v_o|_{\max} = 0.5\frac{Q_i}{C_F} \tag{3.2.32}$$

Then, the figures of merit of the optimal detector system take the following form

$$v_{\mathrm{cusp}} = \frac{Q_i}{2C_F}\left(\frac{\pi}{4\tau}a^2+\frac{\pi\tau}{4}b^2+\frac{1}{2}c^2\right)^{-\frac{1}{2}} \tag{3.2.33}$$

$$\mathrm{ENC_{cusp}} = 2C_F\left(\frac{\pi}{4\tau}a^2+\frac{\pi\tau}{4}b^2+\frac{1}{2}c^2\right)^{\frac{1}{2}} \tag{3.2.34}$$

$$\mathrm{NLW_{cusp}(FWHM)} = \frac{2.355}{e}\varepsilon 2C_F\left(\frac{\pi}{4\tau}a^2+\frac{\pi\tau}{4}b^2+\frac{1}{2}c^2\right)^{\frac{1}{2}} \tag{3.2.35}$$

A time constant $\tau = \tau_c = \frac{a}{b}$ is sometimes termed the "noise corner" time constant. In fact, if we neglect the $1/f$ noise component, τ_c represents an integration time, at which the contributions of the series noise, a, and the parallel noise, b, are equal and both the CR–RC and cusp-like pulse-shaping networks exhibit their best figures of the NLW.

In other frequently utilised pulse shapes, like, for instance, a trapezoid, an optimum integration time does not differ a lot from τ_c. Therefore, the latter can always serve as a benchmark for initial setting of a pulse processing unit.

Another important conclusion, following from (3.2.35), is the fact that an STJ with a relatively large leakage current can still demonstrate an energy resolution close to that of a higher quality device at shorter integration times. Long integration times are not desirable anyway in detector systems intended for microanalysis, especially in mapping applications. Besides, a marginal improvement of the energy resolution can be lost in dealing with severe pile-up problems. Still the development of low leakage devices is vital in order to reduce the tailing and artefacts at low energies. In the majority applications, the maximisation of the peak-to-background ratio on the final energy histogram is the ultimate goal which must be achieved through the improvement of the FWHM as much as the reduction in the background noise.

Noise indexes given by expressions (3.2.25) and (3.2.26) depend on the time constant of a pulse-shaping network. Some scientific publications utilise their dimensionless analogues. In order to derive them, we need to normalise the timescale by the time constant of the pulse-shaping network, that is, time takes the dimensionless form $\frac{t}{\tau}$. In the frequency domain, this is equivalent to changing the frequency scale from ω to $\omega\tau$. With this modification, the mean square voltage at the output of the pulse-shaping network becomes

$$V_n^2 = \int_0^\infty \left(a^2 + \frac{b^2}{\omega^2} + \frac{c^2}{\omega}\right)|H(j\omega\tau)|^2 \mathrm{d}\omega = \frac{a^2}{\tau} I'_a + b^2\tau I'_b + c^2 I'_c \quad (3.2.36)$$

where

$$I'_a = \int_0^\infty \left|H(j\omega\tau)\right|^2 \mathrm{d}(\omega\tau) \quad (3.2.37)$$

$$I'_b = \int_0^\infty \frac{\left|H(j\omega\tau)\right|^2}{\omega^2\tau^2} \mathrm{d}(\omega\tau) \quad (3.2.38)$$

$$I'_c = \int_0^\infty \frac{\left|H(j\omega\tau)\right|^2}{\omega\tau} \mathrm{d}(\omega\tau) \quad (3.2.39)$$

These dimensionless noise indexes do not carry any longer the information about integrating time. They are entirely defined by the shape of the

transfer function or, equally, the step response of the pulse-shaping network. Thus, each type of pulse-shaping networks is characterised by a unique set of I'_a, I'_b and I'_c. For instance, all CR–RC filters have the same indexes given by $I'_a = \frac{\pi}{4}$, $I'_b = \frac{\pi}{4}$ and $I'_c = \frac{1}{2}$. Indexes of other filters can be found, for instance, in [3].

Let us assume that a maximum value of the signal voltage is defined as

$$v_o|_{\max} = \kappa \frac{Q_i}{C_F} \tag{3.2.40}$$

Substituting the equation (3.2.40) into the definition of the ENC, given in section 3.1, we can derive its general expression suitable for all types of pulse-shaping networks, which includes dimensionless noise indexes. It is given by

$$\mathrm{ENC} = \frac{C_F}{\kappa Q_i}\left(\frac{a^2}{\tau} I'_a + b^2 \tau I'_b + c^2 I'_c\right)^{1/2} \tag{3.2.41}$$

So far, we assumed the output signal of the preamplifier to be a perfect step. This means that the collection time of the quasiparticles in an STJ, t_c, is negligibly small. In practice, however, t_c can be a substantial fraction of the pulse-shaping time. Besides, in STJ detectors with large area absorbers the collection time depends on a position of a spot where the interaction with an incoming photon actually occurs. The finite rise time of a pulse obviously modifies the response of a pulse-shaping filter. In fact, it always reduces the maximum value of the signal at the output of the filter, $v_o|_{\max}$. This reduction is quantified by a figure called the "ballistic deficit".

K. Hatch [4] calculated the ballistic deficit for different types of pulse-shaping networks. He found that networks forming a pulse shape with a flat top of the duration longer than the collection time of the detector show the least loss in the peak height. For instance, with a collection time $t_c = 0.2\tau_c$, triangular shape leads to a ballistic deficit of 8.02%, whereas this figure for a truncated triangular (trapezoid) shape reduces to a figure of 3.87%.

The ballistic deficit appears as a systematic error if incoming events do not affect the output current shape and timing of the detector in the energy range of interest. With a proper correction technique, it will result in a slight reduction of the signal-to-noise ratio. Variable shapes and timing produce a random ballistic deficit, which has a double negative effect. It broadens peaks as well as shifts their actual positions

on the energy diagram. The latter makes the elemental analysis more difficult and in some cases even impossible if the overlapping between peaks exceeds a critical recognition value [5].

To summarise, the ballistic deficit, if not dealt with properly, may seriously limit the resolving power of a practical detector system even if detector itself is capable to yield an excellent energy resolution in response to a collimated monochromatic irradiation. Despite that, there has been very little consideration on the problem in the literature with regard to superconducting detectors, in which this effect is very pronounced.

3.2.2 Signal-to-Noise Ratio, ENC and Energy Resolution FWHM of STJ Sensors Coupled to SQUID Current Amplifiers

An equivalent circuit diagram of an STJ detector coupled to a SQUID current amplifier is presented in Figure 3.8. The capacitance, C_d, and the dynamic resistance, R_d, simulate the detector whose shot noise is given by

$$di_d^2 = \frac{1}{2\pi} 2 I_d e \, d\omega \tag{3.2.42}$$

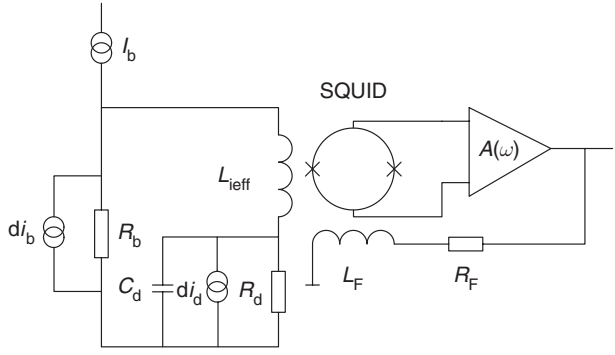


Figure 3.8 An equivalent circuit diagram of an STJ detector coupled to a SQUID current amplifier. The capacitance, C_d, and the dynamic resistance, R_d, simulate the detector, whereas R_b is the bias resistor. The effective input inductance of the SQUID, L_{ieff}, and the detector capacitance form an L–C resonant circuit. If not critically damped, the circuit exhibits the oscillatory behaviour with a long settling time. Therefore, the bias resistor serves a very important additional role of the damping load

R_b is the bias resistor with a thermal noise generator

$$di_b^2 = \frac{4k_bT}{2\pi R_b}d\omega \tag{3.2.43}$$

The effective input inductance of the SQUID, L_{ieff}, and the detector capacitance form an L–C resonant circuit. If not critically damped, the circuit exhibits the oscillatory behaviour with a long settling time. Possible overshots in the preamplifier output voltage can also contribute to the ballistic deficit. High-quality STJs normally demonstrate a large dynamic resistance, $R_d > 10\,k\Omega$. Its finite value certainly helps to partially damp the circuit, but not even near to the critical degree required. Therefore, the bias resistor serves in the circuitry a very important additional role of the damping load. In what follows, we will try to derive a minimum value of R_b, which is adequate for eliminating the oscillatory component from the output signal of the SQUID current amplifier.

The current flowing in the input circuit is described with a second-order differential equation of the form

$$\frac{d^2i}{dt^2} + 2\zeta\omega_n\frac{di}{dt} + \omega_n^2 i = 0 \tag{3.2.44}$$

where $\zeta = \frac{R_d}{2}\sqrt{\frac{C_d}{L_{ieff}}}$ is the damping ratio, $\omega_n^2 = \frac{1}{L_{ieff}C_d}$ is termed the undamped natural frequency of the resonator. Assuming that the damping ratio $0 \le \zeta \le 1$, the characteristic equation derived for the differential equation (3.2.44) is given by [6]

$$x^2 + 2\zeta\omega_n x + \omega_n^2 = \left(x + \zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}\right)\left(x + \zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}\right) = 0 \tag{3.2.45}$$

Its roots are

$$x_1 = -\zeta\omega_n + j\omega_n\sqrt{1-\zeta^2} \equiv -\upsilon + j\omega_d \tag{3.2.46}$$

$$x_2 = -\zeta\omega_n - j\omega_n\sqrt{1-\zeta^2} \equiv -\upsilon - j\omega_d \tag{3.2.47}$$

where

$$\upsilon \equiv \zeta\omega_n \tag{3.2.48}$$

is called the damping factor and $\omega_d \equiv \omega_n\sqrt{1-\zeta^2}$ is the damped natural frequency.

Let us now assume that an incoming event charges the detector capacitance to its initial value of Q_i. This generates a step voltage $V_i = \frac{Q_i}{C_d}$. By applying the initial condition $I(0) = 0$ at $t = 0$ and the functional dependence between the charge across C_d and the current flowing in the input circuit $i(t) = \frac{dQ(t)}{dt}$, we can derive the solutions of equation (3.2.44) for three cases as follows:

1. the underdamped input circuit:

$$i(t) = Q_i\left(\frac{\omega_n^2 e^{-\upsilon t}}{\omega_d}\sin(\omega_d t)\right) \tag{3.2.49}$$

$$Q(t) = Q_i\frac{\omega_n}{\omega_d}e^{-\upsilon t}\sin(\omega_d t + \phi) \tag{3.2.50}$$

with $I(0)$ being an initial value and $\phi = \frac{1}{\tan(\frac{\omega_d}{\upsilon})}$ the phase of the current;

2. the critically damped input circuit:

$$i(t) = Q_i\alpha^2 t e^{-\upsilon t} \tag{3.2.51}$$

$$Q(t) = Q_i(1+\upsilon t)e^{-\upsilon t} \tag{3.2.52}$$

3. the overdamped input circuit:

$$i(t) = Q_i\frac{\omega_n^2}{\sqrt{\upsilon^2-\omega_n^2}}e^{-\upsilon t}\sinh\left(\sqrt{\upsilon^2-\omega_n^2}t\right) \tag{3.2.53}$$

$$Q(t) = Q_i\frac{\omega_n}{\sqrt{\upsilon^2-\omega_n^2}}e^{-\upsilon t}\sinh\left(\sqrt{\upsilon^2-\omega_n^2}t+\phi\right) \tag{3.2.54}$$

The critical damping condition, $\zeta = 1$, ensures the fastest non-oscillatory time response of the circuit. With numerical data $C_d = 100\,\text{pF}$, $L_{ieff} = 260\,\text{nH}$ and $R_d = 10\,\text{k}\Omega$, this leads to a critical bias/damping resistance

$$R_{bcrit} \approx \frac{2\zeta\sqrt{L_{ieff}}}{\sqrt{C_d}} \approx 80\,\Omega \tag{3.2.55}$$

A value of the bias resistor derived from the equation (3.2.55) meets the condition of the voltage bias $R_b \ll R_d$. Thus, in this particular case, the bias resistor adequately provides the voltage bias and the fastest response of the detector.

If a SQUID current amplifier is coupled to a detector with a larger capacitance or to a detector array that contains a number of junctions connected in parallel, a larger bias resistor starts adversely affecting the current sensitivity of the current amplifier. Mears *et al.* [7] suggested to damp the resonance with an additional resistor attached in parallel to the SQUID input inductance. This resistor, however, adds an extra thermal noise $di_R^2 = \frac{4k_bT}{2\pi R}d\omega$. For instance, a 10-Ω resistor generates approximately $1\,\frac{\text{pA}}{\sqrt{\text{Hz}}}$ at a temperature of 0.2 K. This is four times larger than the current noise of a state-of-art SQUID operating at $T = 4.2\,\text{K}$ (Chapter 2).

The signal-to-noise ratio of the STJ–SQUID current amplifier pair is defined in exactly the same way as for the STJ coupled to a transconductance amplifier

$$\upsilon^2 = \frac{\upsilon_o^2|_{\max}}{V_n^2}$$

with exceptions that the input signal cannot be described by a step function and some noise components have truncated spectral densities due to inertiality of the input circuit.

3.2.2.1 Signal

The voltage at the output of the SQUID current amplifier is given by

$$\upsilon_p(t) \approx Q_i \upsilon^2 t R_a \left(e^{-\frac{t}{\tau_i}}\right) \tag{3.2.56}$$

Here R_a is the current-to-voltage transfer function of the amplifier and $\tau_i = \frac{1}{\upsilon}$ is the time constant of the input circuit.

The voltage signal at the output of a pulse-shaping filter in response to the signal given by (3.2.56) can be found using either a convolution theorem in the time domain or the inverse Fourier transform in the frequency domain (Section 3.1) so that

$$\upsilon_o|_{\max} = F^{-1}[V_p(j\omega)H_f(j\omega)]|_{\max} \tag{3.2.57}$$

3.2.2.2 Noise

The system includes three major electronic noise components. These are

1. the current noise of the SQUID

$$\mathrm{d}i_{\mathrm{s}}^2 = \frac{S_I}{2\pi}\mathrm{d}\omega \tag{3.2.58}$$

2. the current shot noise of the detector

$$\mathrm{d}i_{\mathrm{d}}^2 = \frac{2eI_{\mathrm{d}}}{2\pi}\frac{1}{(1+\omega^2\tau_{\mathrm{i}}^2)}\mathrm{d}\omega \tag{3.2.59}$$

3. the current thermal noise of the bias resistor

$$\mathrm{d}i_{\mathrm{b}}^2 = \frac{4k_{\mathrm{b}}T}{2\pi R}\frac{1}{(1+\omega^2\tau_{\mathrm{i}}^2)}\mathrm{d}\omega \tag{3.2.60}$$

Thus, the rms voltage at the output of the filter takes the following form

$$V_{\mathrm{rms}}^2 = \frac{1}{2\pi}R_{\mathrm{a}}^2\int_0^\infty \left\{S_I + \frac{1}{1+\omega^2\tau_{\mathrm{i}}^2}\left[2eI_{\mathrm{d}} + \frac{4k_{\mathrm{b}}T}{R_{\mathrm{b}}}\right]\right\}\left|H_{\mathrm{f}}(j\omega)\right|^2\mathrm{d}\omega \tag{3.2.61}$$

The equation (3.2.61) shows that the SQUID read-out transforms the current noise components into a series type of noise. This is opposite to the case with the transconductance preamplifier where these noise sources formed the parallel type.

In practice, the SQUID current noise can be much larger compared to the shot noise produced by the leakage current of a high-quality STJ and the thermal noise of the bias resistor, that is, $\mathrm{d}i_{\mathrm{s}} \gg \mathrm{d}i_{\mathrm{d}}$ and $\mathrm{d}i_{\mathrm{s}} \gg \mathrm{d}i_{\mathrm{b}}$. Thus, the expression for rms voltage reduces to

$$V_{\mathrm{n}}^2 \approx \frac{1}{2\pi}\int_0^\infty S_I\left|H_{\mathrm{f}}(j\omega)\right|^2 = a^2 I_{\mathrm{a}} \tag{3.2.62}$$

If an amount of charge generated by an STJ detector in response to a deposited mean energy E is not known, the energy resolution FWHM can be directly derived from

$$\mathrm{NLW(FWHM)} = E\frac{2.355}{\vartheta} \tag{3.2.63}$$

In this equation, we assumed that the peak at E on the energy histogram has the Gaussian shape. If a ratio of the mean energy E to the mean charge Q_{I} is available, ENC follows from

$$\mathrm{ENC} = \frac{Q_i}{\vartheta} \tag{3.2.64}$$

Example 3.2

In this exercise, we will evaluate an electronic noise linewidth of a Ta-STJ detector coupled to a SQUID current preamplifier. Let us assume that the detector system incorporates a CR–RC pulse-shaping filter and is exposed to a monochromatic X-ray source ^{55}Fe with a mean energy $E = 5.9\,\mathrm{keV}$. A good quality Ta-STJ with an area of $100 \times 100\,\mu\mathrm{m}^2$ has the following static parameters:

- a capacitance 100 pF;
- a leakage current 0.1 nA; and
- a dynamic resistance 10 kΩ.

The spectral density of a single-stage SQUID current preamplifier is $(S_{\mathrm{I}})^{1/2} = 0.8\ \frac{\mathrm{pA}}{\sqrt{\mathrm{Hz}}}$.

Signal

According to [8], a typical output current produced by such devices in response to 5.9 keV photons has an amplitude approximately $I_{\mathrm{o}}\big|_{\max} = 1\,\mu\mathrm{A}$ and a $1/e$ decay time constant τ_{r} between 1 and 30 μs. Here we choose $\tau_{\mathrm{r}} = 10\,\mu\mathrm{s}$. The signal voltage at the output of the CR–RC pulse-shaping filter is given by

$$v_o(t) = R_{\mathrm{a}} g I_{\mathrm{o}}\big|_{\max} \left[\frac{d}{(d-g)^2} (e^{-gt} - e^{dt}) - \frac{g}{d-g} t e^{-gt} \right] \tag{3.2.65}$$

where R_{a} is the voltage-to-current transfer function of the SQUID current preamplifier, $\mathrm{d} = \frac{1}{\tau_{\mathrm{r}}}$ and $g = \frac{1}{\tau_{\mathrm{F}}}$ with τ_{F} being the time constant of the filter (integration time).

With $\tau_{\mathrm{f}} = 20\,\mu\mathrm{s}$, $v_{\mathrm{o}}\big|_{\max} = 1.75 \times 10^{-7} R_{\mathrm{a}}\,\mathrm{V}$.

Noise

A mean square voltage noise at the output of the pulse-shaping network is given by

$$V_{\mathrm{n}}^2 \approx R_{\mathrm{a}}^2 \frac{\pi}{4} \frac{S_I}{2\pi\tau_{\mathrm{F}}} \approx (6.325 \times 10^{-11} R_{\mathrm{a}})^2 \ \mathrm{V}^2.$$

These give a signal-to-noise ratio $\upsilon = 3.95 \times 10^3$, which leads to an electronic noise linewidth NLW(FWHM) = 3.52 eV with a single-stage SQUID operating at $T = 4.2$ K (the worst case).

The best energy resolution of an STJ with the SQUID read-out was reported to be approximately 29 eV at 5.9 keV [9]. This means that the current noise of the SQUID amplifier would contribute a negligibly small fraction into the total FWHM.

Certainly, STJ detectors are capable to demonstrate a much better energy resolution. For instance, FWHM = 13 eV (at 5.9 keV) was measured in [10] with a future potential to bring the figure down to sub-10 eV. In this case, the current noise of the SQUID amplifier should be reduced. An implementation of a double-stage SQUID with $(S_{\mathrm{I}})^{1/2} = 0.25 \ \frac{\mathrm{pA}}{\sqrt{\mathrm{Hz}}}$ and a CR–RC filter with an integration time $\tau_{\mathrm{F}} = 100 \ \mu\mathrm{s}$ results in NLW(FWHM) = 0.5 eV. An ULT-SQUID with a trapezoidal pulse-shaping network can further improve it.

3.3 NOISE EQUIVALENT POWER, ENERGY RESOLUTION OF SUPERCONDUCTOR MICROCALORIMETERS

As we mentioned in Chapter 1, the electro-thermal feedback aims at maintaining a constant temperature of the TES. This means that the TES responds primarily to a momentary thermal power $\delta P(j\omega)$ applied to the device. For instance, in a microcalorimeter operating in the voltage-biased mode, $\delta P(j\omega)$ generates proportional current variations, $\delta I(j\omega)$, in the input circuitry, further amplified and shaped by the nucleonic channel. The energy of an incident photon is an integral characteristics of $\delta P(j\omega)$ over the pulse duration, τ, and can be derived from the following expression

$$E = \frac{1}{|W_{\mathrm{I}}|} \int_0^{\tau} \delta I(t) \mathrm{d}t \tag{3.3.1}$$

where $W_{\mathrm{I}} = \frac{\mathrm{d}I(t)}{\mathrm{d}P(t)}$ is the responsivity of the microcalorimeter.

Provided that the detector response is linear and the pulse shape does not change with deposited energy, the pulse height, $v_o|_{max}$, can still carry a relative information about E. $v_o|_{max}$, itself depends on the design of the sensor, an operating temperature and at least three time-constants, including the effective time constant of the microcalorimeter, τ_{eff}, the time constant of the SQUID input circuit, τ_I, and, finally, the integration time, τ. Here we assume that the 2π-gain-bandwidth product of the SQUID current preamplifier is much larger compared to all characteristic frequencies of the detector, that is, $\omega_m \gg \frac{1}{\tau_{eff}}$, $\omega_m \gg \frac{1}{\tau_i}$ and $\omega_m \gg \frac{1}{\tau}$.

The minimum 1/exp(1) rise time, Δt_r, of the current pulse δI is ultimately defined by the time constant of the input circuit, τ_i. If, however, an incoming photon interacts with a relatively large area absorber at a distance l from the TES, the diffusion time of electrons given by [11]

$$t_{max} \approx \frac{1}{3}\frac{l^2}{D} \tag{3.3.2}$$

can make a substantial contribution to Δt_r or even dominate it.

The 1/exp(1) decay time of $v_o(t)$ is entirely governed by the effective time constant of the microcalorimeter, which is defined as

$$\tau_{eff} \approx \frac{-\tau_0}{1+A} \tag{3.3.3}$$

where $\tau_0 = \frac{G}{C}$ is the time constant of the microcalorimeter without the electro-thermal feedback and $A = \frac{P_b \alpha}{GT}$ is the DC loop amplification of the electro-thermal feedback.

The mean square voltage noise at the output of the SQUID current preamplifier is given by [12]

$$V_n^2 = R_a^2 \frac{1}{2\pi}\int_0^\infty (S_{PF} + S_T + S_{ISQUID})|H_f(j\omega)|^2 d\omega \tag{3.3.4}$$

where

$$S_{PF} = |W_I|^2 \gamma 4 k_b T^2 G \tag{3.3.5}$$

is the spectral density of the current fluctuation noise resulted from the discrete nature of the random propagation of the energy carriers between the TES and the heat sink,

$$S_{\mathrm{T}} = \frac{4k_{\mathrm{b}}T}{R_{\mathrm{b}} + R_{\mathrm{d}}(\omega)} \tag{3.3.6}$$

is a spectral density of the Johnson noise generated by the TES and the bias resistor, and $S_{\mathrm{ISQUID}} = \frac{S_\Phi}{M_{\mathrm{i}}}$ denotes a spectral density of the current noise of the SQUID preamplifier.

A dimensionless factor $\gamma \approx \frac{n}{2n+1}$ in equation (3.3.5) is called the power flow coefficient. It depends on the temperature gradient along the thermal link between the TES and the heat sink. At mK temperatures, the factor n in the power flow coefficient ranges between $3 < n < 4$. $R_{\mathrm{d}} \approx R(\frac{\tau_0}{\tau_{\mathrm{eff}}})(\frac{1+\omega^2\tau^2}{1+\omega^2\tau_0^2})$ is the dynamic resistance of the TES, whereas R is its static value in the middle of the dynamic range.

The responsivity of the microcalorimeter can be expressed via its other measurable parameters. According to [12],

$$W_{\mathrm{I}} = -\frac{1}{V_{\mathrm{b}}}\frac{A(\omega)}{1+A(\omega)} \approx -\frac{1}{V_{\mathrm{b}}}\frac{1}{1+j\omega\tau_{\mathrm{eff}}} \tag{3.3.7}$$

with V_{b} being the voltage bias.

Since the sensor responds primarily to the power of the input signal, the NEP is a natural figure to characterise its performance. The NEP is defined as the signal power required to obtain a unity signal-to-noise ratio in the presence of the detector noise. In other words,

$$\mathrm{NEP}^2 = \frac{1}{|W_{\mathrm{I}}|^2}(S_{\mathrm{PF}} + S_{\mathrm{T}} + S_{\mathrm{ISQUID}}) \tag{3.3.8}$$

The NEP characterises fully the detector system as a bolometer intended for optical measurements. Some X-ray applications, such as the microanalysis, operate with ENC and the energy resolution. If an energy of an incident event is recovered from the pulse height, the signal-to-noise ratio of the system incorporating an optimum filter is given by [13]

$$\vartheta^2 = \frac{v_{\mathrm{o}}^2|_{\max}}{V_{\mathrm{n}}^2} = \frac{1}{2\pi}\int_0^\infty \frac{|I_{\mathrm{s}}(j\omega)|^2}{S_{\mathrm{PF}}(\omega) + S_{\mathrm{T}}(\omega) + S_{\mathrm{ISQUID}}(\omega)}\,\mathrm{d}\omega \tag{3.3.9}$$

where $I_{\mathrm{s}}(j\omega)$ is the Fourier transform of the signal current pulse. This leads to the energy resolution FWHM

$$\Delta E_{\mathrm{FWHM}} = E\frac{2.355}{\vartheta} \tag{3.3.10}$$

If the total current noise in the input circuitry can be assumed white, the optimum energy resolution becomes directly proportional to the NEP [14]

$$\Delta E_{\mathrm{FWHM}}^{\mathrm{opt}} \approx 2.355\mathrm{NEP}(\omega = 0)\sqrt{\tau_{\mathrm{eff}}} \tag{3.3.11}$$

The optimum filtration is essential in order to achieve an energy resolution predicted by equations (3.3.10) and (3.3.11). The transfer function of an optimum filter can be written down in an explicit form as follows [15,16]

$$H_{\mathrm{f}}^2 = k\left[\frac{I_{\mathrm{s}}^*(\omega)}{S_\Sigma(\omega)}\right]e^{-j\omega\tau} \tag{3.3.12}$$

where k is the normalising factor, $I_{\mathrm{s}}^*(\omega)$ is the complex conjugate of the Fourier transform of the signal waveshape and $S_\Sigma(\omega)$ is the spectral density of the total current noise in the input circuitry, τ means the integration time.

Exercise 3.1

In this exercise, we will evaluate an energy resolution of a high-speed microcalorimeter described in [17]. We assume that the microcalorimeter is exposed to a monochromatic X-ray source ^{55}Fe(5.9 keV). All necessary parameters of the microcalorimeter are summarised in the following table

Normal resistance of TES	$R_{\mathrm{n}} = 0.9\,\Omega$
Thermal conductance	$G = 10\,\mathrm{nWK}^{-1}$
Heat capacitance	$C = 24\,\mathrm{pJK}^{-1}$
Intrinsic time constant	$\tau_0 = 2.4\,\mathrm{ms}$
TES sensitivity	$\alpha = 900$
Operating temperature	$T = 350\,\mathrm{mK}$
Bias voltage	$V_{\mathrm{b}} = 15.6\,\mu\mathrm{V}$

Experimentally measured output signal of the TES can be approximated by

$$i_{\mathrm{s}}(t) = I_0\mathrm{e}^{-\frac{t}{\tau_{\mathrm{eff}}}} \tag{3.3.13}$$

with the maximum current $I_0 = 4\,\mu\text{A}$ and a 1/exp(1) decay time constant $\tau_{\text{eff}} = 15\,\mu\text{s}$. In the frequency domain, it takes the form

$$I_{\text{s}}(j\omega) = I_0 \frac{\tau_{\text{eff}}}{1 + j\omega\tau_{\text{eff}}} \tag{3.3.14}$$

According to equation (3.3.9), the signal-to-noise ratio amounts to

$$\begin{aligned}\vartheta^2 &= \frac{1}{2\pi}\int_0^\infty \frac{I_0^2\tau_{\text{eff}}^2}{\frac{1}{V_{\text{b}}^2}\gamma 4k_{\text{b}}T^2G + 4k_{\text{b}}\frac{T}{R}(1+\omega^2\tau_0^2) + (0.8\times 10^{-12})^2(1+\omega^2\tau_{\text{eff}}^2)} \\ &\approx 0.769\times 10^6\end{aligned}$$

which gives an energy resolution of

$$\Delta E_{\text{FWHM}}^{\text{opt}} = 2.355\frac{5900}{\vartheta} \approx 15.8\,\text{eV}$$

For comparison, other expressions estimate the energy resolution as follows

$$\Delta E_{\text{FWHM}}^{\text{opt}} = 2.355\text{NEP}(0)\sqrt{\tau_{\text{eff}}} \approx 10.4\,\text{eV}$$

and, according to [18],

$$\Delta E_{\text{FWHM}}^{\text{opt}} = 2.355\sqrt{4k_{\text{b}}TC\frac{\sqrt{8n}}{\alpha}} \approx 23\,\text{eV}$$

The measured energy resolution $\Delta E = 20\,\text{eV}$.

Exercise 3.2

In this exercise, we will calculate the energy resolution of the same microcalorimeter with a simple symmetric CR–RC pulse-shaping filter with RC time constant equal to the effective decay time, τ_{eff}.

The signal at the output of the pulse-shaping filter can be found from the inverse Fourier transform as follows

$$v_{\text{o}}(t) = F^{-1}[I_{\text{s}}(j\omega)\cdot H_{\text{f}}(j\omega)] = R_{\text{a}}\frac{I_0 t}{\tau_{\text{eff}}}\left(1 - \frac{1}{2\tau_{\text{eff}}}t\right)e^{-\frac{t}{\tau_{\text{eff}}}} \tag{3.3.15}$$

Its maximum value is $v_o\big|_{\max} \approx 0.92\times 10^{-6}R_{\text{a}}$.

The mean square voltage noise is

$$V_n^2 = \frac{1}{2\pi}\int_0^\infty \left[\frac{\gamma 4k_b T^2 G}{V_b^2(1+\omega^2\tau_{eff}^2)} + \frac{4k_b T}{R}\left(\frac{\tau_{eff}}{\tau_0}\right)^2\left(\frac{1+\omega^2\tau_0^2}{1+\omega^2\tau^2}\right) + (0.8\times 10^{-12})^2\right] \frac{\omega^2\tau_{eff}^2}{(1+\omega^2\tau_{eff}^2)^2}\mathrm{d}\omega \approx 5.63\times 10^{-19} R_a^2$$

Now we can estimate the signal-to-noise ratio

$$\vartheta = \frac{v_o|_{max}}{V_n} \approx 1.23\times 10^3$$

which yields an energy resolution $\Delta E = \frac{2.36\times 5900}{\vartheta} \approx 11.3\,\mathrm{eV}$. Use of optimum pulse-shaping filters can substantially improve the energy resolution of TES spectrometers.

3.4 DYNAMICS AND NOISE OF TIME-VARIANT DETECTOR SYSTEMS

The analogue time-variant pulse processing technique is evolved from a simple CR–RC pulse-shaping network, in which RC low-pass filter was replaced by a true time integrator [19]. A functional block diagram of the system is presented schematically in Figure 3.9. The invention was to introduce an analogue gate between the CR differentiator and the

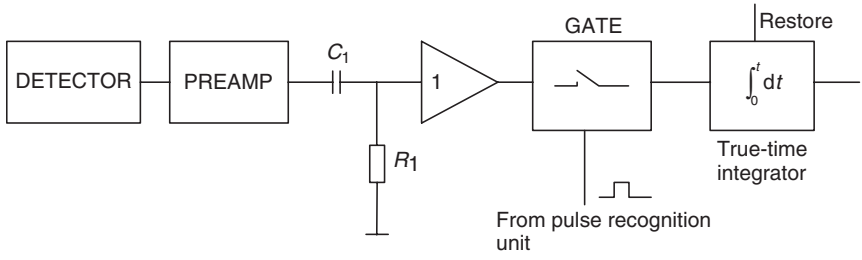


Figure 3.9 A functional block diagram of the time-variant processor. The analogue time-variant pulse processing technique evolved from a simple CR–RC pulse-shaping network, in which RC low-pass filter was replaced by a true time integrator [19]. The invention was to introduce an analogue gate between the CR differentiator and the integrator. The gate opens only after a recognised signal pulse has arrived for the duration of the measurement time τ

integrator. The gate opens only after a recognised signal pulse has arrived for the duration of the measurement time τ. Since the integrator becomes active only during τ, the time-variant pulse processor can discriminate better against noise pulses arising prior to the signal. This yields a noise-to-signal ratio of only 1.6% greater than that of the infinite cusp-response nucleonic channel. For comparison, this figure for a widely utilised trapezoidal pulse-shaping network is approximately a factor of 1.075 times larger compared to the noise-to-signal ratio delivered by the infinite cusp with the same peaking time.

The analogue time-variant processing technique competes successfully with modern digital counterparts in terms of the adaptive filtration. The optimisation of the integration time with respect to the input rate is as easy and takes as long time as the adjustment of the time constants in digital filters. This comes as a serious benefit against digital spectrometers in the combination with the fact that such a system does not place stringent requirements on the performance of ADC, although a possibility of achieving a better signal-to-noise ratio has also led to the development of a digital version of the time-variant processor (see, for instance, [20]).

Time-variant processors keep at present a lower profile compared to their time-invariant counterparts. However, they are very competitive in some applications, particularly those requiring minimum dead time. Therefore, it is still useful to know how to analyse the dynamics and noise properties of these systems employing cryogenic primary detectors, which we will discuss in this section.

Any type of time-variant nucleonic channels represents a special case of adaptive systems with a variable structure. Their performance can be evaluated either in a modified frequency domain, using discrete z-transform, or in the time domain, based on the Campbell theorem. A general introduction to the time domain analysis concerning nucleonic channels incorporating a charge sensitive preamplifier was done in [21,22]. According to it, the electronic noise has a discrete nature caused by the motion of separate charged particles. F. Goulding classified the white noise into two types, which are

1. step-like noise; and
2. delta-like noise.

Electrons, participating in the current flow into the input circuit of a preamplifier, produce the step-like noise. Each electron charges the feedback capacitance, C_F, so that the output voltage is, in fact, the

superposition of the random time distribution of $\frac{e}{C_F}$ voltage steps, with e being the electron charge. Contrary to that, the voltage noise at the preamplifier input does not charge the feedback capacitance. Therefore, it appears at the output terminal as a random sequence of short impulses or delta-functions at a mean rate n per second with a normalised amplitude.

After passing a pulse-shaping network, each individual discrete noise event results in a residual at the measurement time τ. Obviously, residuals of the step-like noise are proportional to the step-function of the shaper, $F(t)$, whereas residuals of the delta-like noise scale with the impulse response, $F'(t) = \frac{dF(t)}{dt}$. According to the Campbell theorem, the mean square effect of fluctuations at the peak time τ is obtained by summing the mean square values of the noise residuals for all steps and δ-pulses preceding the measurement time. Thus, the noise indexes in the time domain are defined as

$$I'_a = \int_0^\infty [F'(t)]^2 dt \tag{3.4.1}$$

for the series delta-like noise and

$$I'_b = \int_0^\infty [F(t)]^2 dt \tag{3.4.2}$$

for the parallel step-like noise.

Normalising the timescale by a factor $x = \frac{t}{\tau}$, we derive the integration time sensitive noise indexes as follows

$$I_a = \frac{1}{\tau} \int_0^\infty [F'(x)]^2 dx \tag{3.4.3}$$

$$I_b = \tau \int_0^\infty [F(x)]^2 dx \tag{3.4.4}$$

Representation of noise sources as the superposition of individual random steps and delta-functions helps to better understand the physics hidden behind expressions (3.4.1)–(3.4.4). However, strictly speaking, it was not necessary to actually derive those expressions in this way.

Mathematically, the equations (3.4.1)–(3.4.4) follow directly from noise indexes obtained in the frequency domain (3.2.25) and (3.2.26), and (3.2.37) and (3.2.38) respectively, using the Parseval theorem.

An adaptive system with the variable structure may change its transfer function a number of times during the working cycle, which we will set to m. Therefore, it is convenient to divide the cycle time into m intervals and calculate residual noise indexes for each of them. The total noise index amounts to the sum of the individual ones, as long as in each interval the system remains linear.

The working cycle of the time-variant pulse processor itself contains at least two intervals: a first one $[-\infty < t < 0]$ and a second one $[0 < t < \tau]$. Thus, the total noise indexes are given by

$$I'_{a1,b1} = I'_{a1,b1}\Big|_{[-\infty<t<0]} + I'_{a2,b2}\Big|_{[0<t<\tau]} \tag{3.4.5}$$

where

$$I'_{a1} = \int_0^{\tau} F'_1(t)\mathrm{d}t \tag{3.4.6}$$

$$I'_{a2} = \int_{t0}^{\tau} F'_2(t)\mathrm{d}t \tag{3.4.7}$$

$$I'_{b1} = \int_0^{\tau} F_1(t)\mathrm{d}t \tag{3.4.8}$$

$$I'_{b2} = \int_{t0}^{\tau} F_2(t)\mathrm{d}t \tag{3.4.9}$$

with $F_1(t)$ and $F_2(t)$ representing the response functions of the pulse shaper at the first and a second time intervals, respectively, $t0$ is the time moment at which signal pulse is occurred.

Exercise 3.3

In this exercise, we propose to find noise indexes of a time-variant pulse shaper described in [23]. The essence of the shaper schematics is shown in Figure 3.10. As can be seen, the system includes a second integrating

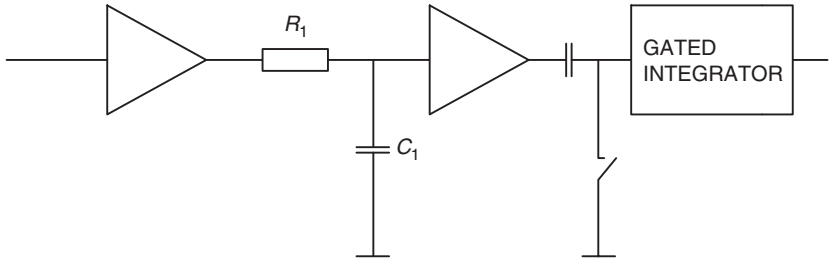


Figure 3.10 A circuit diagram of the pulse shaper of the time-variant processor. The system includes a second integrating RC network in addition to the gated integrator. The pulse response of the combined network follows very closely to a shape of a truncated finite cusp

R_1C_1 circuit. The circuit further modifies the step response of the network in such a way that its shape follows very closely the one of a truncated cusp [16], that is

$$F(t) = \int_0^t \left(1 - e^{-\frac{t}{\tau_1}}\right) \mathrm{d}t = t - \tau_1 \left(1 - e^{-\frac{t}{\tau_1}}\right) \tag{3.4.10}$$

with $\tau_1 = R_1C_1$.

Interested readers can find the detailed analysis of the problem in the Appendix of [16]. Here, we will give the final results of the analysis

$$I_\mathrm{a} = 1.88\tau f^2(\tau)$$

$$I_\mathrm{b} = \frac{0.602}{\tau} f^2(\tau)$$

where $f(\tau)$ is a normalising factor that corresponds to the filter response $F(t)$ at $t = \tau$ for a signal step at $t = 0$.

3.5 SIGNAL-TO-NOISE RATIO OF DETECTOR ARRAYS WITH MULTIPLEXED READ-OUT

Cryogenic systems, providing ultralow temperature and low noise environment for superconductor detectors, have a very limited power dissipation budget. For instance, a typical adiabatic demagnetisation refrigerator can tolerate a heat load of several tens of μW at a base

temperature of 100 mK. This figure raises to approximately 0.4 mW at 220 mK for a ^{3}He single-shot cooler. Within this constraint, it is vital to minimise the amount of wires connecting a multipixel array of devices to the room temperature read-out and processing electronics. The minimisation can be achieved by implementing either a distributed detector array, when a large area absorber is attached to a limited number of sensors, or a multiplexing technique, in which only a part of pixels is addressed at a time.

In this section, we will discuss two types of multiplexing methods. These are:

1. a scheme with a read-out preamplifier successively sampling outputs of a group of sensors one after another; and
2. a matrix read-out of an $M \times N$ pixel array, in which grouped M-row and N-column electrodes are served by continuously attached $M + N$ preamplifiers.

Depending on the type of application, each technique has a different degree effect on the throughput for each individual pixel, the signal-to-noise ratio, detector layout and so on. In what follows, we will try to give comparative evaluation of the dynamics and noise properties of arrays primarily intended for soft X-ray measurements. However, much of the content of the section provides a useful practical guidance for developing large format multiplexed arrays of superconductor detectors, which can be used in a wide range of imaging instruments.

3.5.1 Matrix Read-Out with FET Transconductance Amplifier

The matrix read-out method is illustrated schematically in Figure 3.11. The detector array consists of $M \times N$ pixels. Separate preamplifiers are connected to M top electrodes in each row and N bottom electrodes in each column. In this particular application, the charge sensitive preamplifiers operate in the continuous reset mode. The feedback resistor, R_F, maintains a DC potential difference between inverting and non-inverting inputs close to zero. Thus, the top electrodes can be DC biased by V_b applied to the non-inverting preamplifier input terminal via the signal wire. In a similar way, column preamplifiers maintain quasizero potential on the bottom electrodes.

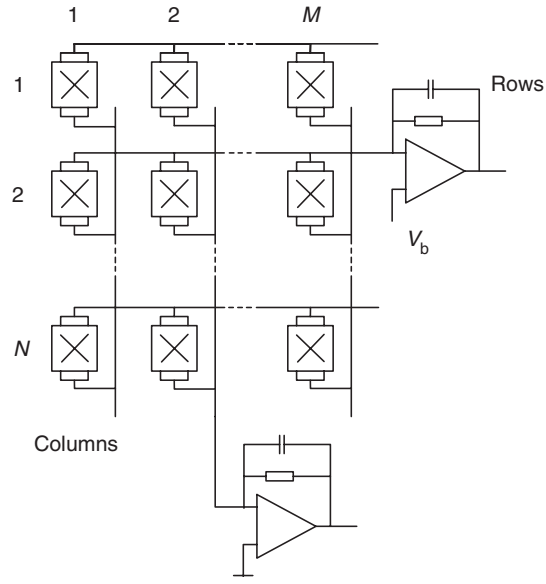


Figure 3.11 Matrix read-out of an array of cryogenic detectors with FET transconductance amplifiers (the technique is reproduced by permission of D. Martin from ESTEC). The detector array consists of $M \times N$ pixels. Separate preamplifiers are connected to M top electrodes in each row and N bottom electrodes in each column. Charge sensitive preamplifiers operate in the continuous reset mode. The feedback resistor, R_F, maintains a DC potential difference between inverting and non-inverting inputs close to zero. Thus, the top electrodes can be DC biased by V_b applied to the non-inverting preamplifier input terminal via the signal wire. In a similar way, column preamplifiers maintain quasizero potential on the bottom electrodes

When an incoming photon interacts with the array, an affected pixel produces a signal in its corresponding row and column. Coincidence measurements between rows and columns help to identify its co-ordinates. This mode of operation reduces the number of wires connecting the array with the room temperature electronics from $M \times N$ required for the parallel interface to just $M + N$.

From the point of view of the bias stability, it is desirable to minimise R_F so that the conditions

$$\frac{R_F}{1 + A(\omega)} \ll \frac{R_d}{M}; \qquad \frac{R_F}{1 + A(\omega)} \ll \frac{R_d}{N} \tag{3.5.1}$$

are met in the whole frequency bandwidth of interest. In this expression, $A(\omega)$ is the amplification factor of operational amplifiers. As we

mentioned in section 2.1.1, the feedback resistor may contribute a substantial portion of the parallel noise proportional to $\frac{4k_bT}{R_F}$. Noise considerations define a minimum tolerable value of R_f so that a reasonable compromise must be found between noise and bias stability conditions given by (3.5.1).

Figure 3.11 shows that the matrix of this type is very convenient to process and operate. For instance, external contact pads need to be wired only to pixels constituting the first and the last rows and columns, whereas other pixels have only connections to their neighbours. This allows the manufacture of chips to be accomplished in fewer layers compared to any other techniques. Thin film wires do not cover the active area of pixels improving the uniformity of their response across the array.

Drawbacks of the arrangement associate with a fact that it does not offer a possibility to control the quality parameters of individual pixels, such as the leakage current or magnetic trap flux and so on. This does not become a serious issue, however, if the manufacturing yield is high and arrays operate in well-controlled conditions.

Figure 3.12 (a, b) presents an equivalent circuit diagram of a pixel, as a part of the array, connected to row and column preamplifiers respectively. According to these diagrams, the input of each preamplifier sees $M(N)$ times increased detector capacitance and only a $\frac{1}{M}(\frac{1}{N})$ fraction of the pixel dynamic resistance. The total capacitances across input terminals of row and column preamplifiers are given by

$$C_{i(row)} = C_p + C_{dstr(raw)}; \qquad C_{i(col)} = C_p + C_{dstr(col)} \tag{3.5.2}$$

$$C_{dstr(raw)} = MC_d + C_{stray}; \qquad C_{dstr(col)} = NC_d + C_{stray} \tag{3.5.3}$$

where C_p is the input capacitance of the preamplifier and C_{stray} is the total capacitance of the stray capacitance.

Spectral densities of the shot noise take the form

$$di^2_{2(row)} = \frac{1}{2\pi} 2MI_g e d\omega; \qquad di^2_{2(col)} = \frac{1}{2\pi} 2NI_g e d\omega \tag{3.5.4}$$

where I_g is the leakage current of individual pixels. The equation (3.5.4) assumes that all junctions have the same value of I_g.

With definitions given by equations (3.5.2)–(3.5.4), we can evaluate the signal-to-noise ratio, ENC and the energy resolution (FWHM) for

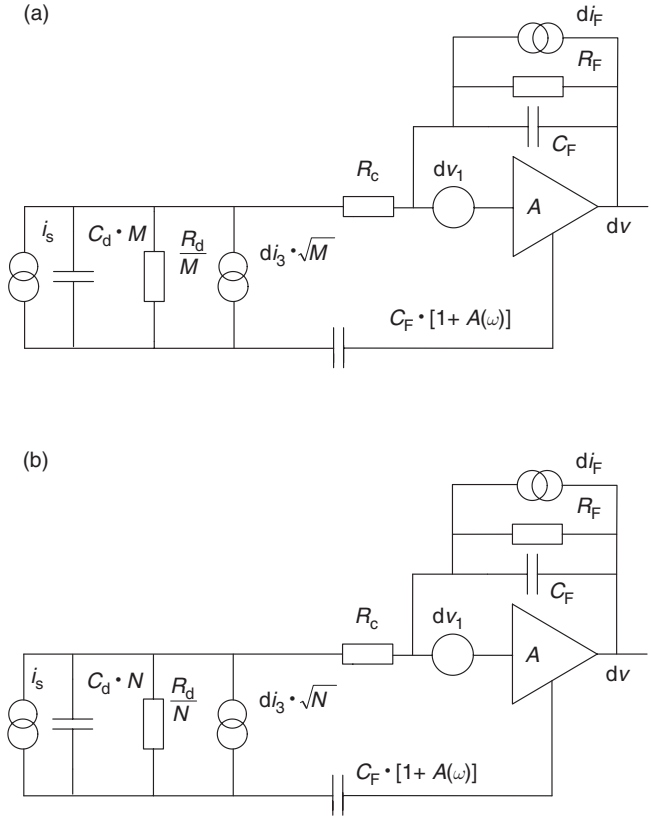


Figure 3.12 An equivalent circuit diagram of a pixel, as a part of the array, connected to row (a) and column (b) preamplifiers respectively. According to these diagrams, the input of each preamplifier sees $M(N)$ times increased detector capacitance and only a $\frac{1}{M}$ $(\frac{1}{N})$ fraction of the pixel dynamic resistance

the matrix read-out using expressions (3.2.29)–(3.2.31). Assuming that C_F and R_F remain constant and taking into account only the major series noise component, the ENC scales with the array dimensions as

$$l_{\text{row}} = \frac{\text{ENC}_{(\text{row})}}{\text{ENC}_{(\text{single-pixel})}} \approx \frac{C_F + MC_d + C_{\text{stray}} + C_p}{C_F + C_d + C_{\text{stray}} + C_p} \approx M \qquad (3.5.5)$$

$$l_{\text{col}} = \frac{\text{ENC}_{(\text{col})}}{\text{ENC}_{(\text{single-pixel})}} \approx \frac{C_F + NC_d + C_{\text{stray}} + C_p}{C_F + C_d + C_{\text{stray}} + C_p} \approx N \qquad (3.5.6)$$

3.5.2 Matrix Read-Out with SQUID Current Amplifiers

A matrix read-out employing SQUID current amplifiers is illustrated schematically in Figure 3.13. Similar to the case with transconductance amplifiers, the SQUID read-out does not require an additional dedicated wire to supply the voltage bias V_b to electrodes. In practice, the input circuitry provides a more stable voltage bias as the loadlines provided with SQUIDs and transconductance amplifiers are in the following proportion

$$R_b + j\omega L_{ieff} \ll \frac{R_F}{1 + A(\omega)} \tag{3.5.7}$$

In order to evaluate the signal-to-noise ratio in the array with the SQUID read-out, we will use an equivalent circuit diagram shown in Figure 3.14. According to it, all junctions in the rows (columns) are connected in parallel so that the input circuit includes the total capacitances: $C_{d\Sigma(row)} = MC_d$ and $C_{d\Sigma(col)} = NC_d$; and the dynamic resistances: $R_{d\Sigma(raw)} = \frac{R_d}{M}$ and $R_{d\Sigma(col)} = \frac{R_d}{N}$. The shot noise is given by equation (3.5.4).

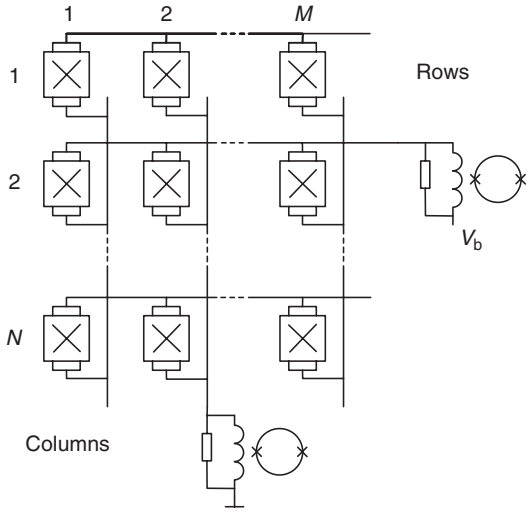


Figure 3.13 A matrix read-out employing SQUID current amplifiers. Similar to the case with transconductance amplifiers, the SQUID read-out does not require additional dedicated wires to supply the voltage bias V_b to electrodes

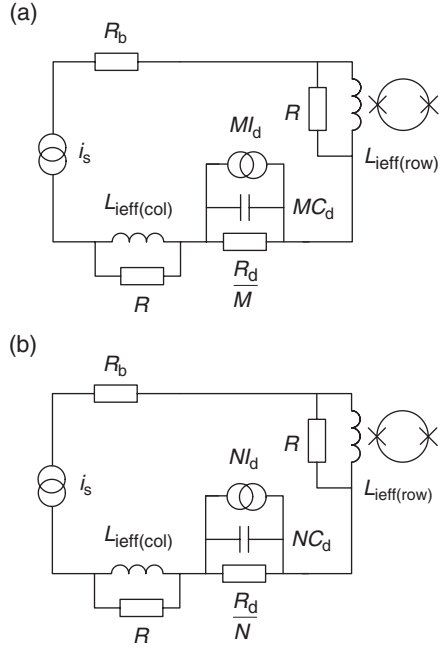


Figure 3.14 An equivalent circuit diagram of a pixel, as a part of the array, connected to row (a) and column (b) SQUID current preamplifiers respectively. C_d, R_d represent the detector capacitance and dynamic resistance. R is the shunt resistor critically damping the $L_{ieff}C_d$ resonance

The suppression of a $L_{ieff}MC_d$ (or $L_{ieff}NC_d$) resonance requires the installation of a dedicated damping resistor across the input coil L_{ieff}. Using for this purpose the bias resistor only would increase its value to a level seriously affecting the voltage bias stability. A large resistor in series with the junctions also limits the current sensitivity of the SQUID current amplifier. With these modifications, the expression (3.2.61) for rms voltage noise at the output of a nucleonic channel connected to the array row takes a new form

$$V_{n(row)}^2 = \frac{1}{2\pi} R_a^2 \int_0^\infty \left(S_I + \frac{2eMI_d}{1+\omega^2\tau_i^2} + \frac{4k_bTM}{R} \right) |H_f(j\omega)|^2 d\omega \qquad (3.5.8)$$

where R is the shunt resistor required to critically damp the single junction $L_{ieff}C_d$ resonance. $V_{n(col)}^2$ can be derived from (3.5.8) by substituting N instead of M.

As we mentioned in Section 3.2, a matrix comprising high-performance junctions and SQUIDs demonstrates the following ratio between major noise contributions:

$$\frac{4k_{\mathrm{b}}TM}{R} \gg S_{\mathrm{I}} \geq 2eNI_{\mathrm{d}} \tag{3.5.9}$$

This yields the scaling factors

$$l_{(\mathrm{raw})}^{\mathrm{SQUID}} \approx \frac{\mathrm{ENC}_{(\mathrm{raw})}}{\mathrm{ENC}_{(\mathrm{single\text{-}junction})}} \approx M^{1/2} \tag{3.5.10}$$

for rows and

$$l_{(\mathrm{col})}^{\mathrm{SQUID}} \approx \frac{\mathrm{ENC}_{(\mathrm{col})}}{\mathrm{ENC}_{(\mathrm{single\text{-}junction})}} \approx N^{1/2} \tag{3.5.11}$$

for columns.

The comparison of expressions (3.5.10) and (3.5.11) with (3.5.5) and (3.5.6) suggests that the SQUID read-out technique becomes competitive against FET transconductance amplifiers as the matrix dimension increases. This is a very positive outcome promising good prospects for applications requiring large format arrays. SQUID current amplifiers with their lower input impedance and closer positioning to the array itself help efficiently combating the cross-talk between junctions as well as the interference problems without introducing any substantial extra heat load on the cryogenics.

3.5.3 Time-Division Multiplexer

The principle of the time-division multiplexer is explained in Figure 3.15. It includes an $M \times N$ sensor array, an $M \times N$ SQUID array, a CMOS multiplexer operating at low temperature and M number of room temperature semiconductor operational amplifiers each combined with an individual bias circuitry. The CMOS multiplexer connects an operational amplifier to one SQUID at a time for the duration of the sampling time t_{s}. At the start of t_{s}, the bias current I_{b} activates a selected SQUID. Once the transient process has dissipated, the system completes

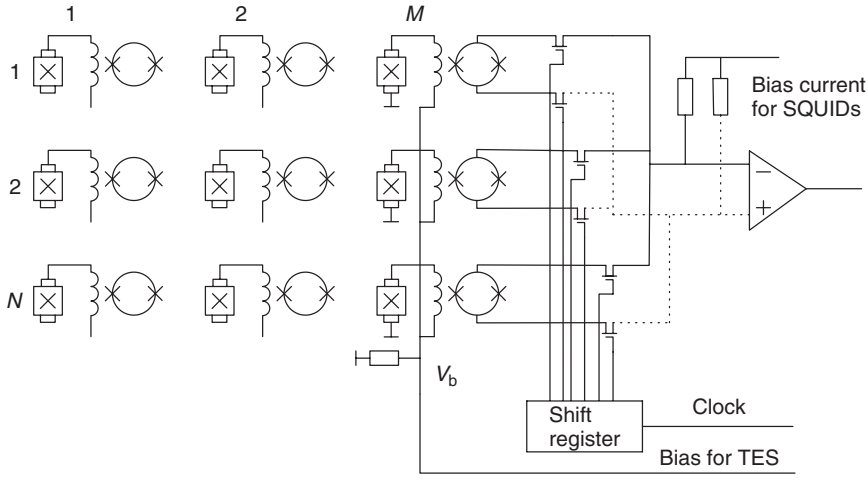


Figure 3.15 A schematic diagram of the time-division multiplexer. It includes an $M \times N$ sensor array, an $M \times N$ SQUID array, a CMOS multiplexer operating at low temperature and M room temperature semiconductor operational amplifiers each combined with an individual bias circuitry. The CMOS multiplexer connects an operational amplifier to one SQUID at a time for the duration of the sampling time t_s. At the start of t_s, the bias current I_b activates a selected SQUID. Once the transient process has dissipated, the system completes configuring a single pixel nucleonic channel, samples its output voltage and switches to a next pixel

configuring a single-pixel nucleonic channel, samples its output voltage and switches to the next pixel.

In order to reduce the cross-talk between channels, the CMOS multiplexer operates in the "break before make" mode. In other words, the clock time includes some "dead" time, t_d, so that the clock frequency is defined as

$$f_{cl} = \frac{1}{t_s + t_d} \tag{3.5.12}$$

The lowest limit of the sampling rate depends on the signal-pulse duration and a number of pixels assigned to one operational amplifier N. In the case of TES microcalorimeter, it depends on the effective time constant of the sensor and is given by

$$f_{cl(min)} = \frac{NK_{min}}{\tau_{eff}} \tag{3.5.13}$$

where K_{min} is the minimum amount of samples needed for the adequate signal pulse restoration.

An upper limit of f_{cl} follows directly from the transient process period, t_{tr}

$$f_{cl(max)} \leq \frac{1}{t_{tr}} \tag{3.5.14}$$

In the small-signal limit, $t_{tr} \approx \frac{5}{f_c}$ with f_c being the gain-bandwidth product of the SQUID current amplifier including the input circuitry.

CMOS and HCMOS transistors in the switched mode of operation practically do not dissipate extra heat. Eight- to sixteen-bit multiplexers were demonstrated to function adequately well down to 1.4 K [24]. Such low-temperature multiplexers enable us to reduce the number of wires to only three per N pixels. The third feedback wire is needed to operate SQUID amplifiers in the flux-locked mode stabilising the bias line.

Dealing with sampling means that we have to comply with the Nyquist theorem in order to prevent the aliasing effect. The output signal of each sensor should be pre-filtered before it reaches the amplification channel. The Nyquist theorem requires a sampling rate of each pixel to be at least twice as much as the noise bandwidth of the filter.

The effective input inductance of the SQUID and the dynamic resistance of TES thermometer form a natural single-pole low-pass filter. Its noise bandwidth is given by

$$f_n = \frac{1}{2\pi}\int_0^\infty \frac{1}{1+\omega^2\tau_i^2}\,d\omega = \frac{R_d}{4L_{ieff}} \tag{3.5.15}$$

The dynamic resistance of the TES as well as effective input inductance may vary with the energy deposited by an incoming photon. Therefore, the Nyquist frequency of individual pixels given by

$$f_{NYC} = \frac{1}{2N(t_s + t_d)} \tag{3.5.16}$$

should be at least twice as large as the maximum noise bandwidth $f_{n(max)}$. The upper and lower limit conditions (3.5.13) and (3.5.14) may require adding some extra inductance, L_{ex}, in series with L_{ieff} in order to match f_n and f_{NYQ}.

Each SQUID amplifier reads out the information from an individual pixel during $\frac{1}{N}$ fraction of the real time duty cycle. Thus, its effective noise bandwidth increases by a factor N. A specified ratio of the current

noise of TES to the current noise of SQUID current amplifier corrected for effective noise bandwidth defines the maximum number of pixels, which can be served by one amplifier, that is

$$N_{\max} \leq \frac{\chi \mathrm{d}i^2_{(\mathrm{TES})}}{\mathrm{d}i^2_{(\mathrm{SQUID})}} \tag{3.5.17}$$

where χ is the margin factor.

For instance, assuming that $\chi = 0.1$, $S_{\mathrm{I(SQUID)}} = 0.8\ \frac{\mathrm{pA}}{\sqrt{\mathrm{Hz}}}$ and the degradation in NEP is less than 1%, then one operational amplifier can multiplex 100 transistor edge sensors described in [17].

One serious limitation associated with the time-division multiplexing technique is the fact that the sampling rate has also a link to the stable operating condition of the TES itself. According to (3.5.13)

$$\tau_{\mathrm{eff}} \geq 4\frac{L_{\mathrm{ieff}}}{R_{\mathrm{d}}} = \frac{1}{f_{\mathrm{n}}} \approx 2Nt_{\mathrm{s}} \tag{3.5.18}$$

With a reasonable sampling time $t_{\mathrm{s}} = 10\,\mu\mathrm{s}$ and $N = 100$, we derive $\tau_{\mathrm{eff}} > 2\,\mathrm{ms}$.

Equations (3.5.14) and (3.5.18) suggest that SQUID current amplifiers with a much faster settling time should be developed to reduce the sampling time, t_{s}, and to use the maximum speed offered by the best superconductor microcalorimeters. This can be partially achieved if the feedback voltage is stored in for each pixel during the whole cycle and utilised as initial values in the next one. In this case, we come to the concept of a SQUID amplifier with the digital feedback loop. The simplest configuration of the loop would contain an ADC, a digital inertial block (PI-controller) and a DAC. Several attempts were undertaken to develop such a system in the past. So far, their noise performance has not been as good as one of analogue amplifiers.

If an application permits the allocation of some time for regular calibration procedure, SQUID amplifiers can be speeded up substantially (several orders of magnitude) by removing the feedback loop. In order to keep an operating point in the middle of the dynamic range, a compensation current should be passed through the feedback inductance during the sampling time. The calibration procedure must store required values of the current for each pixel.

REFERENCES

[1] F.T. Arecchi, G. Cavalleri, E. Gatti and V. Svelto. *Energia Nucleare* **7**, 691 (1960).
[2] W. Press, B. Flannery, S. Teukolsky and W. Vetterling. Numerical Recipes. *The Art of Scientific Computing*. Cambridge University Press. 1989, p. 387.
[3] P. Nicholson. *Nuclear Electronics*. Wiley, 1972, p. 88.
[4] K. Hatch. *IEEE Trans. Nucl. Sci*., NS15, 303 (1968).
[5] P. Statham, In: X-Ray Spectrometry in Electron Beam Instruments, edited by D. Williams, J. Goldstein and D. Newbury. Plenum Press, New York, 1995, p. 101.
[6] J. Di Stefano, A. Stubberud and I. Williams. Feedback and Control Systems, McGraw-Hill, 1990, p. 48.
[7] C. Mears, S. Labov, M. Frank *et al*. High resolution superconductor X-ray spectrometers. Preparing UCRL-JC-125225, 1996.
[8] D. Martin, P. Verhoeve, A. Peacock and D. Goldie. *Proceedings of SPIE* **4008**, 3340 (2000).
[9] M. Frank. *Nucl. Instrum. Methods Phys. Res*. **A444**, 375 (2000).
[10] L. Li, L. Frunzio, C. Wilson *et al*. *J. Appl. Phys*. **90**, 3645 (2001).
[11] D. Goldie. *Private communication*, 1999.
[12] S. Lee, J. Gildmeister, W. Holmes *et al*. *Appl. Optics*, **37**, 3391 (1998), J.C. Mather, *Appl. Optics*, **21**, 1125 (1982).
[13] E. Gatti, M. Sampietro and P. Manfredi, *Instrum. Methods Phys. Res*. **A287**, 513 (1990).
[14] S.H. Moseley, J. Mather and D. McCammon. *J. Appl. Phys*. **56**, 1257 (1984).
[15] V. Radeka. *Instrum. Methods Phys. Res*. **52**, 86 (1967).
[16] J. Llacer. *Instrum. Methods Phys. Res*. **130**, 565 (1975).
[17] T. Morooka, K. Tanaka, A. Nagata *et al*. *Supercond. Sci. Technol*. **15**, 133 (2002).
[18] K. Irwin and B. Cabrera. US Patent 5,641,961 (1997).
[19] K. Kandiah. Active integrators in spectrometry with radiation detectors. Report No. AERE-R 5019, 1966.
[20] Technical Note of Princeton Gamma-Tech, Inc. http://www.pgt.com/DPP2.html.
[21] M. Deighton. *Instrum. Methods Phys. Res*. **58**, 201 (1968).
[22] F. Goulding. *Instrum. Methods Phys. Res*. **100**, 249 (1972).
[23] K. Kandiah, A. Smith and G. White. Proc. 2nd ISPRA Nuclear Electronics Symp., May 1975.
[24] S.V. Uchaikin. *Instruments and Experimental Techniques* **40**, 581 (1997).

4

Pulse Processing Electronics

INTRODUCTION

Previous chapters have dealt mainly with the analogue part of spectrometers and the analogue representation of the pulse propagation through the pulse-shaping network. As we mentioned before, modern detector systems tend to process information digitally and perform a number of other operations with acquired pulses. These are, for instance, the selection by height and counting in multichannel analyser (MCA) to present measurement results in a histogram form, the pulse recognition, the pile-up inspection, the rise time discrimination and others.

The modern designs of signal processing systems tend to start the digitisation of the detector signal at the earliest possible stage, which is the output of the preamplifier. Such systems are called digital spectrometers, as opposite to their analogue counterparts in which the conversion into the digital form is performed after the pulse-shaping filter. Early digitisation offers serious advantages and new possibilities in the nuclear instrumentation. These are the cost-effectiveness, use of transfer functions of shaping networks with improved noise indexes, the absence of fluctuations and drifts inherent to the analogue circuitry, an improved count rate capability and so on. To make the statement more illustrative, we will give the following example. In the digital spectrometer, after a pulse has been converted into a shaped peak by a digital filter, the captured peak maximum is immediately available as a digital value. It can be used as an address in the histogramming process without introducing an extra delay. In other words, the multichannel analyser in its

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin
 ISBN: 0-470-85759-5

bulky, complex form practically disappears as does associated with it the "dead" time.

At present, general purpose, universal digital pulse processing (DPP) units, capable to recognise and process signals either from pulse-reset or continuous-reset preamplifiers, are commercially available. They normally employ an adaptive structure enabling the automatic optimisation of system parameters as a function of the input count rate, the base line instability, background noise and so on. In practice, this convenience often comes at a price of adding a few eV to the energy resolution. It is still left up to a user to provide a perfect matching of the preamplifier with a DPP, eliminate parasitic ground loops and perform fine adjustments of the bias, the integration time, the gain, discriminator thresholds and so on, if a measurement requires preserving the ultimately achievable accuracy. The latter is usually the case when dealing with superconducting detectors. Their relatively high cost and complexity is fully justifiable only in applications that cannot be fulfilled with any other conventional analytical instruments.

Following the introduction, the sections of the current chapter will concentrate on selected problems and critical components of spectrometers, which have a strong effect on resulting measurement accuracy, rather than giving a detailed description of the DPP principle. Readers, interested in the latter, should refer to other comprehensive books, such as [1–3]. In the section 4.3, we will briefly discuss a potential fast superconductor DPP, which can operate at low temperature. Such a DPP is particularly necessary in the conjunction with large format detector arrays in order to eliminate multiple wires connecting them to the room temperature electronics.

4.1 PULSE PROCESSING TECHNIQUES

In previous chapters, we discussed a variety of superconductor primary detectors. Each of them is capable to operate in more than one biasing mode and match with different types of preamplifiers. In every individual combination, a detector produces a set of unique pulse shapes in response to an incoming event of interest. For example, let us consider a current-biased STJ coupled to a FET transconductance preamplifier and exposed to four incoming photons of the same energy and timing. Figure 4.1 shows three pulse shapes at the output of a preamplifier employing a pulse-reset technique (b), the continuous discharge via the feedback resistor (c), and, finally, the continuous discharge through the forward-biased input-FET (d). In Figure 4.1 (c), we emphasise the fact that the time constant of the feedback circuit is large enough (due to the large R_F) to consider the decay

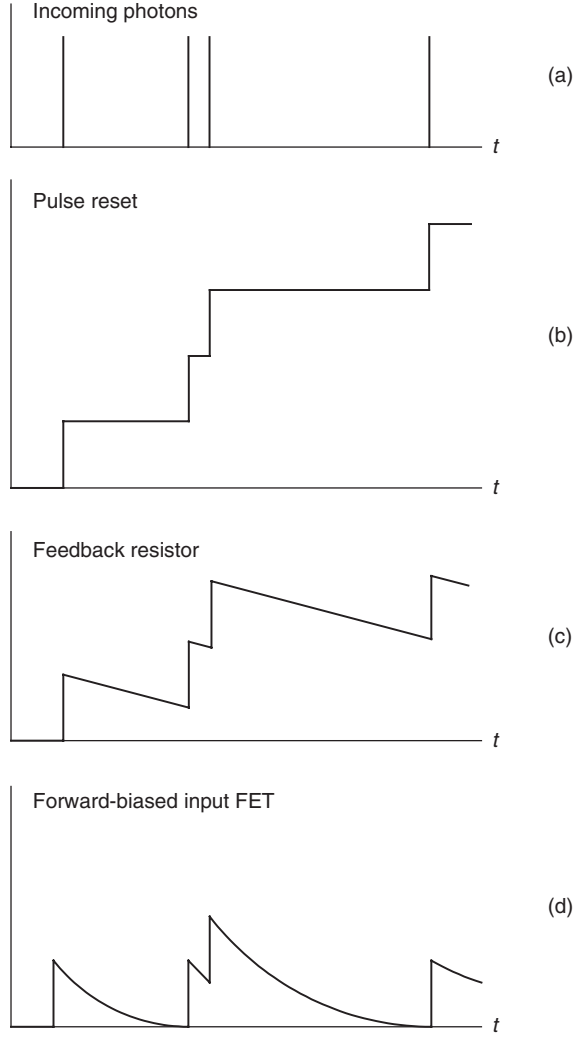


Figure 4.1 Three pulse shapes at the output of a preamplifier employing a pulse-reset technique (b), the continuous discharge via the feedback resistor (c), and the continuous discharge through the forward-biased input FET (d). Pulse shapes are so obviously different that it would not be possible to process all of them with the same digital spectrometer, at least, without introducing of a some signal conditioning unit and/or tuning the firmware and software algorithms

being linear at a reasonable input count rate. Pulse shapes are so obviously different that it would not be possible to process all of them with the same digital spectrometer, at least, without introducing some signal conditioning unit and/or tuning the firmware and software algorithms. A similar

problem may occur even with the use of one type of the preamplifier. Some superconductor detectors with large area absorbers and detector arrays produce output pulses with a shape and timing changing as a function of the position sensitivity. This can seriously affect the final energy resolution caused by the incomplete charge collection and the ballistic deficit unless the digital spectrometer takes into account details of the detector arrangement, specifics of an application and introduces necessary corrections into the final result.

From this short introduction, one can conclude that the development of a digital spectrometer with adequate performance requires a creative approach involving multidisciplinary expertise in system engineering, electronics, software and applications.

The detector spectrometer performs five major functions, which involve:

1. the pulse recognition;
2. the separation of a pulse of interest produced by a primary sensor from the parasitics and pile-up;
3. accurate measurement of its height or the area assuming these parameters are proportional to the energy of an incoming X-ray photon;
4. assigning a digital number (address) to each measurement; and, finally,
5. adding a count to the corresponding channel in an MCA for histogramming.

These functions lead us to a generic functional block diagram presented in Figure 4.2. The signal of the preamplifier arrives at the input of a signal conditioning unit (SCU) and a discriminator. The SCU modifies the signal in such a way that its noise and amplitude match the resolution and the dynamic range of the analogue-to-digital converter (ADC). The Nyquist pre-filter incorporated between the SCU and the ADC minimises the aliasing effect caused by the digitisation process. The task of the discriminator is to ensure that only one photon is measured at a time. It includes a filter, comparator and trigger. In the digital spectrometers, the signal discrimination can be performed digitally using the hardwired logic or/and the computation power of the digital signal processor (DSP). In this case, the input of the discrimination unit should be connected to the output of the ADC (shown by the dash line).

The efficiency of discriminators is very important in producing a final spectrum free of artefacts such as, for instance, sum peaks. These peaks appear when two X-rays arrive short after each other. If a discriminator

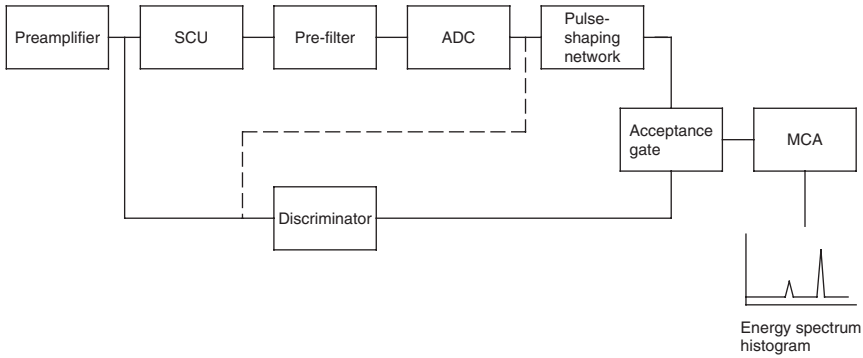


Figure 4.2 A generic functional block diagram of the digital spectrometer. The signal of the preamplifier arrives at the input of a signal conditioning unit (SCU) and a discriminator. The SCU modifies the signal in such a way that its noise and amplitude match the resolution and the dynamic range of the ADC. The Nyquist pre-filter incorporated between the SCU and the ADC minimises the aliasing effect caused by the digitisation process. The task of the discriminator is to ensure that only one photon is measured at a time. In the digital spectrometers, the signal discrimination can be performed digitally. In this case, the input of the discrimination unit should be connected to the output of the ADC (shown by the dash line)

does not distinguish between the peaks, the system counts them as one (for instance, with the double energy in response to a monochromatic irradiation). Sum peaks cause the identification error; that is an operator may see peaks corresponding to elements that are not present in the sample. The problem becomes even more serious if discriminators are not efficient in detecting pile-up caused by low-energy photons. The whole spectrum of a composition specimen (containing light chemical elements) will be distorted, as sum peaks will be sitting on the high side of every peak even at a quite modest count rate.

Pile-up grows at the square of the input count rate [4]. Therefore, the presence of sum peaks can be determined by changing their intensities as a function of the current of the excitation beam. Another method of the pile up identification was proposed in [5]. It employs a random oscillator producing pulses with fixed amplitude and a shape similar to the one of the primary detector. The number of pulses injected by the oscillator is compared with a number of counts in the artificial peak. The difference yields the fraction of lost pulses due to pile-up with real events into the sum peaks. By changing the amplitude of the pulses, it is possible to monitor the efficiency of the discriminator in the whole energy range of interest and even introduce quantitative corrections for errors caused by the pile-up.

High-quality spectrometers employ multichannel discriminators to ensure an efficient pile-up protection over the wide range of the detection energy: fast, intermediary and slow ones. The speed is defined by the time constant of the filter incorporated into each discrimination channel. Fast discriminators deal with high-energy peaks when the system must ensure the high throughput of data. Low-energy X-rays can be distinguished from noise only with the help of a filter with longer time constants. Therefore, they are usually processed with the slow channel, which offer a modest count rate capability.

The timing functionality of a discrimination unit depends on whether a spectrometer employs a pulse-shaping network with a fixed or variable integration time, T_p. In the system with a fixed T_p, the acceptance gate will reject all pulses arriving within T_p, as shown in Figure 4.3(a). T_p in the systems with the adaptive pulse-shaping network is defined by the time interval between pulses (Figure 4.3b). An operator sets up only the minimum processing time, T_{pmin}. If a following event occurs within T_{pmin}, the acceptance gate rejects it. Other pulses will be counted in bins of the MCA.

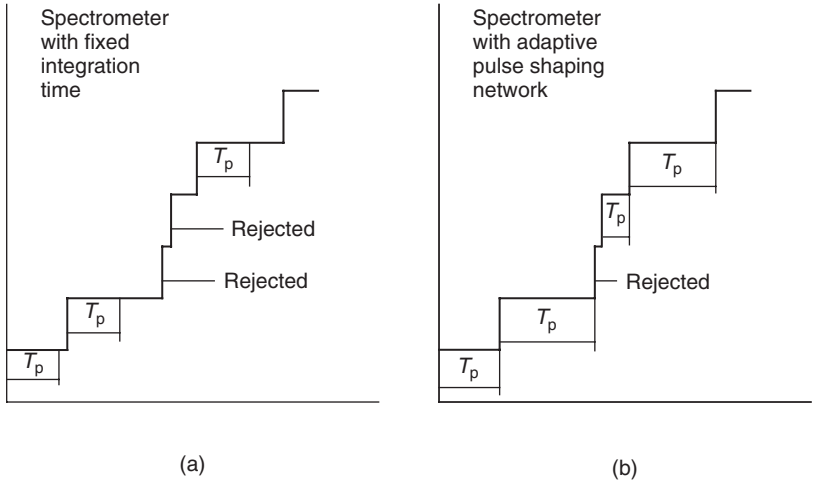


Figure 4.3 The output voltage of a pulse-reset preamplifier as a function of time explaining the principle of spectrometers with a fixed (a) and variable (b) integration time. In the system with a fixed T_p, the acceptance gate will reject all pulses arriving within T_p. T_p in the systems with the adaptive pulse-shaping network is defined by the time interval between pulses. An operator sets up only the minimum processing time, T_{pmin}. If a following event occurs within T_{pmin}, the acceptance gate rejects it. Other pulses will be counted in bins of the MCA

The spectrum acquired with the variable integration time contains a number of Gaussian shapes with a distribution of resolutions determined by the distribution of arrival intervals of events on the voltage ramp. Therefore, the final resolution depends on the rate and, in fact, two consecutive measurements can yield different figures. With the sufficient statistics, however, these variations normally smooth out.

At a low input count rate the resolution of the adaptive spectrometer comes close to the resolution of a system with the fixed longest integration time. At the highest input count rate it becomes comparable with results obtained with a spectrometer at the shortest fixed time constant. In between, the adaptive spectrometers deliver a better energy resolution, as it uses the maximum of the available measurement time.

All digital spectrometers can be classified based on different types of the signal conditioning units into three major categories:

1. AC-coupled spectrometers that employ a CR-differentiating network between the output of the preamplifier and the pre-filter circuit;
2. DC-coupled spectrometers in which the output of the preamplifier is directly connected to the input of the pre-filter unit; and
3. DC-coupled spectrometers in which the low-frequency component of the preamplifier output signal is subtracted by analogue means in a controlled way.

In addition to that, each configuration can utilise some kind of a deconvolver either for reducing the effect of the ballistic deficit or for the base line correction. In what follows, we will discuss all three spectrometer configurations in more detail.

4.1.1 AC-Coupled Digital Spectrometers

In an AC-coupled system, the output signal of the preamplifier is passed through a C_dR_d differentiator, as shown in Figure 4.4. The transfer function of the preamplifier is given by

$$H_p(j\omega) \approx \frac{K_p(\omega)}{1 + j\omega\tau_p} \tag{4.1.1}$$

where $K_p(\omega) = \frac{A(\omega)}{1+A(\omega)}\frac{1}{C_p}$, $\tau_p = R_pC_p$ and $R_p = \frac{R_iR_f}{R_i+\frac{R_f}{A(0)+1}}$.

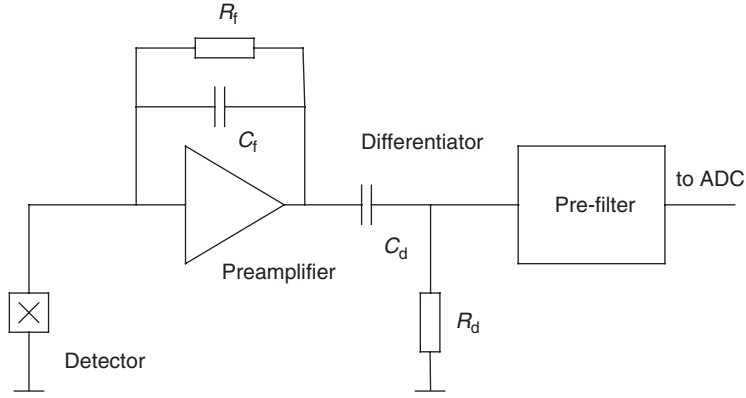


Figure 4.4 A schematic diagram of the analogue channel of the AC-coupled spectrometer system. The output signal of the preamplifier is passed through the C_dR_d-differentiating network removing the DC component

The inverse Fourier transform of expression (4.1.1) yields the impulse transfer function of the preamplifier, $h_p(t)$. With usual assumptions $A \gg 1, R_i \to \infty$, the time constant τ_p becomes equal to the feedback time constant $\tau_p \approx \tau_f = R_fC_f$ and $K_p \approx \frac{1}{C_f}$. This results in the exponentially decaying $h_p(t)$

$$h_p(t) \approx \frac{1}{C_f} e^{-\frac{t}{\tau_f}} \tag{4.1.2}$$

The response of the preamplifier to the detector current $I_d(t)$ can be found directly in the time domain using the convolution integral as follows [6]

$$v_p(t) = \int_0^{T_c} I(t') \frac{1}{C_f} e^{-\frac{t-t'}{\tau_f}} dt' = \frac{1}{C_f} e^{-\frac{t}{\tau_f}} \int_0^{T_c} I(t') e^{\frac{t'}{\tau_f}} dt' \tag{4.1.3}$$

where T_c is the duration of the current pulse from the detector. The integral $\int_0^{T_c} I(t') e^{\frac{t'}{\tau_f}} dt'$ is constant. Therefore, at $t > T_c$, the decay rate of v_p is defined by the time constant of the feedback network of the preamplifier, as shown in Figure 4.5.

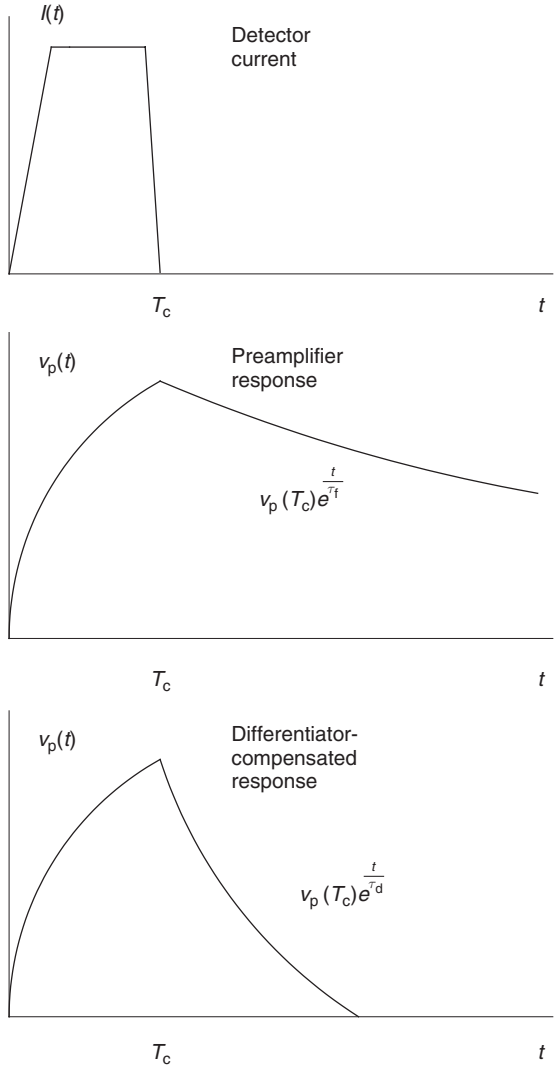


Figure 4.5 Output signal of the detector, preamplifier and a CR-differentiating network with the pole-zero correction

With a long time constant τ_f, the output signal of the preamplifier may exceed the dynamic range of ADC at a certain input count rate. The C_dR_d differentiator prevents that. Its transfer function is given by

$$H_d(j\omega) = \frac{j\omega}{1 + j\omega\tau_d} \tag{4.1.4}$$

with $\tau_d = C_d R_d$. The impulse response function of the preamplifier–differentiator pair can be found from the inverse Fourier transform of the product $H_p(j\omega)H_d(j\omega)$. The result of calculation has the form

$$h(t) = \frac{1}{C_f}\left(\frac{w}{w-1}e^{-\frac{t}{\tau_d}} - \frac{1}{w-1}e^{-\frac{t}{\tau_f}}\right) \tag{4.1.5}$$

where $w = \frac{\tau_f}{\tau_d}$.

In order to simplify the pulse processing and minimise the ballistic deficit it is desirable to change the response function, given by the equation (4.1.5), to a single-pole response. An example of the single-pole response function is shown in Figure 4.5. Analytically, it can be written as

$$h(t) = \frac{1}{C_f}\frac{w}{w-1}e^{-\frac{t}{\tau_d}} \tag{4.1.6}$$

This operation is often called the pole-zero correction. There are two ways to achieve this correction by employing either the compensation technique or the deconvolution of the differentiator output signal straight after its digitisation.

The compensation of the term $\frac{1}{w-1}e^{-\frac{t}{\tau_f}}$ is achieved by adding of a $\frac{1}{w-1}$ fraction of the preamplifier output voltage to the output voltage of the differentiation network, as shown in Figure 4.6. This basic pole-zero cancellation circuit ensures a single-pole response at any desired decay time produced by the time constant of the differentiator, τ_d.

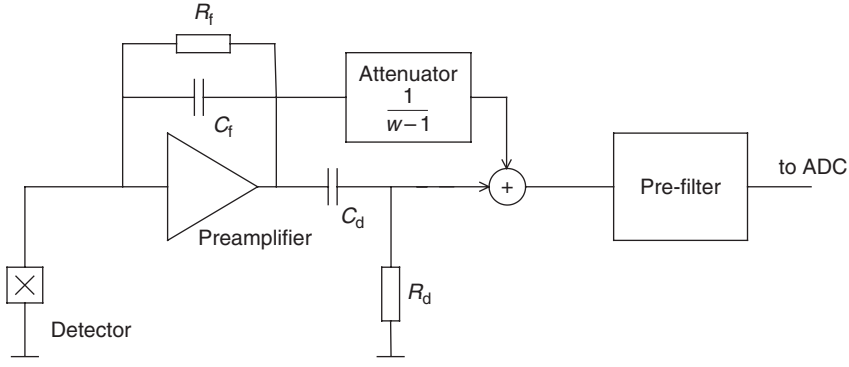


Figure 4.6 A functional circuit diagram illustrating the pole-zero correction by employing the compensation technique

A first spectrometer incorporating an automatic pole-zero corrector was described in [7]. In this design, the corrector adjusts a division ratio of a digitally controlled attenuator until the signal at the output of the pulse-shaping network does not show an overshot or an undershot with respect to the zero line. Originally, automatic pole-zero adjustments employed consecutive iteration algorithms required a certain amount of the calibration time and X-ray pulses before a spectrometer is ready for actual pulse processing. This works quite well with preamplifiers in which the continuous discharge rate is fixed by a known feedback resistor, R_f. In this case, the re-calibration procedure would only need to be repeated after the exchange of the preamplifier by another one with a different time constant, τ_f.

The problem becomes much more complex when spectrometer should process the output signal of a preamplifier with the continuous discharge via the forward-biased input transistor. The resistance of the forward bias junction is a non-linear function of the voltage between the gate and source electrodes. This generates a multi-pole response of the preamplifier that results in an impulse function with a complex decay tail. In addition, the voltage across the gate–source electrodes and, hence, the forward-biased resistance as well as the input capacitance depend strongly on the rate and the energy of the incoming events. In order to process such signals the spectrometer should monitor their shapes continuously and employ more sophisticated techniques that provide an instant error correction.

All instant correction techniques are performed digitally on the basis of the signal deconvolution theorem. The digital data accumulated before the pulse-shaping stage are deconvolved so that to eliminate the poles and zeroes introduced by a part or even the whole analogue channel. This simplifies the signal itself, which may look like, for instance, a pure exponential decay, and the algorithm of the digital pulse-shaping network. An example of a real time recursive algorithm removing the impulse response of the high pass C_dR_d network is described in [8].

Clearing the data of all extra poles introduced by the whole analogue channel leads to restoring the original shape of pulses generated by a radiation event in the primary detector. First, this immediately recovers the signal from the ballistic deficit, which is important for superconductor detectors with their substantial variations in the rise time. Second, the pulse-shaping of the original charge distribution enabled us to achieve the best compromise between the signal-to-noise ratio, the resolving time and the high throughput without the degradation of the energy

resolution. This approach is currently at its initial development stage. It will require seeking new optimum transfer functions of the pulse-shaping networks and introducing radical changes into pile-up rejection techniques.

4.1.2 Deconvolution Techniques

The idea of the total signal deconvolution and processing the restored shape of the detector output signal is expected to be widely accepted in the future digital radiation spectrometry. This is particularly true for systems based on primary superconductor sensors such as the STJ and the TES attached to large area absorbers. As we mentioned before, these devices generate position-sensitive shapes and timing. The pulse peaks convoluted with the impulse response of the analogue circuitry depend non-linearly on energies of corresponding events. Unless properly corrected, this non-linearity may cause a substantial degradation of the final energy resolution and shifts peak positions on the energy histogram.

An example of a functional diagram of a digital spectrometer involving a deconvolution unit is shown in Figure 4.7. The deconvolver itself is normally attached straight to the output of the digitiser. The numerical strings of restored signals follow further down the usual root through the pulse-shaping network, the gate to the MCA. The discriminator measures time between two consecutive pulses in order to determine whether closely spaced peaks affect the accuracy of the deconvolution unit or/and a pile-up is going to occur in the pulse-shaping network.

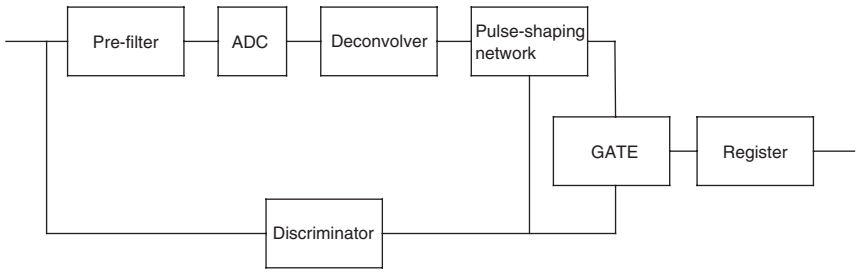


Figure 4.7 An example of a functional diagram of a digital spectrometer involving a deconvolution unit. The deconvolver itself is normally attached straight to the output of the digitiser. The numerical strings of restored signals follow further down the usual root through the pulse-shaping network, the gate to the MCA. The discriminator measures time between two consecutive pulses in order to determine whether closely spaced peaks affect the accuracy of the deconvolution unit or/and a pile-up is going to occur in the pulse-shaping network

The signal at the input of the ADC is given by

$$v_{\mathrm{p}}(t) = \int_{-\infty}^{\infty} g(t) f(t-\tau)\mathrm{d}\tau \tag{4.1.7}$$

where $g(t)$ is the output signal of a detector, $f(t)$ represents the impulse response of the analogue channel that includes a preamplifier, a pre-filter and a CR-differentiating circuit in the AC-coupled spectrometers. The ADC essentially transforms the equation (4.1.7) into a digital convolution integral in the form

$$v_{\mathrm{p}}(it_{\mathrm{s}}) = \sum_{j=0}^{\infty} g(jt_{\mathrm{s}}) f(it_{\mathrm{s}} - jt_{\mathrm{s}}) \tag{4.1.8}$$

where t_{s} is the sampling interval. With the timescale normalised by t_{s}, we can further modify (4.1.8) as follows

$$v_{\mathrm{pi}} = \sum_{j=0}^{\infty} g_i f_{i-j} \tag{4.1.9}$$

In this expression, $v_{\mathrm{p}i}$, g_i and f_{i-j} become dimensionless numerical strings representing samples of the output voltage of the analogue channel, the detector response and the impulse response function respectively.

An ultimate goal of the deconvolution procedure is to resolve the system of infinite number of linear equations (4.1.9) with respect to g_i. In practice, this can only be achieved by elimination of those equations containing insignificant strings. It is desirable to develop a fast deconvolution algorithm that does not require a large computational power and provides continuous real time data flow in the spectrometer. One example of such an algorithm, called Moving Window Deconvolution, was proposed in [9]. Authors suggest to select an initial finite time window at the beginning of the acquisition process. Within this window, one deals with finite number of numerical strings with a length equal to the size of the window. For instance, if the duration of the window is mt_{s}, then the equation (4.1.9) takes the following truncated form [9]

$$v_{\mathrm{pi}} = \sum_{j=0}^{m} g_i f_{i-j} \tag{4.1.10}$$

The solution of equation (4.1.10) with respect to g_i yields the total charge released by a detector during the selected time window mt_s as follows

$$g(1) = \sum_{i=0}^{m} g_i \tag{4.1.11}$$

Sliding the window to the right by one sampling interval t_s from $j = 1$ to $j = m + 1$ requires to solve one equation in order to find the charge increment Δg released during this interval. Therefore, total charge released by the detector by the moment $m + 1$ becomes

$$g(2) = g(1) + \Delta g|_{m+1} \tag{4.1.12}$$

Performing similar computations for all consecutive steps, until the final one is reached, we derive the sought charge distribution function of the detector. In order to further speed up resolving the inverse problem for g_i the impulse response function can be coded as a part of the spectrometer firmware. The function $f(t)$ itself can be either defined analytically, if possible, or calculated from measured actual transfer function of the analogue chain between the detector and the input of ADC.

4.1.3 DC-Coupled Digital Spectrometers

DC-coupled spectrometers do not employ the CR-differentiating network. As a result, they do not require a pole-zero correction, but need to use other means for matching the signal swing with the dynamic range of the ADC. One example of a DC-coupled spectrometer is shown in Figure 4.8. It works with a pulse-reset preamplifier producing a staircase of non-decaying voltage steps. The reset is applied each time as the voltage exceeds the dynamic range of the ADC. After anti-aliasing prefiltering, the preamplifier output is sampled continuously by the ADC and X-ray pulse heights are derived by subtracting the weighted average of a set of values before X-ray events from the weighted average of a set of values after X-ray events. In fact, a set of values of a previous event may serve as a base line for the following one.

This concept of the pulse-reset spectrometer, if it can be implemented, achieves the best compromise among various figures of merit, primarily

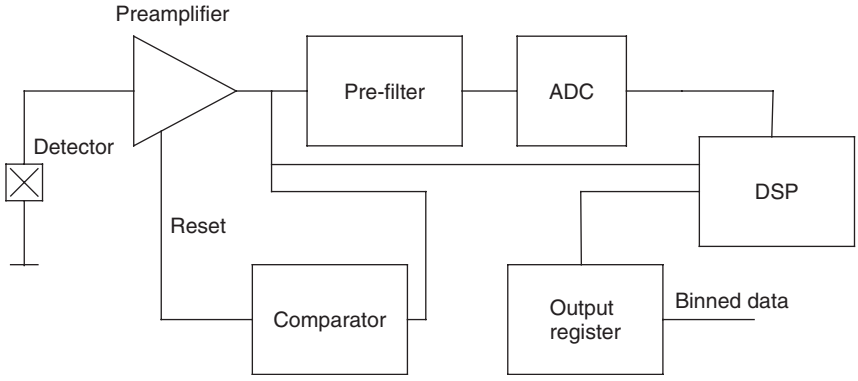


Figure 4.8 An example of a DC-coupled spectrometer. It works with a pulse-reset preamplifier producing a staircase of non-decaying voltage steps. The reset is applied each time as the voltage exceeds the dynamic range of the ADC. After anti-aliasing pre-filtering, the preamplifier output is sampled continuously by the ADC and X-ray pulse heights are derived by subtracting the average of a set of values before X-ray events from a set of values after X-ray events. In fact, a set of values of a previous event may serve as a base line for the following one

between the accuracy and the acquisition rate. There is a number of factors characterising this compromise. These are:

1. The concept does not require a deconvolver that otherwise would be needed to eliminate introduced poles. This reduces the overall processing time and always-possible errors associated with any inverse problem. The latter is very important because the deconvolution technique indeed does not work well with the pulse-reset preamplifier. The digitisation of flat steps produces a string of correlated samples. This correlation adds a substantial amount of noise to the recovered distribution function $g(t)$ due to non-linearities and the quantisation noise of the ADC [10].
2. By measuring the voltage ramp the base line is recovered instantly and automatically.
3. The peak position on the energy histogram does not shift with the count rate and energy of photons. The peak stability is a particularly important feature for the X-ray mapping with the following quantitative analysis of small features in the large area samples.

The leakage current of the primary detector and the input-FET of the transconductance preamplifier cause a finite slope of the ramp voltage, as shown in Figure 4.9. Firmware should take it into account, for

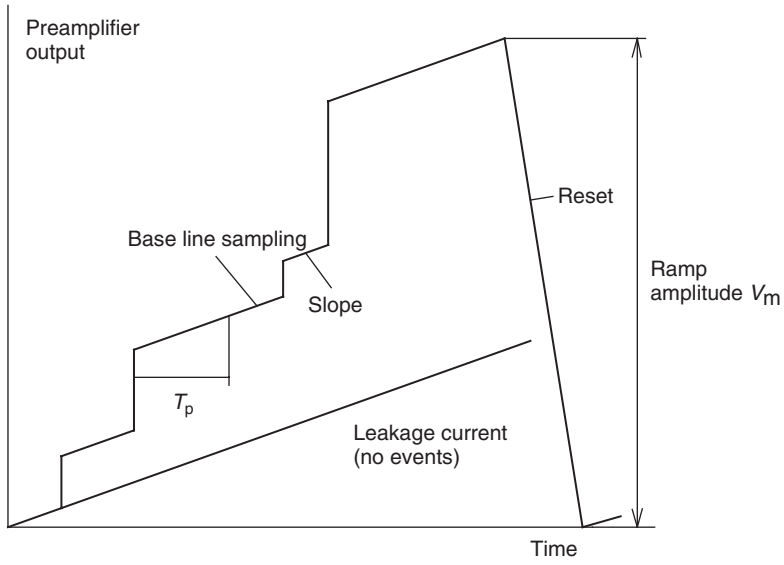


Figure 4.9 The leakage current of the primary detector and the input FET of the transconductance preamplifier cause a finite slope of the ramp voltage. Firmware should take it into account by introducing correction factors into weighting coefficients of the pulse-shaping network. The correction factors stay constant as long as the base temperature, the bias and other operating conditions of the primary sensor do not change. If that is not the case, sampling the zero base line, while events are not present, allows one to continuously update the values of correction factors during the sample analysis

instance, by introducing correction factors into weighting coefficients of the pulse-shaping network. The correction factors stay constant as long as the base temperature, the bias and other operating conditions of the primary sensor do not change. If that is not the case, sampling the zero base line, while events are not present, helps to continuously update the values of correction factors during the pulse analysis.

A serious disadvantage of this type of DC-coupled spectrometers is a rather inefficient usage of the dynamic range of the ADC. The dynamic range should accommodate not just step heights produced by individual events, but also the whole amplitude of the ramp, V_m. Since the ramp accommodates a certain amount of steps, a reset rate is governed by a number of bits in the ADC, the input count rate and the leakage current. Signal-to noise considerations do not allow V_m to be larger than a couple of hundred mV even with 14- and 16-bit ADCs. If either leakage current or the input count rate, or both of them happen to be large, resetting may contribute a substantial portion of the dead time

limiting the maximum possible throughput of the system. In this respect, the DC-coupled spectrometers relay heavily on the major progress of high-speed precision ADCs. Semiconductor companies commit serious investments in this field to primarily satisfy the demand of the communication market. As a result, a wide selection of fast, relatively inexpensive ADCs is now available (see, for instance a selection in [11]). Modern ADCs make the configuration, shown in Figure 4.8, quite competitive among other processing techniques. The future competitiveness is linked with the future general progress in ADCs.

4.1.4 A DC-Coupled Digital Spectrometer with an Analogue Signal Conditioner

A general purpose digital spectrometer may be required to accommodate voltage ramps that extend well above the dynamic range of any ADC available. Such a signal has to be conditioned with some analogue means before the conversion into digital strings. W. Warburton and B. Hubbard [12] proposed a very inventive conditioning technique that, similar to the CR-differentiating network, removes the low-frequency component of the preamplifier output, but maintains the analogue channel to be DC-coupled to the ADC. In other words, the method preserves advantages of both the AC- and DC-coupled spectrometers eliminating, however, major problems associated with the pole-zero correction as well as the dynamic range and non-linearities of the ADC. Figure 4.10 shows an example of the functional block diagram of the

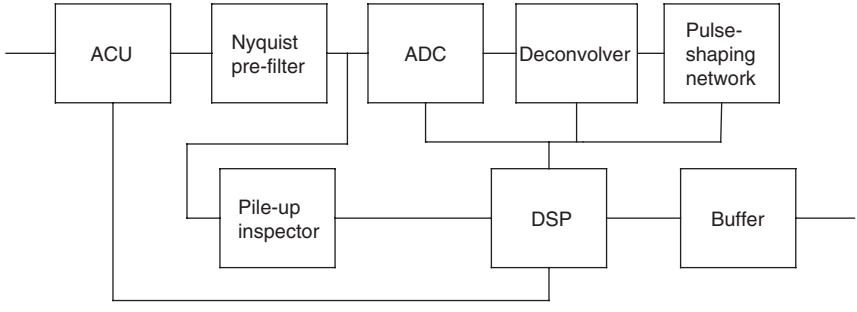


Figure 4.10 An example of the functional block diagram of the digital spectrometer including an analogue conditioner (the concept is reproduced by permission of W. Warburton). The analogue conditioning unit (ACU) follows directly the output of the preamplifier

digital spectrometer including an analogue conditioner. The analogue conditioning unit (ACU) follows directly the output of the preamplifier. In the Warburton *et al.* technique, it contains a waveform generator, which is digitally controlled by the DSP. The waveform generator constructs a sawtooth voltage to subtract it from the preamplifier output signal, as illustrated in Figure 4.11. It can be seen that the swing

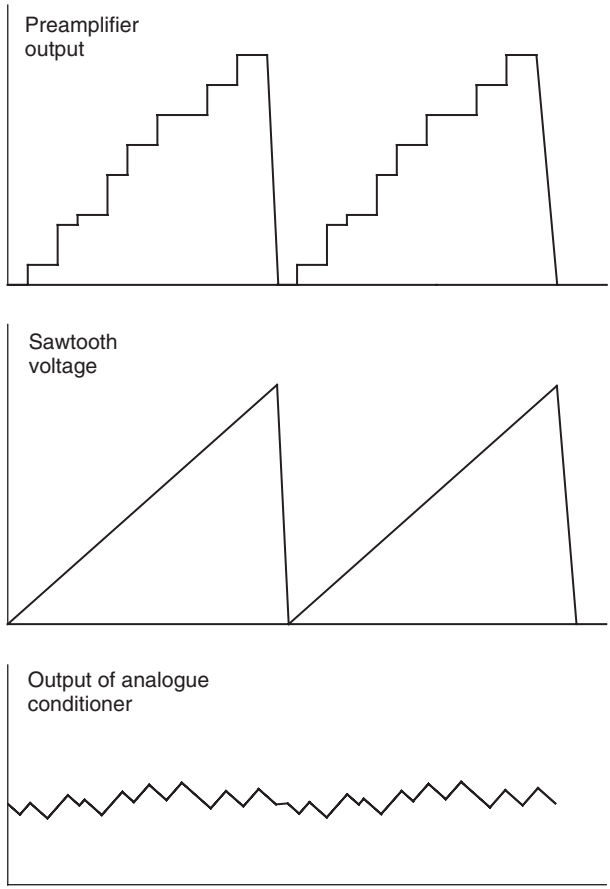


Figure 4.11 In the Warburton *et al.* technique [12], the analogue conditioning unit contains a waveform generator, which is digitally controlled by the digital signal processor (DSP). The waveform generator constructs a sawtooth voltage to subtract it from the preamplifier output signal. It can be seen that the swing of the conditioned signal becomes smaller and does not depend on the maximum voltage of the ramp. At the same time, it contains all the information about the energy of incoming events. Knowing the slope of the constructed sawtooth voltage it can be restored with a non-complicated deconvolution algorithm (the concept is reproduced by permission of W. Warburton)

of the conditioned signal becomes smaller and does not depend on the maximum voltage of the ramp. At the same time, it contains all the information about the energy of incoming events. Knowing the slope of the constructed sawtooth voltage it can be restored using a non-complicated deconvolution algorithm.

The average slope of the ramp is given by

$$\text{Slope}_{\text{ave}} = \frac{I_{\text{inave}} + I_{\text{leakage}}}{C_{\text{f}}} \tag{4.1.13}$$

where I_{inave} is the average output current of the detector during ramp time defined by the energy and the input rate of incoming events and I_{leakage} is the total leakage current flowing into the input terminal of the front-JFET. According to the equation (4.1.13), it may vary significantly during the measurement time. Therefore, the DSP continuously analyses the ADC output to prevent the difference signal departing from the ADC allowed input range. In fact, it is adjusted in such a way that the waveform generator produces almost a replica of the average slope of the ramp.

Digital spectrometers comprising a signal conditioning unit can employ very fast ADCs having a modest number of bits. This does not affect the accuracy of the processing due to the reduced swing of the incoming signal, but with less frequent resets increases dramatically the maximum throughput that the spectrometer can handle. This property is particularly valuable when a digital spectrometer has to operate with a detector array and when a lot of background noise and parasitic events are present in the detector output.

4.1.5 Digital Pulse-Shaping Filter

As we mentioned before, the pulse-shaping filters of digital spectrometers operate with strings of discrete values representing the output voltage of a preamplifier rather than with the voltage itself like their analogue counterparts. Yet the objective remains the same: to determine the height of a step produced by an incoming event with the best possible signal-to-noise ratio. Digital filters accomplish that by taking a weighted average over the points before the step and subtracting it from the average over the points after the step. This allows us to write down the

filter function values $\{T_i\}$ of a stream data $\{d_i\}$ at time moments $\{i\} = \left\{\frac{t}{T_s}\right\}\Big|_{\text{integer}}$ in the form [12]

$$T_i = -\sum_{k=i-2L-G+1}^{i-L-G} w_k d_k + \sum_{k=i-L+1}^{k} w_k d_k \qquad (4.1.14)$$

where the length of the averaging, $L = \left\{\frac{T_p}{T_s}\right\}\Big|_{\text{integer}}$ is the measurement time, T_p, normalised by the sampling interval, T_s. It is equivalent to the peaking time in the analogue filters. $G = \left\{\frac{T_g}{T_s}\right\}\Big|_{\text{integer}}$ is the normalised gap interval. The gap interval takes care of the ballistic deficit due to finite charge collection time and includes the rise and settling times of the step caused by the preamplifier. Discrete factors $\{w_i\}$ represent weighting constants of the filter. They essentially determine the type of average being computed, which is equivalent to the impulse response function in the analogue domain. For instance, when the weighting coefficients decrease with the separation from the step, the filter has a "cusp-like" response function. When weighting factors are constant, the filter becomes trapezoidal. Both the cases are illustrated in Figure 4.12.

If the gap is negligibly small compared to the length, the trapezoidal filter becomes triangular. It provides the best signal-to-noise ratio with respect to the series part of the noise. The "cusp" filter is more efficient in reducing the parallel noise. In the both cases though the noise indexes grow as the gap increases. Therefore, one has to look for a compromise in terms of the width of the gap to balance effects of the ballistic deficit and noise indexes on the worsening of the final energy resolution of the detector system.

The equation (4.1.14) requires the storage of the most recent $2L + G$ data points and a large amount of time consuming adding operations. This can be avoided by using the recursion relation between neighbouring values T_i and T_{i-1}. Thus, the practical pulse-shaping algorithm takes the following form [10,12]:

$$T_i = T_{i-1} + w_i d_i - d_{i-L-G} + d_{i-2L-G} \qquad (4.1.15)$$

Further noting that the partial sum $d_{i-L-G} + d_{i-2L-G}$ is equal to the partial sum $d_i - d_{i-L}$ evaluated $L + G$ time steps earlier, T_i can be computed using only two FIFO memories. The first one with dimensions $C_w \times L$ to hold L values for making sums $d_i - d_{i-L}$ and the second one with dimensions $(C_w + 1) \times (L + G)$ for actual calculating T_i. Here C_w is the bit width of words of a DSP utilised.

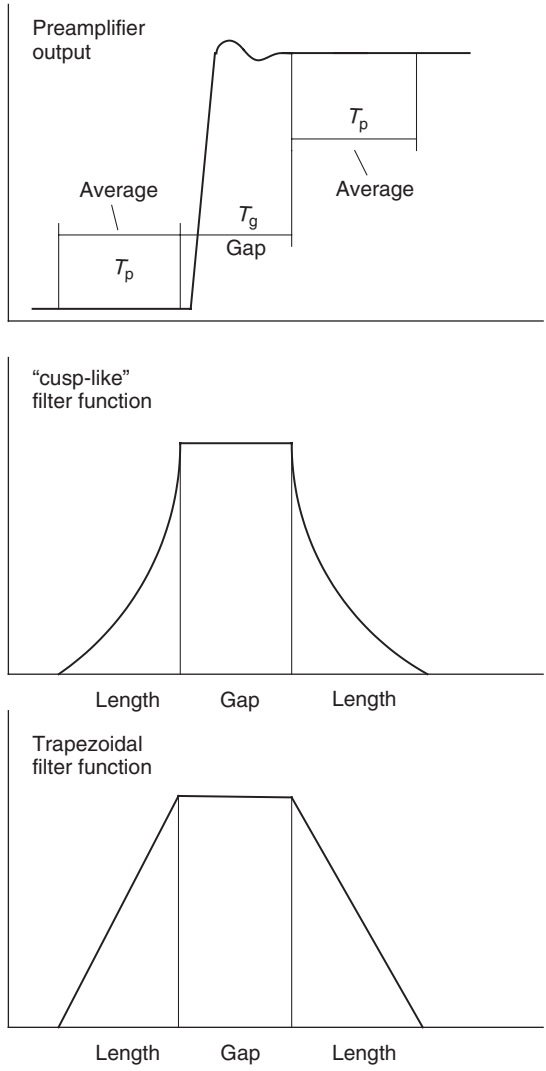


Figure 4.12 Weighting constants of the digital filter essentially determine the type of average being computed, which is equivalent to the impulse response function in the analogue domain. For instance, when the weighting coefficients decrease with the separation from the step, the filter has a "cusp-like" response function. When weighting factors are constant, the filter becomes trapezoidal

Digital filters, unlike their analogue counterparts, do not have decaying tails on the completion of the basewidth $2L + G$. Therefore, digital spectrometers have a greater potential in terms of the throughput without the degradation of the energy resolution.

4.1.6 Throughput of Digital Spectrometers

4.1.6.1 Input count rate

There are three types of the input count rates (ICR) in every detector system. These are the incoming ICR, the true ICR and the measured ICR. The incoming ICR represents a rate at which radiation photons interact with the primary detector. Due to the finite time constant of the detector and the random nature of these interactions, a fraction of events will be indistinguishable as separate ones. This happens when the arrival times of, say, two photons are so close that they produce a joint single pulse with a height proportional to the sum of their energies. A second type of ICR is the rate at which pulses appear at the output of the preamplifier or the input of the spectrometer. From the point of view of a spectrometer, this is the true ICR. The system attempts to count all preamplifier pulses with a height exceeding a threshold set up in the fast discrimination channel. But, again, due to the pile-up in the fast channel some of the pulses will be missed. Remaining pulses, which the system manages to count and report to an operator, constitute a third type of the ICR. It is called the measured ICR. Figure 4.13 helps to visualise all three types of ICRs in the detector system.

Assuming that the Poisson statistics is applied and the condition $2\tau_F \mathrm{ICR}_{\mathrm{true}} \ll 1$ is met, all three types of ICR relate to each other as follows [4]:

$$\mathrm{ICR}_{\mathrm{true}} = \mathrm{ICR}_{\mathrm{incoming}} e^{-\tau_{\mathrm{det}} \mathrm{ICR}_{\mathrm{incoming}}} \tag{4.1.16}$$

$$\mathrm{ICR}_{\mathrm{measured}} = \mathrm{ICR}_{\mathrm{true}} e^{-\tau_F \mathrm{ICR}_{\mathrm{true}}} \tag{4.1.17}$$

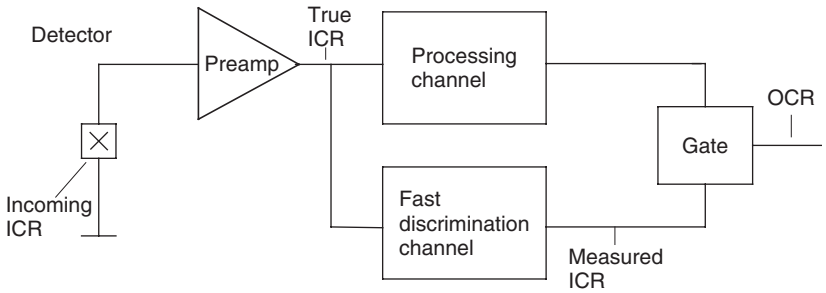


Figure 4.13 A functional circuit diagram of a spectrometer illustrating different types of input count rates (ICR)

where τ_{det} and τ_F represent the time constants of the primary detector and the filter in the fast discrimination channel respectively.

4.1.6.2 Output count rate

The time interval, which a system requires to process an incoming pulse, is often called the dead time, τ_d. The reason for that comes from the fact that from the moment a pulse has arrived at the input of the spectrometer during the processing time τ_d, the spectrometer does not accept any following pulses. The minimum dead time is defined by the basewidth of the pulse-shaping filter

$$\tau_{d\,min} = 2T_p + T_g \tag{4.1.18}$$

In practice, the dead time is always set up with some margin accommodating the intervals required to perform the deconvolution, various corrections, binning and so on. The relation between the $ICR_{measured}$ and OCR is given by

$$OCR = ICR_{measured} e^{-\tau_d ICR_{measured}} \tag{4.1.19}$$

Expressions (4.1.16) and (4.1.17) should be used to introduce corrections into the equation (4.1.19) if other types of ICR are to be used.

As the ICR increases, the number of events arriving within the interval $2T_p + T_g$ and, therefore, rejected by the fast discrimination channel becomes larger. As a result, at some incoming rate, the OCR reaches its maximum saturation value. The maximum value of the output count rate, OCR_{max}, at a fixed basewidth $2T_p + T_g$ characterises the maximum throughput of the spectrometer. OCR_{max} can be found from the plot of the OCR as a function of $ICR_{measured}$, as shown in Figure 4.14. If the plot fits the equation (4.1.19), the maximum throughput takes the form

$$OCR_{max} = \frac{1}{e\tau_d} \tag{4.1.20}$$

All spectrometers employ an internal system clock. This gives to an operator an option to set up the "live" time duration of the acquisition rather than using the real time to eliminate the need for explicit time corrections into the final spectrum. The "live" time is a measure with a meaning opposite to the earlier discussed "dead" time, that is, the time duration that it takes for the spectrometer to capture and processes only

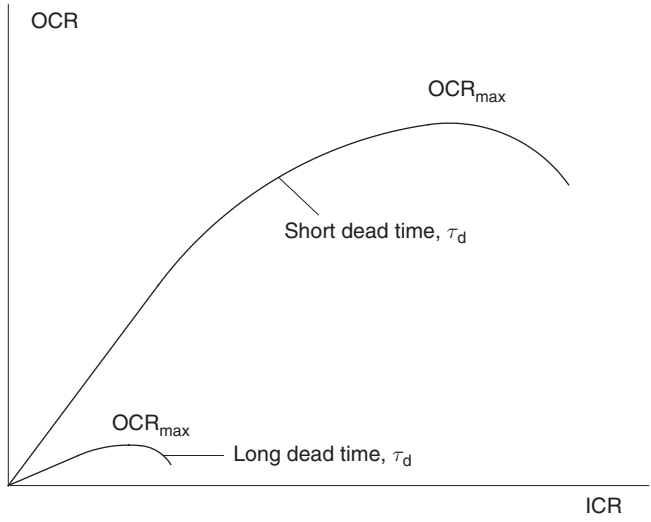


Figure 4.14 The output count rate (OCR) as a function of the input count rate (ICR) characterising the throughput of a spectrometer at short and long processing times, τ_d

accepted and binned pulses. One needs to notice that the definitions of the live time are somewhat different for analogue and digital spectrometers. Analogue spectrometers normally employ an input gate. The input gate blocks following pulses from reaching the pulse-shaping network while an accepted pulse is being shaped and digitised. In this case, the live time is simply the time over which the input gate was held open during the entire acquisition. In digital spectrometers, however, the preamplifier signal is continuously digitised and pipelined. The live time becomes a measure of the system availability to capture X-ray peaks from the output of the digital filter. It can also be considered as the time over which the virtual acceptance gate, shown in the functional block diagram (Figure 4.2), is open.

4.2 ANALOGUE-TO-DIGITAL CONVERSION

4.2.1 ADC: Basic Information

The Analogue-to-Digital Converter is possibly the most critical component of the detector system. In early analogue spectrometers, it was a part of the MCA dealing with a perfectly conditioned, free from pile-up pulse-shaping filter signal. In such configuration, an implementation of

a relatively slow ADC with a modest number of bits could still result in a reasonably good overall system performance. Modern digital spectrometers, however, tend to digitise the analogue signal at an early stage. Digital discrimination units and the pulse-shaping filter operate with digital strings rather than the analogue signals. As a result, an efficient rejection of the pile-up requires the conversion time of the ADC to be much shorter compared to the smallest time constant of the discrimination unit. In digital spectrometers, the converter must handle noisy and poorly conditioned signals, especially coming from non-synchronised pulse-reset preamplifiers. At the same time, moving the ADC towards the front stages of the system means that its noise, jitter, aliasing phenomenon and non-linearities have a much more pronounced effect on the final energy spectrum. To summarise all that, one can conclude without the exaggeration that a selection of an ADC ultimately defines the compromise between the energy resolution and the count rate capability of the whole system.

A functional diagram of the analogue-to-digital converter is shown in Figure 4.15. It accepts an analogue voltage, V_{I}, and produces an output binary word $b_1 b_2 \ldots b_N$ of a fractional value, ΔV_1 such that

$$\Delta V_1 = b_1 2^{-1} + b_2 2^{-2} + \cdots + b_N 2^{-N} = \frac{V_{\mathrm{i}}}{\kappa V_{\mathrm{r}}} \qquad (4.2.1)$$

where logic coefficients b_κ take values 0 or 1, κ is an appropriate scaling constant called the ADC gain, and V_{r} is the internal reference voltage. The equation (4.2.1) suggests that the conversion can be thought of as a two-step process. The first step includes sampling the input signal, V_{I}, in time. It is usually performed at regularly spaced intervals, T, which yields the sampling frequency defined as $f_{\mathrm{s}} = \frac{1}{T}$. The quantisation or the digitisation of each taken sample in the amplitude occurs at the second step. The dynamic range of an ADC is divided into 2^N sub-ranges called the least significant bit (LSB), ΔV_1. The resolution of the ADC with the parallel output code, N, determines the number of output leads, as shown in Figure 4.15.

4.2.2 Quantisation and Thermal Noise of ADC

Due to a finite resolution the ADC is unable to distinguish different levels of the input voltage within the LSB. Therefore, the maximum error

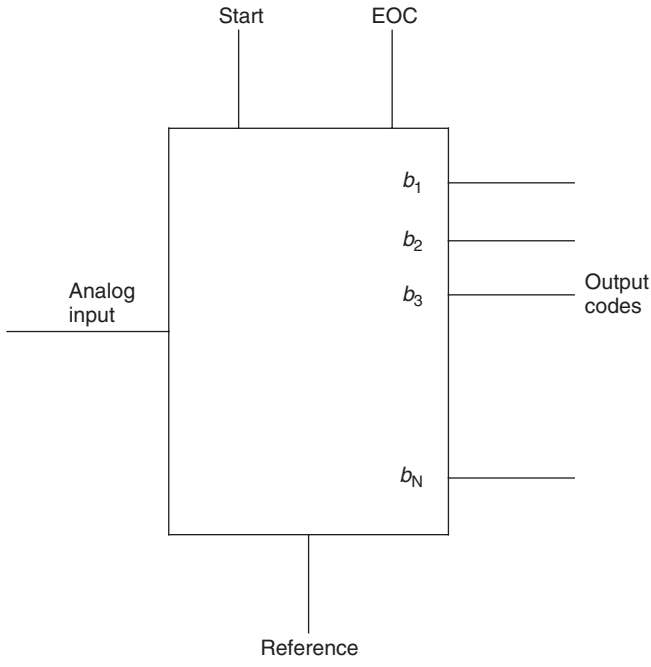


Figure 4.15 A functional diagram of the ADC

of an ideal device can be as much as $\pm\frac{1}{2}$ LSB. The difference between the actual value of the input voltage and a value of the assigned output code is called the quantisation error, q_N. Its example is shown in Figure 4.16. As can be seen, the error is an inherent feature of the A-to-D conversion process. In other words, it cannot be eliminated. At the end of this section, we will discuss briefly the sliding scale method, which allows some reduction of q_N at the expense of the total conversion time. Yet, the mainstream minimisation of the quantisation error is achieved by increasing the ADC resolution.

Let us assume that all channels of an ideal ADC have an equal probability to sample the input signal. In this case, for each channel in the range $\left[-\frac{\Delta V_i}{2} \text{ to } \frac{\Delta V_i}{2}\right]$ the probability distribution $p_i = \frac{1}{\Delta V_i} = \frac{1}{\Delta V}$, where $\Delta V = 2^{-N}$. The rms quantisation noise can be written as [13]

$$q_n^{rms} = \left[2\int_0^{\frac{\Delta V}{2}} px^2 dx\right] = \frac{\Delta V}{\sqrt{12}} \tag{4.2.2}$$

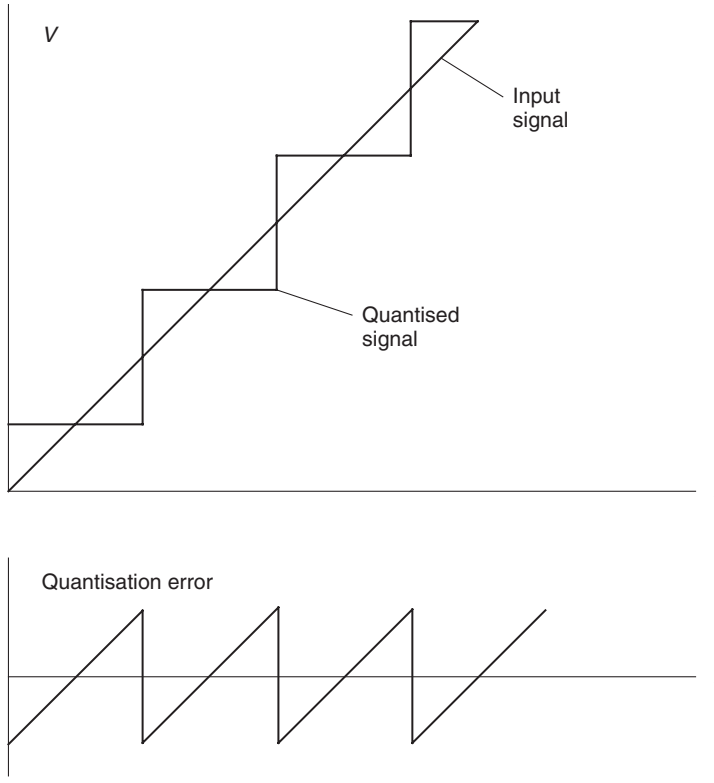


Figure 4.16 The quantisation error of the ADC

Similar result can be achieved if the error within the channel can be approximated by the linear function

$$e(t) = \Delta V\left(\frac{t}{T} - \frac{1}{2}\right) \tag{4.2.3}$$

as illustrated in Figure 4.17. In this case, the rms value of $e(t)$ is given by

$$q_{\mathrm{n}}^{\mathrm{rms}} = \left[\frac{1}{T}\int_0^T \Delta V^2\left(\frac{t}{T} - \frac{1}{2}\right)^2 \mathrm{d}t\right]^{\frac{1}{2}} = \frac{\Delta V}{\sqrt{12}} \tag{4.2.4}$$

Figure 4.18 illustrates the presence of the quantisation noise in the explicit form in the digital spectrometer as a source of the noise codes at the output of the ADC.

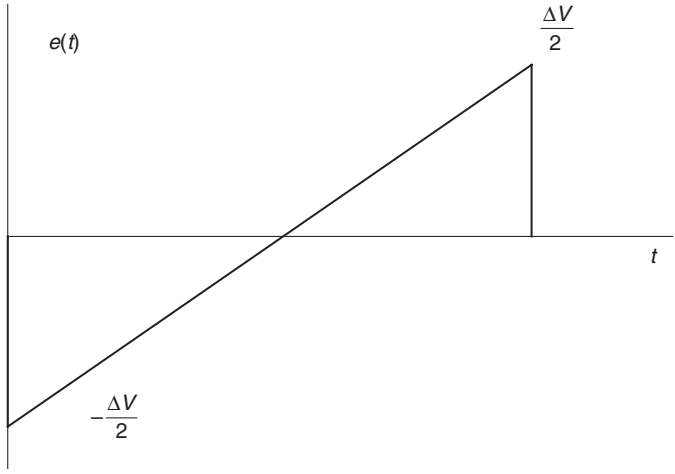


Figure 4.17 The linear approximation of the quantisation error within one channel

The signal-to-noise ratio of the ADC is defined as

$$\mathrm{SNR(dB)} = 20\log_{10}\left(\frac{V_{\mathrm{i}}^{\mathrm{rms}}}{q_{\mathrm{n}}^{\mathrm{rms}}}\right) \tag{4.2.5}$$

It depends on the shape of the input signal. In the literature, the figure is often quoted for the sinusoidal voltage, $V_{\mathrm{i}}^{\max}\sin(\omega t)$. This results in the SNR due to quantisation error given by

$$\mathrm{SNR(dB)} = 20\log_{10} = \left(\frac{V_{\mathrm{i}}^{\max}}{\sqrt{2}} - \frac{2^{N-1}\sqrt{12}}{V_{\mathrm{i}}^{\max}}\right) = 6.02N + 1.76 \tag{4.2.6}$$

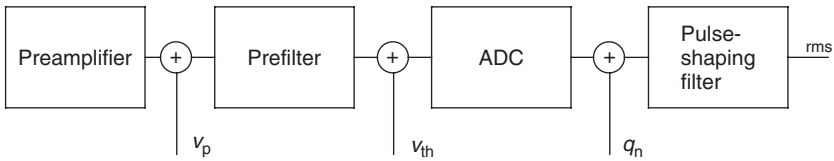


Figure 4.18 A functional block diagram of a detector spectrometer with noise sources: v_{p} is the noise of the preamplifier, v_{th} is the thermal noise of the ADC, and q_{n} is the noise generated by the quantisation error

Real ADCs have a number of other noise sources. These are the thermal noise, the shot noise and the flicker noise of comparators, the fluctuations of the reference voltage and passive components. The thermal noise, v_{nth}, can be represented by the equivalent input referred generator with the effective noise resistor, R_{eff}, so that

$$v_{\mathrm{th}}^{\mathrm{rms}} = \sqrt{4k_b R_{\mathrm{eff}} B_{\mathrm{eff}}} \tag{4.2.7}$$

where B_{eff} is the effective resolution bandwidth. B_{eff} is derived from the 3-dB attenuation point on the plot of SNR as a function of frequency.

An energy resolution that a detector system is expected to provide directly translates into the ADC gain and the dynamic range requirements. As we learned in previous chapters, the detector–preamplifier couple generates a pulse step with a height, V_{s}, in response to an incoming radiation photon. In addition, the signal contains a noise component with a mean square value denoted as V_{p}^2. After the digitisation, the total noise component in strings is given by the sum

$$V_{\mathrm{n}} = \sqrt{V_{\mathrm{p}}^2 + (V_{\mathrm{th}}^{\mathrm{rms}})^2 + (q_{\mathrm{n}}^{\mathrm{rms}})^2} \tag{4.2.8}$$

Neglecting for a moment the noise of ADC, we derive that the digitisation adds less than 1% to V_{p} if the condition

$$\frac{V_{\mathrm{p}}}{\Delta V_1} \geq 6.7 \tag{4.2.9}$$

is met. Equation (4.2.9) immediately sets up the gain of amplifier stages preceding the ADC. Sometimes it can be expressed in the units $\left[\frac{V}{\mathrm{bit}}\right]$ if ΔV_1 is replaced by 1 bit.

Once ΔV_1 has been defined then the ADC dynamic range is set by the need to cover the maximum voltage delivered by the preamplifier or the signal conditioning unit, V_{m}. A number of bits required to process analogue signals without the saturation errors is given by

$$N = \frac{\log\left(\dfrac{V_{\mathrm{m}}}{\Delta V_1}\right)}{\log(2)} \tag{4.2.10}$$

Example 4.1

Let us assume that a preamplifier generates a voltage step $V_s = 50\,\text{mV}$ corresponding to a photon with an energy of $E = 5.9\,\text{keV}$ (MnK_α). A detector system is expected to show at this line an energy resolution FWHM $\Delta E_{\text{FWHM}} = 20\,\text{eV}$. Based on this information we can derive the rms noise at the output of the preamplifier as

$$V_p = \frac{\Delta E_{\text{FWHM}} V_s}{2.35E} = \frac{20\,\text{eV}}{2.35 \times 5900\,\text{eV}} \approx 72\,\mu\text{V} \tag{4.2.11}$$

According to equation (4.2.9), one bit of the ADC resolution should correspond to approximately

$$\Delta V_1 = \frac{V_p}{6.7} \approx 10.7\,\mu\text{V}$$

The maximum voltage at the output of preamplifier $V_m = 1\,\text{V}$ gives the requirement for the dynamic range. Following the equation (4.2.10),

$$N = \frac{\log\left(\dfrac{1}{10.7 \times 10^{-6}}\right)}{\log(2)} \approx 16$$

Setting an inspection time of the fast discrimination channel, $T_F = 1\,\mu\text{s}$, we derive a sampling rate of the ADC as

$$R_{\text{ADC}} \geq 5\frac{1}{T_F} \approx 5\,\text{MsPs}$$

If such an ADC is not available then either the output voltage of the preamplifier should be conditioned to reduce V_m or the time constant of the discrimination channel made longer. For instance, the reduction of V_m to a level of 0.2 V would allow the implementation of a 14-bit ADC. Fourteen-bit ADCs exist in a large variety from chip manufacturers with a sample rate exceeding 100 Msps.

Like any other noise components, n_q enters the final noise budget and affect the signal-to-noise ratio of the spectrometer. If the probability distribution of samples is uniformly distributed within each quantisation

interval, the rms error generated by n_q at the output of the pulse-shaping digital filter takes the form [14]

$$\varepsilon_{rms} = \sqrt{\frac{1}{12}\frac{1}{2^{2N}}\sum_i w_i^2} \tag{4.2.12}$$

where w_i represent weighting coefficients of the nonrecursive digital pulse shaping filters.

Provided that the probability distribution of the quantisation error is Gaussian (which is true in the most of practical cases), its noise linewidth at the FWHM is given by

$$\mathrm{FWHM}_q = 2.36 \times \varepsilon_{rms} \tag{4.2.13}$$

The equation (4.2.12) works well only when the sampling is fast enough to accumulate sufficient statistics for accurate averaging. If this condition is met then the minimisation of the quantisation error is achieved by the mere increase of the ADC resolution and selecting filter algorithms with low weights w_i. If the number of accumulated samples is low, the quantisation error becomes variable and well above the value defined by (4.2.2), (4.2.4) and (4.2.12).

4.2.3 Non-linearities of ADC

After an ADC has passed the selection criteria on the resolution and dynamic range, it needs to be examined on the differential and integral non-linearities. The importance of this step results from the fact that non-linearities modify the entire filter weighting function and affect the final pulse classification in the energy histogram. Without adequate corrections, the modified weighting functions make the whole processing channel less efficient in terms of the signal-to-noise ratio with respect to all noise sources, the pile-up inspection, the final energy resolution, tailing and the peak stability.

4.2.3.1 Differential non-linearity

In all real ADCs, the probability distribution function for each channel is not uniform. This non-uniformity generates the differential non-linearity

(DNL). Assuming the maximum DNL in terms of the channel width to be δ, the probability distribution function takes the following modified form

$$p_{\mathrm{r}} = \frac{1}{\Delta V}(1 \pm \delta) \tag{4.2.14}$$

Substituting the equation (4.2.14) into (4.2.2), we derive the variance of the quantisation noise that takes into account the maximum differential non-linearity

$$(q_{\mathrm{n}}^{\mathrm{rms}})^2 = \frac{\Delta V^2}{12}(1 \pm \delta) \tag{4.2.15}$$

In reality, the maximum DNL does not quite characterise the performance of the ADC. Due to the statistical nature of the error it is better to utilise the variance of the DNL. Assuming that the probability distribution function of the differential non-linearity is uniform from $-\delta$ to δ its variation can be estimated as follows [9]

$$\mathrm{var}[\mathrm{DNL}] = \left[2\int_0^{\nu} px^2 \mathrm{d}x\right] = \frac{\delta \Delta V^2}{12\sqrt{3}} \tag{4.2.16}$$

where $|\nu| = \delta\frac{\Delta V^2}{12}$ is the function of the quantisation noise added due to the differential non-linearity.

The purpose of the discussion was to show that the differential non-linearity should be treated like any other source of noise discussed in the previous chapter. Therefore, its final impact has to be found at the output of the pulse-shaping filter using either the equation (4.2.12) or by adding to a, b, c components from the preamplifier in quadrature and calculating the overall signal-to-noise ratio, ENC and FWHM.

Figure 4.19(a) shows an example of the A-to-D conversion function with the presence of differential non-linearity.

4.2.3.2 Integral non-linearity

The integral non-linearity can be visualised with the help of Figure 4.19(b). Unlike DNL, it generates a stationary systematic error that can be corrected or compensated.

Let us assume that the maximum error is a μ fraction of the channel width. The error changes within an interval ranging from $-\mu\frac{\Delta V}{T}$ to $\mu\frac{\Delta V}{T}$. Its probability distribution due to the integral

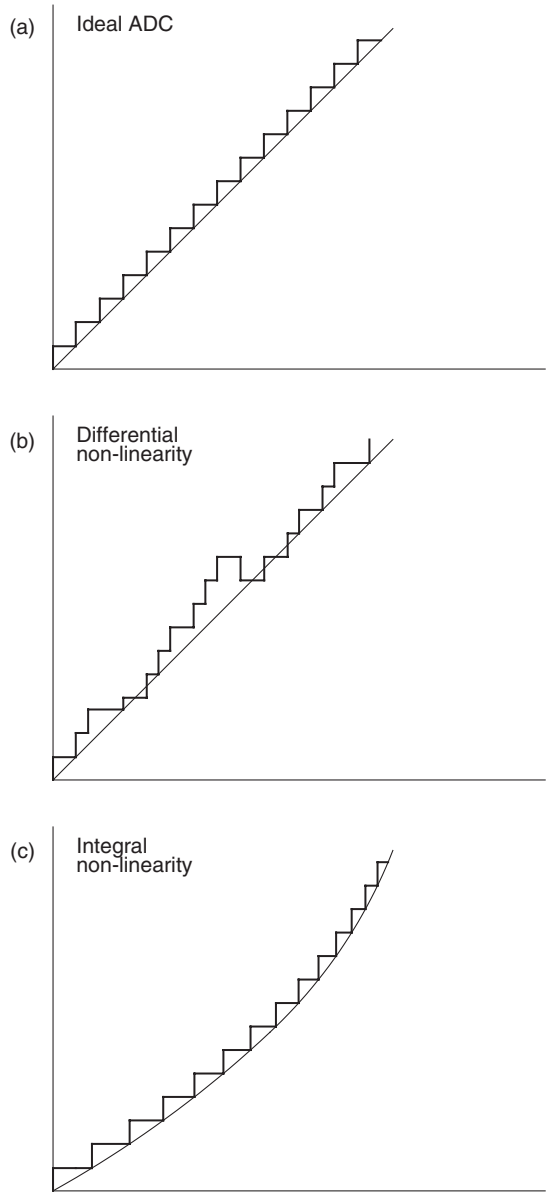


Figure 4.19 The analogue-to-digital conversion function of an ideal ADC (a), an ADC with a differential non-linearity (b) and an ADC with an integral non-linearity

non-linearity (INL) takes the form $p = \frac{T}{2\mu\Delta V}$, which results in the variance of the INL noise given by

$$\mathrm{var}[\mathrm{INL}] = 2\int_0^{\mu\frac{\Delta V}{T}} px^2 \mathrm{d}x = \frac{\mu^2 \Delta V^2}{6T^2} \tag{4.2.17}$$

where T is the integration time of the pulse-shaping filter. The integral non-linearity error referred to the input ADC appears as a series of step-like noise pulses with amplitude inversely proportional to the duration of the integration time. In this respect, the effect of INL on the final energy resolution is usually smaller than DNL especially at long integration times.

4.2.3.3 Experimental evaluation of differential and integral non-linearities

Manufacturers normally specify typical values of the DNL and the INL of the commercial ADCs. This enables us to derive a rough estimate of their influence on the energy spectrum histogram. If the influence is substantial, ADCs need to be characterised individually in order to correct for INL or/and select a chip with the best performance.

Figure 4.20 depicts a functional block diagram of a system measuring non-linearities of an ADC being a part of a digital spectrometer. The input

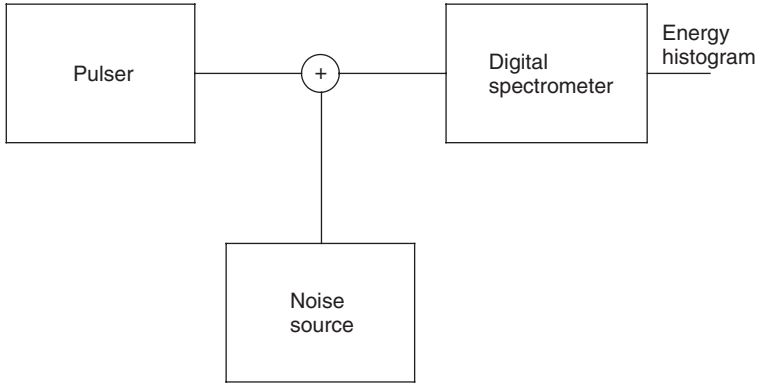


Figure 4.20 A functional block diagram of a system measuring non-linearities of an ADC being a part of a digital spectrometer. The input of the ADC is connected to two oscillators. The pulse oscillator simulates the detector output. It generates pulses with the adjustable amplitude. A calibrated noise source gives the measure for deriving absolute figures of the quantisation noise caused by the DNL and the INL

of the ADC is connected to two oscillators. The pulse oscillator simulates the detector output. It generates pulses with the adjustable amplitude. A calibrated noise source gives the measure for deriving absolute figures of the quantisation noise caused by the DNL and the INL.

At the beginning, the system digitises the zero input signal. This results in the "true" reference zero peak on the energy histogram. The next step is to sample pulses with an amplitude $n\Delta V$. The effect of non-linearities can immediately be quantified from the difference between the zero peak at $n\Delta V$ amplitude and the true zero peak in quadrature. Changing the pulse height in steps so that n takes values from 0 to N allows characterisation of the whole range of the ADC. In these measurements, the DNL will appear as an additional noise increasing the width of zero peak, whereas INL causes a peak shift as a function of n.

There are other techniques for evaluating the quality of ADCs, for instance, those using reference samples. Details of the measurement set-up and some experimental results can be found in [9].

4.2.4 Aperture Time and Uncertainty

The sampling theorem assumes that values of the input signal are captured instantly at well-defined time moments. This can be expressed as

$$f_s(n) = \int_{-\infty}^{\infty} f(t)\delta(t - n\tau)\mathrm{d}t \tag{4.2.18}$$

where $\delta(t - n\tau)$ is the Dirac delta function, $f(t)$ is the input signal and $f_s(n)$ is the sampled function $F(t)$. In practice, however, the sampling process takes a certain period of time and there is an uncertainty in an actual instant at which samples have been taken. Following [15], the sampling function can be represented in a modified form as follows

$$x_s(n) = \int_{-\infty}^{\infty} f(t)\Psi(t - n\tau)\mathrm{d}t \tag{4.2.19}$$

where $\Psi(t)$ is already a continuous function of time, which includes random variable responsible for the jitter of the sampling moments.

Both the phenomena contribute to the signal degradation. In fact, they may become dominant sources of noise in the digitisation process with high resolution ADCs.

4.2.4.1 The effect of non-zero aperture time

A simplified sample-and-hold circuit is shown in Figure 4.21. In this circuit, R is the output resistance of the follower 1. When the switch is closed the output voltage takes the following form

$$V_{\text{out}}(t) = \int_{-\infty}^{t} V_{\text{in}}(\tau) e^{\frac{\tau - t}{RC}} \mathrm{d}\tau \tag{4.2.20}$$

which implies that the function $\Psi(t)$ is defined as

$$\Psi(t) = e^{\frac{t}{RC}}, \quad t < 0, \qquad \Psi(t) = 0, \quad t > 0$$

Expanding functions f_s and x_s as Taylor series and subtracting one from another we can write down the expression for the error caused the non-zero aperture time, t [15]:

$$\varepsilon_{\text{nza}}(t_0) = f'(t_0) \int_{-\infty}^{0} t e^{\frac{t}{RC}} \mathrm{d}t \tag{4.2.21}$$

According to equation (4.2.21), the error generated by the non-zero sampling time is proportional to the product of the slew rate of the input signal and the aperture time.

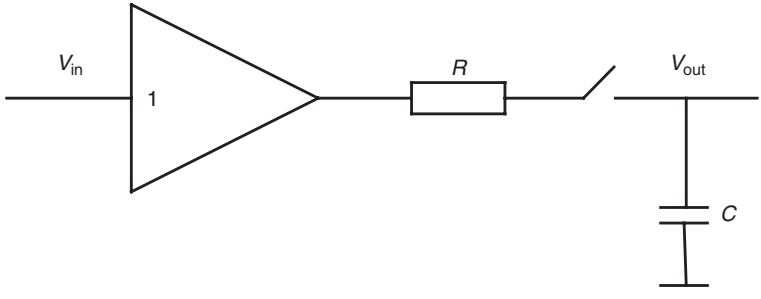


Figure 4.21 A simplified circuit diagram of the sample-and-hold circuit

4.2.4.2 The effect of aperture uncertainty

The sampling instant uncertainty is often called the aperture uncertainty or the aperture jitter. The aperture uncertainty is the more inclusive term that takes into account all possible causes including noise, variations in the signal amplitude-dependent delay (pronounced in the flash ADCs), temperature, the clock instability and so on. The aperture jitter is usually meant to result from the thermal noise only.

The presence of the aperture jitter can be described using the sampling function in the form

$$\Psi(t) = \delta(t + t_{\mathrm{j}}) \tag{4.2.22}$$

where t_{j} represents the uncertainty in the sampling instant. Following the same procedure as that for the non-zero aperture time, we derive the error generated on each sample

$$\varepsilon_{\mathrm{aj}} = f'(t_0)t_{\mathrm{j}} \tag{4.2.23}$$

In other words, the error due to the jitter is proportional to the slew rate of the input signal multiplied by the jitter time.

In order to characterise the quality of the conversion we must average error voltages given by the equation (4.2.23) over the whole sampling period of the input signal. The rms value of the voltage error will be the product of the rms value of the slew rate and the rms value of the time jitter, as follows

$$\varepsilon_{\mathrm{aj}}^{(\mathrm{rms})} = f'(t_0)^{(\mathrm{rms})} t_{\mathrm{j}}^{(\mathrm{rms})} \tag{4.2.24}$$

Since the aperture uncertainty acts as a wideband noise on the clock, it appears at the digitisation stage as wideband noise due to mixing effect. The spectrum of the noise is periodic and repeated around the sample rate. In this respect, the clock stability becomes of paramount importance in order to reduce the degradation of the noise floor performance of the ADC. However, even an ideal clock does not eliminate this noise source completely.

The theoretical signal-to-noise ratio for an ADC limited by the aperture uncertainty is given by [16]

$$\mathrm{SNR}_{\mathrm{au}} = -20\log_{10}\left[2\pi f_{\mathrm{analog}} t_{\mathrm{j}}^{(\mathrm{rms})}\right] \tag{4.2.25}$$

where f_{analog} is the frequency of the input signal.

Example 4.2

Let us assume that we sample a full-scale sinusoidal signal with $f_{\text{analog}} = 10\,\text{MHz}$ with the rms aperture jitter of 3 ps. Then, according to the equation (4.2.23), the expected error would limit the signal-to-noise ratio to a value of 74.5 dB. This corresponds to a signal-to-noise ratio of an ideal 12-bit ADC which has the quantisation noise only.

4.2.5 Reduction of Differential Non-Linearity with Sliding Scale Method

Sliding scale method was proposed first in [17]. This ingenious technique enables the substantial reduction of the differential non-linearity in all types of ADCs. Sometimes it is called "dithering". The principle of the method is to add a small noise to the input signal, V_{I}, of an ADC and subtract it digitally from the output strings. In this case, even for impulses of a constant height, V_{I}, the coding will be spread over a number of channels due to the added noise. The statistical distribution of the final channel-number numbers will be determined by a weighted average over part of the ADC range and will, therefore, be less sensitive to the width of an individual channel. Assuming the channel profiles to be rectangular (an ideal ADC) and that the width of a kth channel is ΔV_k, then the effective width of the kth channel after smoothing, $\Delta V'_k$ is given by

$$\Delta V'_k = \frac{1}{m+1} \sum_{l=0}^{m} \Delta V_{k+1} \tag{4.2.26}$$

where m is a number of bits which are necessary to cover the peak-to-peak amplitude of the added random signal. According to equation (4.2.26), the variations in the width of the channels will be reduced by a factor of $(m+1)$.

An example of a practical implementation of this method is shown in Figure 4.22. The system includes a random code generator that is transformed into the voltage noise with a digital-to-analogue converter (DAC). In doing so, noise codes are immediately available for the digital subtraction. Thus, the method introduces negligible time delay into the total processing channel. Yet, there is a price to pay for the sliding scale

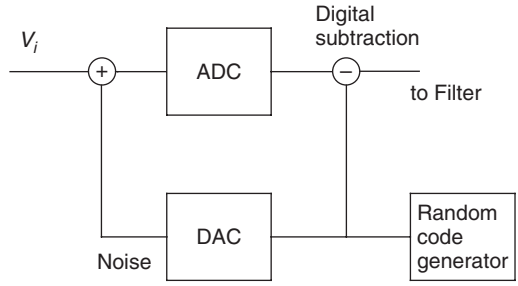


Figure 4.22 An example of a practical implementation of the sliding scale method. The system includes a random code generator that is transformed into the voltage noise with a digital-to-analogue converter (DAC). In doing so, noise codes are immediately available for the digital subtraction. The method introduces negligibly small time delay into the total processing channel. However, an ADC is required to have up to m extra bits of the resolution to accommodate the smoothing random signal

effect. An ADC may be required to have up to m extra bits of the resolution to accommodate the smoothing random signal.

It was shown in [18] that the quantisation error can also be significantly reduced if the arrival time of pulses is not synchronised with the ADC acquisition time. In this case, the output of the digital filter is not a single value for a fixed amplitude of the input pulse plus the quantisation noise, but a quasi-Gaussian distribution around a some average value. In other words, the random arrival time acts as a low-pass filter for the quantization noise. It can be numerically evaluated using (4.2.12) with appropriate indexes. An improvement achieved with this method can be counteracted by a reduction in the signal-to-noise ratio due to the aperture uncertainty. This will certainly happen if an ADC must sample input voltages with high slew rates.

To summarise topics discussed in this section, we will write down a generalised equation for the signal-to-noise ratio of a practical ADC that takes into account the thermal noise, the differential non-linearity and the aperture jitter [16]:

$$\mathrm{SNR}_{\Sigma} = -20\log\left[\left(2\pi f_{\mathrm{analog}} t_{\mathrm{j}}^{(\mathrm{rms})}\right)^2 + \left(\frac{1+e}{2^N}\right)^2 + \left(\frac{V_{\mathrm{noise}}^{(\mathrm{rms})}}{2^N}\right)^2\right]^{\frac{1}{2}} \quad (4.2.27)$$

where e and $V_{\mathrm{noise}}^{\mathrm{rms}}$ represent the average differential non-linearity and the thermal noise of a converter both expressed in units of the LSBs.

4.3 DIGITAL RISE (FALL) TIME DISCRIMINATION

The major factors responsible for deviations in the pulse rise time at the output of superconductor detectors result from a relatively long diffusion time, τ_d. This makes the rise time position sensitive. For instance, in the TES microcalorimeters, variations in τ_d lead to variations in a fraction of the photon energy lost into the substrate and current leads. In the particle detectors, τ_d causes an incomplete charge collection due to a finite lifetime of particles. The incomplete charge collection means that only a fraction of created quasiparticles are swept to the counter-electrode and/or to the preamplifier input. As a result, the charge (current) signal amplitude is lower than expected and the energy peaks on the histogram shift towards lower values producing the measurement error. It also often appears as a tail on the low energy side of the peak. An example of such a spectrum is shown in Figure 4.23.

There are two other mechanisms that have effect on the incomplete charge collection in the particle detectors with a superconducting absorber. These are trapping quasiparticles and escaping high-energy phonons into the substrate, counter-electrode and contact leads. Trapping centres are essentially volumes with a suppressed energy gap created by imperfections of the crystalline structure and/or by Abrikosov vortices. Both types of trapping centres tend to be correlated as the same crystalline imperfections form pinning centres for vortices. In the large area absorbers, the film thickness and surface defects may also seriously influence the diffusion dynamics of quasiparticles.

The incomplete charge collection broadens primarily low energy peaks on the final histogram. However, worsening the energy resolution is not the only consequence of the phenomenon. Non-Gaussian peak shapes

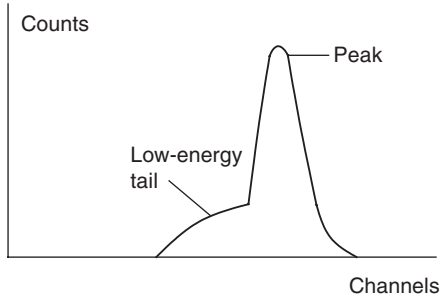


Figure 4.23 An example of the energy spectrum, in which a peak is misshaped by the incomplete charge collection. Some loss mechanisms can generate a tail on the low energy side of the peak

further distorted by the presence of tails make existing fitting algorithms and together with them the whole quantitative analysis erroneous.

One universal way to reduce the incomplete charge collection is to collimate detector area to a size when a quality of the energy spectrum becomes acceptable. Collimators are normally used with all types of detectors (superconductor and non-superconductor) to eliminate edge effects. Apart from that, any further reduction in the sensitive area is a serious drawback and necessarily narrows the detector applicability. An X-ray optics helps, to some extend, in restoring the solid angle. At the moment of preparing this book, the optics is quite expensive and commercially not available in all the variety required.

The low energy tails that remain after the collimation can be further reduced electronically using rise time discriminators. There are two types of the rise time discriminators:

1. those that analyse the leading edge of the pulse; and
2. those measuring the cross-correlation factor between the shape of an incoming pulse and a written in memory expected pulse shape.

In this section, we will briefly discuss both techniques.

4.3.1 Rise Time Discriminators Analysing the Leading Edge of the Pulse

An example of the functional block diagram illustrating the technique is shown in Figure 4.24. The rise time is estimated using the counter-routine. It amounts to a number of clock intervals between the leading

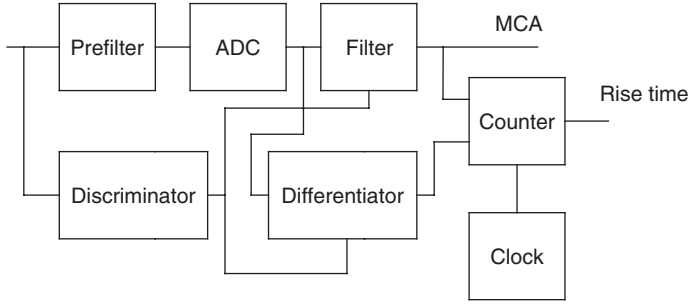


Figure 4.24 A functional block diagram of a spectrometer employing the rise time discrimination technique based on the analysis of the leading edge of the pulse

edge of the pulse and its maximum at the output of the pulse-shaping filter or another specially designed filter. Measuring the leading edge may become a formidable task for some pulse shapes. Following a standard procedure of mathematical analysis of complex function, a double differentiation of the peak helps to identify the steepest region.

4.3.2 Cross-Correlation Pulse Shape Discriminators

An implementation of this type discriminator in the digital spectrometer is illustrated in the functional block diagram in Figure 4.25. The cross-correlation factor between an incoming pulse and an expected pulse shape is given by [19]

$$C_{xz} = \frac{\frac{1}{M}\sum_{j=1}^{M}(x_j - \bar{x})(z_j - \bar{z})}{\sigma_x \sigma_z} \tag{4.3.1}$$

where $\sigma_x^2 = \frac{1}{M}\sum_{j=1}^{M}(x_j - \bar{x})^2$ and $\sigma_z^2 = \frac{1}{M}\sum_{j=1}^{M}(z_j - \bar{z})^2$ with x_j and z_j representing the digitised input pulse and the predetermined pulse shape respectively. The cross-correlation factor C_{xz} ranges from $+1$ for the perfect correlation to -1 for the perfect anticorrelation. A practical acceptance value varies and depends on a type of application.

This technique is particularly useful for STJ detectors often showing double energy peaks for monochromatic X-rays, which correspond to the absorption of phonons at the upper or lower electrode. The absorption of phonons by different electrodes affects slightly the rise time itself [20]. Thus, the rise time discriminators can detect variations in the

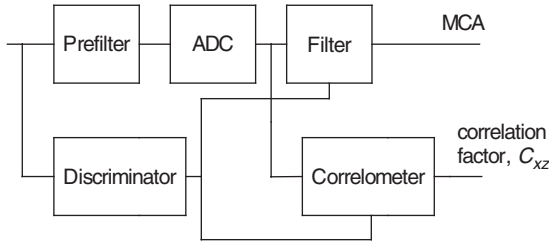


Figure 4.25 A functional block diagram of a spectrometer employing cross-correlation pulse shape discriminators

delays only in the cooperation with very fast current preamplifiers. Otherwise, the cross-correlation technique is more efficient, of course, at the expense of extending the processing time.

It will not be exaggerating to state that the loss mechanisms in superconductor detectors affect every single pulse arriving at the input of the preamplifier. Therefore, the rejection of identified pulses is not an option, as this would reduce the throughput of the detector system. Instead, the digital spectrometer must implement some kind of the amplitude correction algorithm. The development of such algorithms is a creative process sometimes requiring a deep knowledge of the detector physics on the microscopic level, as well as details of an intended application. The prime task of any correction procedure is to distinguish between the incomplete charge collections caused by different loss mechanisms, as each of them may require dissimilar correction factors. A practical way to do that is to use the bi-parametric representation of the preamplified pulses proposed in [21]. On the bi-parametric plot, each pulse is represented by a dot with the co-ordinates defined as the pulse rise time and the channel number (Figure 4.26). Different mechanisms responsible for the incomplete charge collection are expected to form defined lines, curves or densely populated areas. The heights of pulses that belong to a line or a curve can be corrected using a fitted function derived at a calibration stage with a known monochromatic

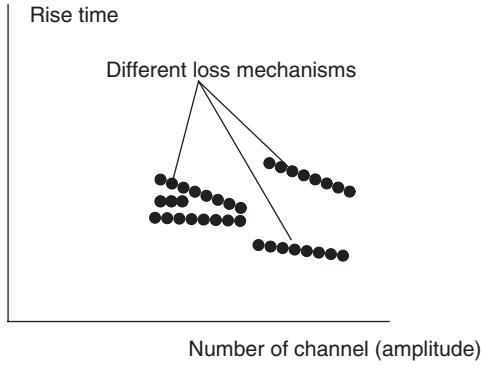


Figure 4.26 A bi-parametric representation of the preamplified pulses proposed in [21]. On the bi-parametric plot, each pulse is represented by a dot with the co-ordinates defined as the pulse rise time and the channel number. Different mechanisms responsible for the incomplete charge collection are expected to form defined lines, curves or densely populated areas. The heights of pulses that belong to a line or a curve can be corrected using a fitted function derived at a calibration stage with a known monochromatic radiation source

radiation source. In the electron-probe microanalysis, the incomplete charge correction algorithms will be a part of the matrix corrections adapted for superconducting detectors [22].

As an example, the position sensitivity of the superconductor detectors with large absorbers was analysed in [23]. Assuming a constant loss rate, the incomplete charge collection as a function of the distance from the point where the interaction of the incoming photon with the absorber takes place to the collection electrode of the STJ produces a straight line on the bi-parametric plot. In this case, we can apply a linear sliding method proposed in [21]. According to the method, the correction factor has the following form

$$\alpha_{\text{cor}} = \frac{A_{\text{cor}}}{A} = m - \frac{t_{\text{rise time}}}{Am} \tag{4.3.2}$$

where A is the amplitude of the measured pulse, A_{cor} is the expected amplitude, $t_{\text{rise time}}$ is the rise time of the pulse derived from the digital processing algorithm and m is the slope of the fitting line of the bi-parametric function.

As we mentioned before, practical bi-parametric functions often cannot be fitted with straight lines. Therefore, the amplitude correction procedures as well as the correction factors take more complex forms. There is a number of publications where one can find generic approaches and specific examples in dealing with the problem in semiconductor detectors. Interested readers should refer, for instance, to [6] and [24] for further details.

4.4 SUPERCONDUCTOR DIGITAL SPECTROMETER

Superconductor electronics has a full range of components to reproduce the functionality of practically any analogue or digital device available in semiconductors. Earlier, in Chapter 3, we discussed high-resolution analogue current amplifiers based on DC SQUIDs. Similar to diodes and transistors, the Josephson junctions on their own and single- or multiple-junction SQUIDs can also operate as digital components. In fact, there are even two types of the superconductor logic families:

1. The voltage latching logic in which Josephson gates switch between a superconducting state and a voltage state corresponding to the logical states “0” and “1” [25]. The functionality of the gate is illustrated in Figure 4.27.

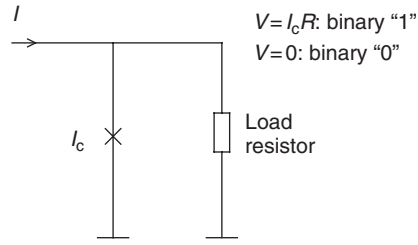


Figure 4.27 An example of the voltage latching logic in which a Josephson gate switches between a superconducting state and a voltage state corresponding to the logical states "0" and "1" [25]

2. Single flux quantum non-latching logic operating with very short voltage pulses.

Figure 4.28 explains the second principle. It shows an interferometer incorporating a single Josephson junction with the critical current I_c. A magnetic flux, Φ_e, applied to a single junction interferometer evokes a supercurrent I circulating in the loop. Once the supercurrent exceeds a critical value I_c, the junction switches into a normal state to let the flux quantum enter the interferometer. After the flux quantum has entered the loop the junction returns to the superconducting state. During this

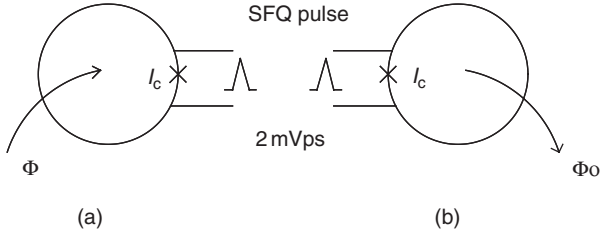


Figure 4.28 A single junction interferometer representing a component of the single flux quantum non-latching logic operating with very short voltage pulses. A magnetic flux, Φ_e, applied to a single junction interferometer evokes a supercurrent I circulating in the loop. Once the supercurrent exceeds a critical value I_c the junction switches into a normal state to let the flux quantum enter the interferometer. During this transient process, the interferometer generates a single flux voltage pulse (a). A single flux quantum (SFQ) pulse applied to the junction changes the magnetic flux inside the loop by the Φ_0 (b). In other words, the single junction interferometer can perform a function of a basic superconductor analogue-to-digital and digital-to analogue converters

transient process, the interferometer generates a single flux voltage pulse with the area given by the Faraday's law as

$$\int_0^t V(t)\mathrm{d}t = \Phi_0 \equiv 2\,\mathrm{mVps} \tag{4.4.1}$$

The logic uses dynamic representation of information. The arrival of the single flux quantum pulse between two consecutive clock pulses means the binary "1", otherwise it is the binary "0".

The equation (4.4.1) works both ways. A single flux quantum (SFQ) pulse applied to the junction changes the magnetic flux inside the loop by the Φ_0 (Figure 4.28). In other words, the single junction interferometer can perform a function of a basic superconductor ADC and DAC.

The duration of the SFQ pulse represents the aperture time of the ADC. Being very short (1–2 ps), it sets up the fundamental limit for the product of resolution and the bandwidth, which is much higher compared to that of conventional semiconductor ADCs. It is also expected that superconductor digital spectrometers will be able to eliminate the preamplifier stage from the analogue chain and digitise directly the detector output. The development of such ADCs (sometimes they are called digital SQUIDs) was pioneered by Fujimaki in 1988 [26]. They employ the delta modulator architecture, as shown in Figure 4.29. The concept was practically abandoned by the semiconductor industry in favour of the delta–sigma ADCs because of the high differential nonlinearity and missing codes. But with a very short aperture, a limited bandwidth required for superconductor detectors and oversampling technique, it may provide adequate performance.

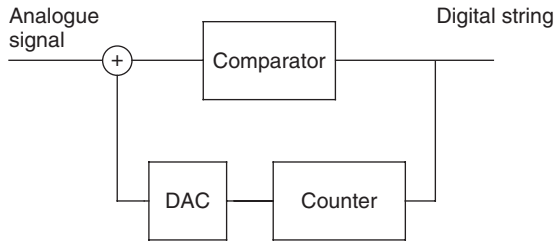


Figure 4.29 A functional block diagram representing a superconducting ADC, which normally employs the delta modulator architecture. Sometimes, such an ADC is called a digital SQUID amplifier. The digital SQUID will combine functions of the preamplifier and the ADC converting directly the output signal of a superconducting detector into a digital form

An operation at a low temperature, which would be considered as a serious drawback for a general purpose device, becomes one of the most attractive features in the digital spectrometer processing pulses of cryogenic detectors. It becomes almost a necessity when operating with large format detector arrays.

Superconductor digital spectrometers are still at an early development stage. The functionality of all principal components has been proven experimentally for both the types of logic. The progress in the field has been captured in a number of excellent review papers. The most recent of them are [27–29].

4.5 SELECTED TOPICS ON THE HARDWARE DESIGN

In this section, we will examine some hardware design techniques that addresses the non-ideal behaviour of electronic components, matching issues between different parts of the circuitry, grounding, electromagnetic interference and so on. A distinctive feature of these and other similar design aspects is the fact that they are invisible for standard design tools. In other words, they do not enter directly into any of the equations describing the system functionality. Yet neglecting them at the development stage proves to be very costly: running late and well above the budget projects or the acceptance of a degraded performance. For instance, the incorporation or reshaping the ground plane may require re-laying poly chlorinated biphenyls (PCBs), whereas a wrong grounding or the need in an extra shield will probably result in alterations of the mechanical construction of a prototype.

The unpredictability and irreproducibility of the phenomena involved make the discussion of the subject quite difficult. Logical sequences of equations do not exist and, even if invented, will be of a little assistance in the majority of practical situations. In what follows, we will try to formulate some generic hardware design rules. We will pay particular attention to those that are applicable to digital spectrometers operating with superconducting detectors.

4.5.1 Noise Reduction in Systems with Switching Power Supplies

Switching power supplies (SPS) have become a common choice in the modern electronics. They are more than twice as efficient compared to

linear power transformers. The high efficiency means that they dissipate less power, do not require bulky heat sinks, are of smaller size, lighter weight and, finally, lower cost. These benefits come with series drawbacks, most notably adding a large noise to the output voltage. Spikes generated by switching process create harmonics that extend over a broad band of frequencies. It is common to measure their traces even at 100 MHz and higher. The noise propagates through connecting leads as an electrical signal as well as through the air as the electromagnetic radiation. If no measures are taken, the peak-to-peak amplitude of noise is such that it will certainly affect the sensitive analogue circuitry of detector systems. The high-frequency radiation can suppress the functionality of superconductor devices incorporating Josephson junctions.

Whenever we have to deal with an interference problem it is advisable to break it into three major components. One needs to locate *the source* of noise, *the conduction–radiation paths* and *the receipting points*. The minimisation of the overall effect of the interference on the system performance assumes primarily working with each of these factors on individual basis. In what follows, we will itemize some of the generic recommendations:

1. Design around *the source* should include filtering, proper grounding and perhaps some local shielding.
2. A proper grounding, shielding and the physical separation of the source from a receipting point can attenuate the radiation and conduction paths.
3. A filtering of power lines, grounding, the use of amplifiers with differential inputs as well as differential outputs rejecting the common mode signals help to improve immunity of the low noise analogue circuitry.

Further, we will discuss all outlined techniques in more details.

4.5.1.1 Filtration

An SPS, as a part of the spectrometer, includes two types of filters. The first one attenuates the interference signal propagating through the AC power line and the second one reduces the noise at the output of the supply. The configuration is illustrated schematically in Figure 4.30. The power line filter represents a conventional single- or multiple-stage *L*, *C*, *R* filter with some variations, which are governed by particular applications. They are designed to combat the common mode and differential

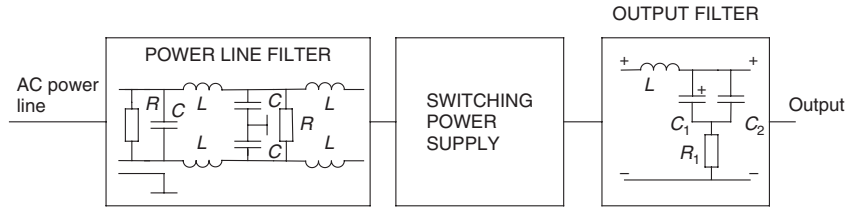


Figure 4.30 A functional block diagram of a switching power supply including two types of filters. The front filter attenuates the interference signal propagating through the AC power line and a second one reduces the noise and ripples at the output of the supply

mode spurious signals. The common mode signal is the signal that is found on both power lines with the same amplitude and polarity. The differential mode signal is present only on one of these lines. A series inductances, L, create a high impedance path for the common mode noise whereas capacitances act only against the differential mode noise. Resistors damp possible LC resonance modes.

Experimenting with power lines is dangerous and, therefore, should be avoided by the unauthorised personal. Efficient power line filters are commercially available in very compact and convenient packages combined with the entrance socket for the power line. Some extra filtering of the high-frequency common mode noise can be achieved by using a choke made of essentially a few turns of the power cable wound around a large ferrite.

The output filter includes two types of capacitors connected in parallel: a polarised electrolytic capacitor, C_1, and a non-polarised ceramic capacitor, C_2. This is due to the fact that starting from a certain frequency each capacitor departs from its ideal behaviour caused by the presence of the equivalent series resistance (ESR) and the equivalent series inductance (ESL). ESL is normally characterised by the frequency at which the component switches from a capacitive to inductive response function. Figure 4.31 shows a SPICE model of the real capacitor network. A rough guidance to the capacitance selection is given in Table 4.1. According to the Table, the electrolytic capacitors help to combat the low-frequency noise components, whereas the multilayer ceramic capacitors rectify the high-frequency part of the noise spectrum. Tantalum and ceramic capacitors are available in small sizes suitable for the surface mount.

An additional filtration of the power supply output is achieved by employing linear voltage regulators. This must be traded off against the power supply efficiency. Even a type of a chosen linear regulator may

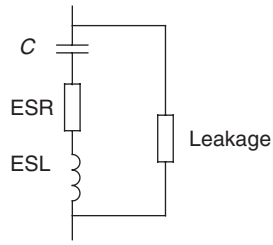


Figure 4.31 A SPICE model of a real capacitor network. Starting from a certain frequency each capacitor departs from its ideal behaviour caused by the presence of the equivalent series resistance (ESR) and the equivalent series inductance (ESL). ESL is normally characterised by the frequency, at which the component switches from a capacitive to inductive response function

Table 4.1 Capacitor performance as a function of frequency

	Aluminium Electrolytic	Tantalum Electrolytic	Ceramic Multilayer
ESR	~1 Ω at 100 kHz	~0.1 Ω at 100 kHz	~0.1 Ω at 1 MHz
ESL	~100 kHz	~1 MHz	1 GHz

affect the trade-off. For instance, low dropout devices do not reject noise at frequencies above a few kHz.

All sensitive analogue and digital active components normally include localised decoupling filters attached closely to power supply leads. They are intended to further suppress high-frequency external ripples and to provide the return path for AC current generated by the component itself. Signal lines can also contain low-pass filters. Often these filters are integrated with standard D-type or round connectors. These filtered connectors are also available commercially. It is possible to order a customised design at an extra cost when each line will have a specified corner frequency.

4.5.2 PCB Layout

As we mentioned earlier, designing a low-noise detector system must include a stage at which all potential sources, paths and sensitive receptors of electromagnetic signal are identified. It is advisable at this stage to divide the circuitry into the smallest functional units and assign them to one of layout blocks. These blocks can be, for instance, low level analogue signals, high level analogue signals, low current digital signals,

high current digital busses and so on. From the previous section we learned that the sampling clock in digital spectrometers requires a special attention to avoid the jitter. The clock is as perceptive to the noise as the low-level analogue signals, but causes as much trouble as any other digital signals. Therefore, associated with it circuitry should, if possible, be kept separate from other blocks as well. Once it is done, the next step will be to organise individual ground planes, power lines/planes for each of these blocks. The best option is to use a multilayer PCB design so that at least one complete layer is dedicated to the ground plane. The complete ground plane layer acts as low-impedance return path for AC currents generated locally by active components and minimises emissions of the high-frequency radiation by generating mirror images of currents flowing in signal tracks. Signal traces embedded between the ground and power planes have the minimum inductance. In fact, the arrangement forms high-frequency transmission lines. The return current path for high-frequency signal on a trace is located directly above and below the trace on the ground and power planes. The high-frequency signal is contained entirely inside the PCB.

Transmission lines must be properly terminated to avoid reflections. Their characteristic impedance is given by

$$Z_0 = \frac{87}{\sqrt{\varepsilon_r + 1.41}} \ln\left[\frac{5.98d}{0.89w + t}\right] \tag{4.5.1}$$

where ε_r is the dielectric constant of a material separating metallic layers, d is the thickness of the board between metal layers (mm), w and t represent the width and the thickness of metal trace respectively, in mm. The propagation velocity of the signal has the form

$$v_0\left[\frac{m}{ns}\right] \approx \frac{1}{42\sqrt{0.475\varepsilon_r + 0.67}} \tag{4.5.2}$$

If the ratio of track length to the rise time of the pulse exceeds the propagation velocity defined by (4.5.2), the track should be treated as a transmission line and requires the termination on both sides.

The debugging and troubleshooting of the multilayer PCB layout is virtually impossible. Besides, a low-quality dielectric material with the variable dielectric constant separating layers creates extra parasitic capacitances that may affect the stability of some amplifiers. High-frequency low-level oscillations are particularly difficult to detect as attaching an oscilloscope probe to the amplifier terminals changes

oscillating conditions. Therefore, several modifications of the PCB layout should be envisaged at the design stage to address the problems. A generic example of the PCB layout is shown in Figure 4.32. We can see that the system has two separate analogue and digital ground planes. Both the planes are kept separate until they reach the "star" ground point. Paths connecting individual grounds to the "star" ground point are vital for the minimisation of the interaction between different blocks. The paths are made of either multiple bus bars or wide copper brads for the minimum resistance and inductance. The back-to-back Schottky diodes are often used between grounds to prevent accidental transient voltage gradients that could damage on-board components.

It is obvious that some electronic devices do not belong entirely to one of the specified layout blocks. A good example of such a device is the ADC that, by definition, must combine both analogue and digital circuits. Since the ADC is a very important part of the digital spectrometer we will discuss their driving, layout and grounding in more details.

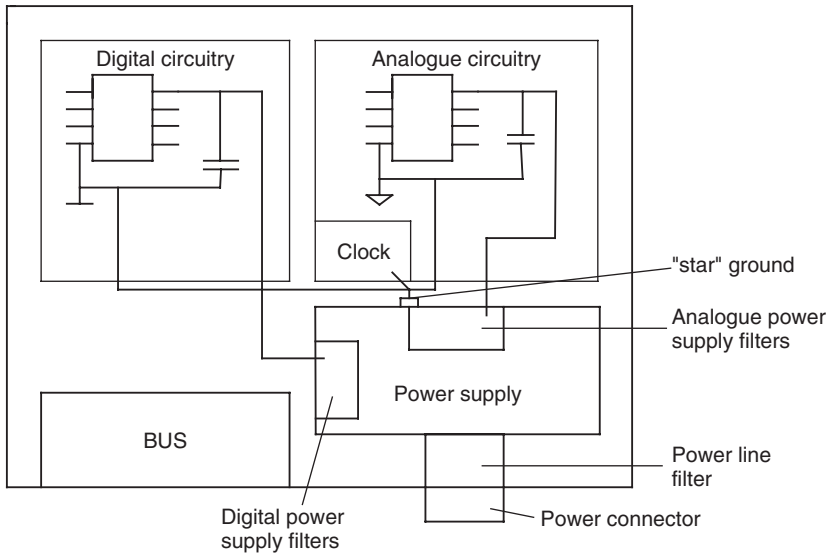


Figure 4.32 A generic example of the PCB layout. The system has two separate analogue and digital ground planes. Both the planes are kept separate until they reach the "star" ground point. Paths connecting individual grounds to the "star" ground point are vital for the minimisation of the interaction between different blocks. The paths are either made of multiple bus bars or wide copper brads to minimise their resistance and inductance

4.5.3 Layout, Decoupling and Grounding of ADCs

High-performance ADCs such as the AD7671 with the true bipolar input and AD7677 with the true differential input (16-bit resolution, 1-MHz bandwidth) have five specified ground pins. These are:

1. INGND to sense the analogue input signal;
2. REFGND belongs to the reference voltage;
3. AGND is the analogue circuitry ground;
4. DGND is the ground pin for digital circuitry of the ADC; and
5. OGND is the input–output interface and digital power ground.

This variety of available ground terminals gives a designer enough flexibility to achieve in practice the best performance of the devices by assigning their separate parts to different layout blocks. What is important that the assignment can change depending on an application and a surrounding circuitry and a layout. For instance, if a PCB has relatively small dimensions and an ADC is placed in close proximity to the "star" ground, all grounds can be divided into three categories. The first one combines INGND, REFGND and AGND that are joined together externally to the analogue ground plane. The second one, DGND, is attached to the digital ground plane. It is advisable to always keep separate the power digital grounds including the OGND from the rest of the circuitry.

In general, the name "DGND" on an ADC means literally that this pin connects to the digital part of the device circuitry rather than it must be attached to the digital ground plane of the detector system. In fact, running an ADC with separate grounds can inject a substantial amount of the digital noise into the analogue part of the device via stray capacitances, especially if it involves long ground tracks. In this case, a manufacturer would rather suggest to tie the DGND pin to the analogue ground plane as well with the minimum lead length [30]. Any extra impedance in the DGND connection causes a rapid deterioration in the ADC performance.

When the AGND and DGND pins are joined together, it is vital to prevent the ADC output driving high current loads. Using an isolation buffer latch reduces the output currents and the back reaction of the noise on the digital ground and the data bus on the analogue part of the system.

The presence of some digital noise and a certain amount of electromagnetic interference signal on the ADC input pins is inevitable. Employing ADCs with true differential inputs can further reduce the

common mode component of these parasitic signals. This technique, however, places particularly stringent requirements on the dynamics and noise of the buffer converting single-ended output of the preamplifier into the differential ADC input. In order to maintain the overall system accuracy, its noise, settling time and total harmonic distortion (THD) must be considerably better than that of the ADC itself. A 180° phase difference must be maintained with a good stability within the ADC analogue bandwidth. The net signal-to-noise degradation due to a signal-ended output-to-the-differential input buffer is given by [30]

$$SNR_{\text{loss}} = 20\log\left(\frac{N}{\sqrt{N^2 + \pi f_{-3\text{dB}}\left(\frac{2.5e_{\text{n}}}{V_{\text{i}}^{\max}}\right)^2}}\right) \tag{4.5.3}$$

where $f_{-3\text{dB}}$ is the 3-dB bandwidth of the ADC in MHz (or the cut-off frequency of the Nyquist pre-filter), e_{n} is the equivalent input noise voltage spectral frequency of the operational amplifiers utilised in the buffer, $V_{\text{i}}^{\max}$ is the full-scale input span of the ADC in V.

The resolution of the ADCs essentially defines what the settling time of analogue components involved in the analogue channel must be. For instance, an implementation of a 16-bit ADC would require the voltage signal to be settled to an accuracy of 0.0015% before the sampling occurs. Most data sheets specify the settling time to only 0.1% of the full-scale. Therefore, it is up to a designer to verify the figure experimentally at the component selection stage of a development project.

4.5.4 Grounding Aspects of the System Design

Figure 4.33 shows the optimum ground arrangement in the superconductor detector system. All internal PCB grounds of the detector, the preamplifier and the spectrometer are referred to the common "star" ground point organised in the preamplifier enclosure. The internal ground plane of the spectrometer and the power ground are coupled via a small value capacitor, *C*. This is to prevent high-frequency oscillations in the spectrometer. A single-ended analogue communication between the external equipment sitting on the power ground and the detector system should be accomplished via a forward referencing isolation amplifier, shown in Figure 4.34. The common mode rejection

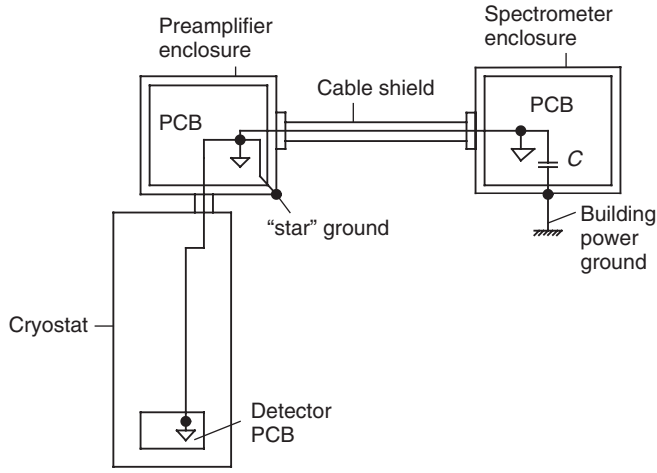


Figure 4.33 An optimum ground arrangement in the superconductor detector system. All internal PCB grounds of the detector package, the preamplifier unit and the spectrometer are referred to the common "star" ground point organised in the preamplifier enclosure. The internal ground plane of the spectrometer and the power ground are coupled via a small value capacitor, *C*. This is to prevent high-frequency oscillations in the spectrometer

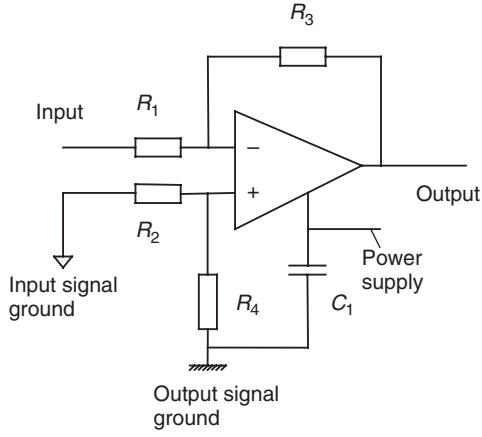


Figure 4.34 A forward referencing isolation amplifier. The common mode rejection ratio of the ground noise in this amplifier is defined by how closely matched the ratios of R_2 to R_1 and R_4 to R_3. R_1 must include a value of the output impedance of the signal source. The power supply for the isolation amplifier must be referenced to the output signal common

ratio of the ground noise in this amplifier is defined by how closely matched the ratios of R_2 to R_1 and R_4 to R_3. R_1 must include a value of the output impedance of the signal source. The power supply for the isolation amplifier must be referenced to the output signal common (see capacitance, C_1, in Figure 4.34).

Very often the detector system must have two power ground connections. This happens, for instance, in microanalysis, when the cryostat with the detector is attached to the grounded electron microscope. Moreover, there are many situations where the preamplifier and the spectrometer are separated by long cables under formidable man-made noise conditions. The metallic frame of a building is also not an equipotential structure. If the system is grounded at two different points of the frame, the AC and DC potential difference between them create the ground loop current in the cable shield, as shown in Figure 4.35. The current couples into central conductors through the mutual inductance and thus generates ground loop voltage noise at the input of the spectrometer buffer or pre-filter. A design of a cable affects the spectral characteristics of this noise. With high-quality cables it will be concentrated mainly in the low-frequency region. In this case, the differential amplifier becomes an efficient tool in combating the problem. The differential portion of the ground loop noise can be minimised by selecting a proper value of the resistor R_2, as shown in Figure 4.36. The resistor R_1 matches impedances of the preamplifier output and the cable for the minimum reflection of the signal pulses.

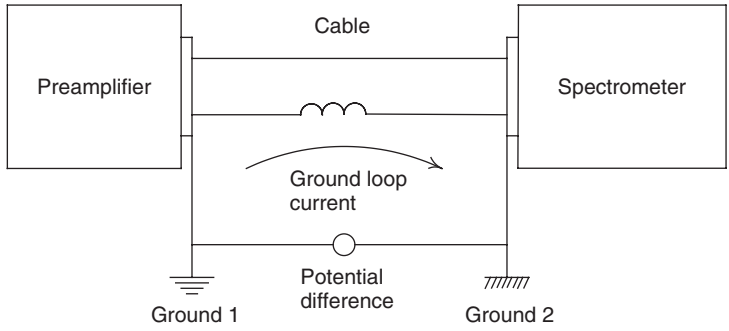


Figure 4.35 An example of a system grounded at two different points of the building ground. The AC and DC potential difference between them create the ground loop current in the cable shield. The current couples into central conductors through the mutual inductance and thus generates ground loop voltage noise at the input of the spectrometer buffer or pre-filter

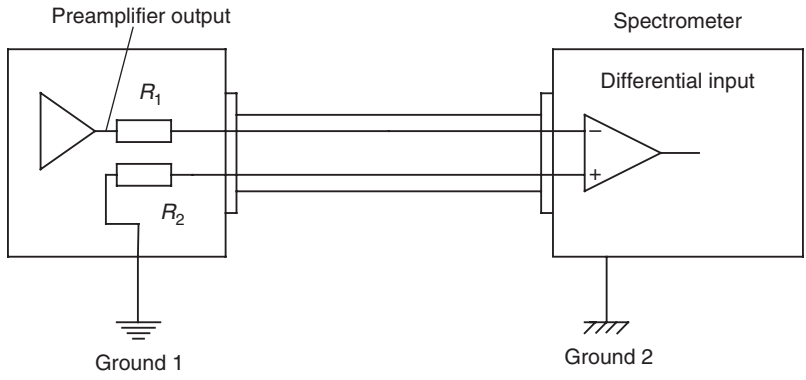


Figure 4.36 A ground arrangement in detector systems involving long connecting cables and the differential amplifiers. A differential portion of the ground loop noise can be minimised by selecting a proper value of the resistor R_2. The resistor R_1 matches impedances of the preamplifier output and the cable for the minimum reflection of the signal pulses

REFERENCES

[1] A. Oppenheim and R. Schafer. Digital Signal Processing, Prentice-Hall, 1975.
[2] Uwe Meyer-Baese. Digital Signal Processing with Field Programmable Gate Arrays. Springer-Verlag, Berlin, 2001.
[3] P. Lapsley, J. Bier *et al*. DSP Processor Fundamentals. J. Wiley & Sons, Inc., 1997.
[4] G. Knoll. Radiation Detection and Measurement. Wiley, 1989, p. 614.
[5] P. Johns and M. Yaffe. *Nucl. Instrum. Meth*. **A255**, 559, 1987.
[6] P. Nicholson. Nuclear Electronics. Wiley, 1972.
[7] J. Cover *et al*. *IEEE Trans. Nucl. Sci*. **29**, 609, 1982.
[8] V. Jordanov, G. Knoll, A. Huber and J. Pantaris. *Nucl. Instrum. Meth. Phys. Res*. **A353**, 261, 1994.
[9] A. Georgiev and W. Gast. *IEEE Trans. Nucl. Sci*. **40**, 770, 1993.
[10] V. Jordanov. *Nucl. Instr. and Meth. Phys. Res*. **A351**, 592, 1994.
[11] High Speed Converters, The Analog Devices Solution Bulletin, May 2003.
[12] W.K. Warburton and B. Hubbard. US Patent 5,870,051, 9 Feb. 1999.
[13] V. Tikhonov and M. Moronov. Markovian Processes. Radio I svjaz', 1980.
[14] G. Ripamonti, A. Castoldi and E. Gatti. *IEEE Trans. Nucl. Sci*., **41**, 1109, 1994.
[15] M. Sauerwald. Application Note AD-3, National Semiconductor, May 1994.
[16] B. Brannon. Application Note AN-501. Analog Devices, 1994.
[17] C. Cottini, E. Gatti and V. Svetlo. *Nucl. Instrum. Meth*. **24**, 241, 1963.
[18] E. Gatti and S. Cava. The quantization cell profile: a useful concept for a deeper insight in analogue to digital conversion. Proceedings of the International Conference on Measurement systems and Distributed Processing. University of Pavia, 15–16 Oct. 1979, p. 121.
[19] M. Al-Hadda, C. Lin, W. Miller and R. Berliner. *IEEE Trans. Nucl. Sci*. **41**, 1765, 1994.

[20] T. Ikeda, Y. Nakai, C. Otani *et al*. X-ray Spectroscopy of Highly Charged Ions using Superconducting Tunnel Junctions. Presentation at ISEC'2001, Osaka, 2001.
[21] J. Cardoso, J. Simoes, T. Menezes and C. Correia. *Nucl. Instrum. Meth. Phys. Res*., 2003 (in print).
[22] Reed, S. *Electron Microprobe Analysis*, Cambridge Press, 1975.
[23] H. Kraus, F. v. Felitsch, J. Johum, R. Mossbauer *et al*. *Phys. Lett*. **B321**, 195, 1989.
[24] B. Keele *et al*. *IEEE Trans. Nucl. Sci*. **43**, 1365, 1996.
[25] H. Haykawa. Computing. In: *Superconducting Devices* eds S. Ruggiero and D. Rudman, Academic Press, 1990.
[26] N. Fujimaki. IEEE J. Solid State Circuits Conf., WAM3.1, 1988.
[27] D. K. Brock. Intern. *J. of High Speed Elect. and Sys*. 11, 2307, 2001.
[28] P. Bunyk, K. Likharev and D. Zinoviev. ibid., 257.
[29] T. van Duzer and G. Lee. Digital Signal processing. In: *Superconducting Devices*, S. Ruggiero and D. Rudman Academic Press, 1990, p. 197.
[30] A. Guery, Ch. Kitchin. *Analog Dialogue*, 36-06, 2002.

5

Applications of Systems Based on Superconducting Detectors

In the previous four chapters, we reviewed principles of superconductor primary detectors, examined the signal propagation through the analogue channel of the spectrometer, the digital signal processing and, finally, discussed important aspects of the hardware design including the electromagnetic interference problem and grounding. Now it is time to show examples on how all this knowledge can be implemented in the end-to-end development of the complete measurement systems with advanced performance. We will mostly concentrate on two instruments representing an immediate commercial interest. These are: the superconducting energy dispersive spectrometer (SEDS) for the X-ray electron-probe nanoanalysis operating as a part of the high-resolution Field Emission Gun Scanning Electron Microscope (FEG SEM) and the Time-of-Flight mass-spectrometer (TOFMS) intended for the analysis of very large mass biopolymers.

The electron-probe analysis is a technique tailored for the investigation of the textural context and the chemical composition of various substances. The fundamental principle involves firing a beam of accelerated electrons onto a surface of a specimen of interest. The beam induces an emission of characteristic X-rays from within the specimen, which are subsequently measured by X-ray spectrometers.

An idea of the X-ray microanalysis was originally introduced by R. Castaing in his PhD thesis (University of Paris, 1951). He also developed the basic theory of the electron–X-ray interaction and demonstrated the principle in combining the electron microprobe with the Wavelength Dispersive Spectrometer (WDS).

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin
 ISBN: 0-470-85759-5

The era of the electron-probe microanalysis using the semiconductor energy dispersive spectrometer (EDS) started in 1968 with the publication of the dedicated paper by R. Fitzgerald, K. Keil and K. Heinrich [1]. Fast and efficient semiconductor EDSs had become an ideal supporting tool for rapid parallel developments in the imaging microscopy that took place in the late 1960s with the advent of high-resolution Scanning electron microscopes (SEM).

A further evolution in the electron-probe analysis will likely involve the EDS based on cryogenic detectors. NIST scientists were first to demonstrate new opportunities in the field [2]. In order to derive, say, a chemical composition in a nanostructure or super-thin multilayer interface, microscopes must operate with a low energy excitation and low current beam. The electron-probe analysis has to rely entirely on low intensity and closely positioned *L*, *M* peaks, where a high-energy resolution and the energy-dispersive mode of operation of superconductor detectors are at a premium.

Time-of-Flight mass-spectrometers utilise the calorimetric property of superconductor detectors (STJs and TESs). Their 100% efficiency is maintained up to virtually any size and mass of biopolymer molecules, even those that struggle to eject a measurable amount of electrons in conventional ionisation multiparallel plates. Cryogenic detectors offer new exciting opportunities in the high-precision mass-spectrometry. For instance, they are capable to actually measure the kinetic energy of molecules and thereby discriminate between a single molecule and specimen fragments, single- and multiple-charged molecules.

Certainly, there are many more potential applications for cryogenic detectors. The fundamental research in astronomy, the whole probe fluorescence analysis of light elements, crystallography, remote sensing and so on will benefit from their unsurpassed energy resolution. The overall evolution picture in these emerging technologies is very dynamic. At this stage, it is sensible to follow current publications and review papers presented at specialised conferences such as the Applied Superconducting Conferences (ASC), the International Superconductor Electronics Conferences (ISEC), Low Temperature Conferences (LT) and so on.

The information about superconductor detector systems will be incomplete without mentioning the accompanying components: the X-ray optics, the ultra low temperature cryogenics, low-temperature packaging, superconductor shielding and so on. We cannot describe them here in detail, but will give a short review of relevant papers where interested readers can find a more complete treatment.

5.1 ELECTRON-PROBE NANOANALYSIS WITH SUPERCONDUCTOR DETECTORS

5.1.1 Principle of the Electron-Probe Analysis

The electron-probe analysis is based on the measurement of the energy and the intensity of characteristic X-ray photons, which are generated during the interaction of the electron beam with a specimen. Figure 5.1 illustrates the interaction process. It presents a simplified shell model of a generic atom with *K*, *L* and *M* energy shells fully occupied by electrons. It is known that electrons tend to hold primarily the lowest possible energy states in the atom where they incidentally would have the lowest possible potential energy. For that reason, if a high-energy electron from the beam ejects one of the inner electrons, the appeared vacancy is quickly filled in by another electron from an outer shell. In Figure 5.1(a), the beam electron interacts with the *L*-shell electron transferring to the latter its kinetic energy, E_0. If the transferred energy exceeds the binding energy, E_L, the electron is forced to leave the shell. Figure 5.1(b) shows the atom in a state when the *L*-shell position is vacant straight after the ionisation process has been completed. If the vacancy is filled in with the *M*-shell electron, its potential energy decreases from E_M to E_L.

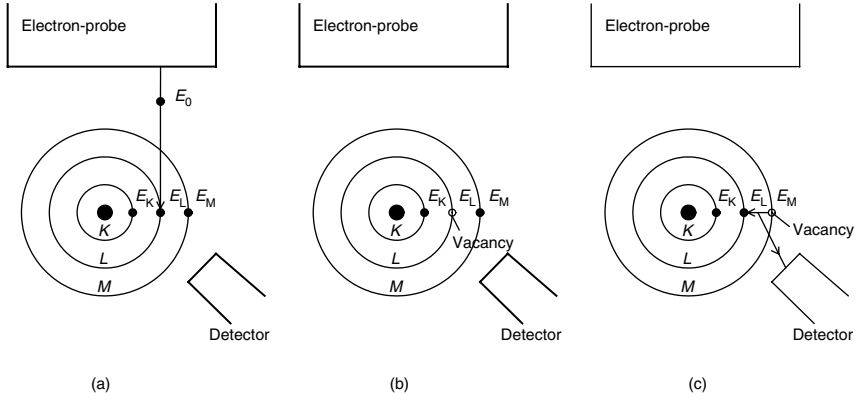


Figure 5.1 A simplified shell model of a generic atom with *K*, *L* and *M* energy shells fully occupied by electrons. The beam electron interacts with the *L*-shell electron transferring to the latter its kinetic energy, E_0 (a). If the transferred energy exceeds the binding energy, E_L, the electron is forced to leave the shell. (b) shows the atom in a state when the *L*-shell position is vacant straight after the ionisation process has been completed. If the vacancy is filled in with the *M*-shell electron, its potential energy decreases from E_M to E_L. The difference is released as a characteristic photon with the energy $\sim (E_M - E_L)$ (c)

The difference is released as a characteristic photon with the energy $\sim (E_M - E_L)$ (Figure 5.1c).

A qualitative representation of peaks in the energy histogram generated by transitions between *L–K* shells, *M–L* shells and *N–M* shells, is captured in Figure 5.2. In practice, the atomic shell structure is more complex and, therefore, peaks can contain a number of satellite lines (see details, for instance, in [3]). Yet, positions of each individual peak as well as of all satellite lines on the energy scale are unique for each atom. Thus, the measurement of characteristic X-rays helps to restore the chemical composition of the specimen.

The identification of chemical elements from peak positions is called *qualitative* analysis. Intensities of the characteristic peaks constitute the input data for the *quantitative* analysis. After instrumental corrections, the subtraction of the background and matrix corrections quantitative analysis yields the concentration ratio of chemical elements in a specimen.

Instrumental factors that may have a substantial effect on the line intensities include:

– X-ray collimation;
– windows and filters;
– trapping of stray electrons;

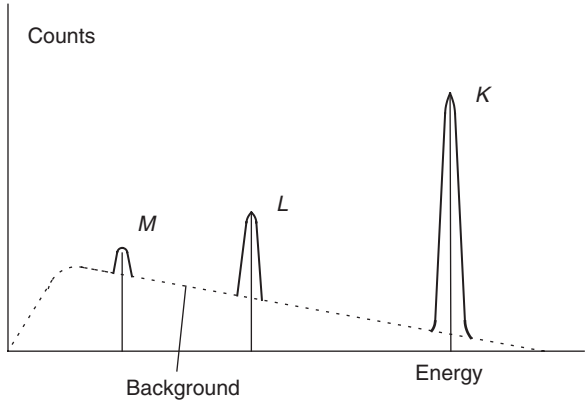


Figure 5.2 A qualitative representation of peaks in the energy histogram generated by transitions between *L–K* shells, *M–L* shells and *N–M* shells. In practice, the atomic shell structure is more complex and, therefore, peaks can contain a number of satellite lines [3]. Positions of each individual peak as well as of all satellite lines on the energy scale are unique for each atom. Thus, the measurement of characteristic X-rays helps to restore the chemical composition of the specimen. Characteristic peaks sit on the top of the background noise called bremstrahlung

– extraneous noise in the system;
– pile-up effect;
– Abrikosov vortices in the case of STJ detectors; and
– performance degradation of a detector with time due to thermo-cycling, building up stresses in interfaces, changing diffusion times, icing and so on.

5.1.1.1 Background noise in the energy histrogram

Figure 5.2 shows that characteristic peaks sit on the top of the background noise signal. It originates from elastic collisions of probe electrons with charged nuclei. During the process of the deceleration, the deflection electrons bound to lose their kinetic energy generating X-ray photons. The phenomenon is called bremstrahlung. Since free electrons do not have quantum energy states, the energy spectrum of the bremstrahlung photons is continuous extending from 0 to the excitation energy of the beam, E_0. The distribution of photons as a function of energy is given by [4]

$$N(E) = 2 \times 10^{-9} Z \frac{E_0 - E}{E} \tag{5.1.1}$$

At a lower energy range the background decreases due to the absorption of low-energy photons in the excitation volume of the sample. Figure 5.2 indicates that these photons cannot reach the detector rather than the fact that they are not being generated.

The energy loss to the bremstrahlung is inherent to the electron-probe analysis. Quantitatively it takes only a fraction of a per cent of the incident beam energy. Yet, the created continuum imposes a fundamental limitation to the low-energy analysis and often sets up the minimum detection limit.

5.1.1.2 Matrix corrections

The matrix correction procedure represents essentially a solution of the inverse problem involving the relationship between corrected line intensities and the chemical composition. It must take into account all factors that influence intensities of characteristic peaks. According to [5] these are:

– the absorption of characteristic X-rays;
– the secondary fluorescence generated by absorbed characteristic X-rays and the background;

– the backscattering of incident probe electrons;
– the variations in the efficiency of the X-ray production caused by the "stopping" power of the specimen and so on.

An illustration of all listed factors is given in Figure 5.3. For instance, the absorption of characteristic X-rays occurs when their energy is used to eject other electrons in the same (Auger effect) or other atoms. This reduces a, so-called, fluorescence yield of the sample. At the same time, transitions of electrons to occupy vacancies created by the absorption effects enhance lower energy peaks in the spectrum. These peaks also need to be corrected if they participate in the quantitative analysis. Further details on the matrix correction technique can be found in [5,6].

Figure 5.3 shows that there are two excitation volumes in a specimen. The primary volume of the direct interaction with the e-beam is surrounded by a larger volume of secondary excitations, which also affect

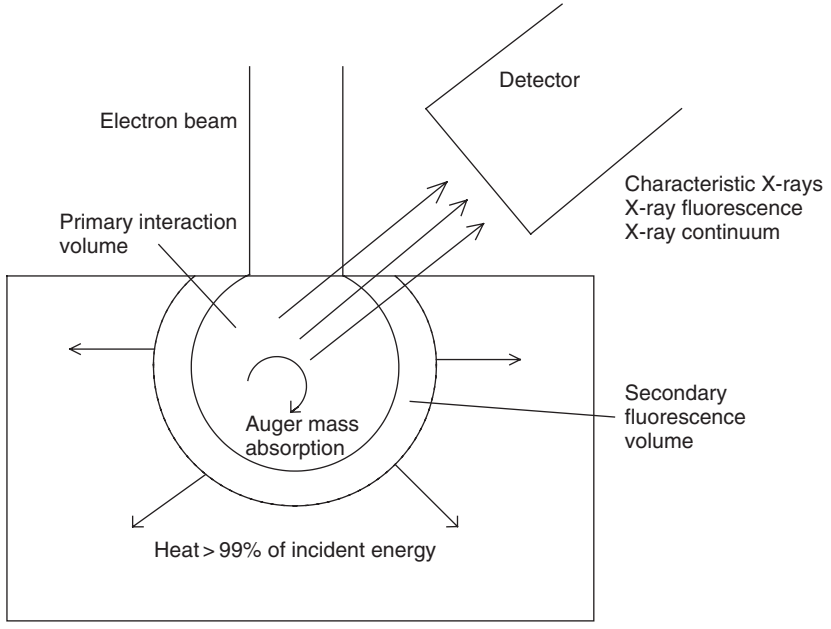


Figure 5.3 Schematic representation of interaction of the electron beam with a specimen. There are two excitation volumes in a specimen. The primary volume of the direct interaction is surrounded by a larger volume of secondary excitations, which also affect the final spectrum of the X-ray spectrometer. The electron-probe nanoanalysis requires both the volumes to be minimised so that they could be contained well within a nanostructure under study or a thickness of layers in multilayer thin films

the final spectrum of the X-ray spectrometer. The electron-probe nanoanalysis requires both the volumes to be minimised so that they could be contained well within a nanostructure under study or a thickness of layers in multilayer thin films. The characteristic dimensions of the volumes such as the spot diameter and the penetration depth of the incident electrons depend on the e-beam diameter, d, and the accelerating voltage, E_0, respectively. According to [5], the minimisation of d can be achieved by reducing the beam current, I_p, as

$$d = \left[\frac{i_p}{\frac{3}{16}\pi^2 \beta C_s^{-\frac{2}{3}}} \right]^{\frac{3}{8}} \tag{5.1.2}$$

where β is the brightness of the image, C_s is the spherical aberration coefficient of the lens. With typical values of $\beta = 2.8 \times 10^4\,\mathrm{A\,cm^{-2}\,sr^{-1}}$ and $C_s = 50\,\mathrm{mm}$, the equation (5.1.2) suggests that the operation with a maximum excitation current of a few pA will be needed to maintain a spot size of approximately 10 nm. The number of electron arriving at the specimen surface per second can be estimated as $n_0 \approx 6.25 \times 10^{18} i_p s^1$. Most of the beam energy is dissipated into the heat. In practice, only a fraction of a per cent of the beam electrons generates informative characteristic X-rays. From a rough estimation, the generation rate of X-ray photons available for the detection into the whole semisphere (a solid angle of 2π) above the specimen is expected to be just a few thousand counts per second. Such a low photon generation rate raises concerns regarding the statistical error, which, incidentally, has to be kept either well below the energy resolution of the spectrometer or low enough to avoid severe overlapping between spectral peaks. The solid angle and the quantum efficiency of the detector must be maximised in order to be able to complete the data acquisition within a reasonable period of time.

As we mentioned before, the penetration depth of the electron beam into a sample is primarily a function of the accelerating voltage and the mass density of the material. According to [7], the analytical resolution follows a relationship of the form:

$$R[nm] = 2.76 \frac{A}{\rho Z^{0.89}} \left(E_0^{1.67} - E_c^{1.67} \right) \tag{5.1.3}$$

where E_c is the critical excitation energy for the characteristic X-rays of interest (keV), A is the atomic weight (g/mol), Z is the atomic number

and ρ is the density of the specimen (g/cm^{-3}). The intensity of characteristic X-rays is given by [5]

$$I = 3.48 \times 10^4 \left(\frac{\omega L}{Ac}\right)(U-1)^{1.67} \tag{5.1.4}$$

where ω is the fluorescence yield of the structure of interest, L is a factor taking into account the loss of the ionisation resulting from the electron backscattering, $c = 3 \times 10^5\,\mathrm{keV^2\,cm^{-2}\,g^{-1}}$, $U = \frac{E_0}{E_c}$ is the overvoltage factor. The recommended value of the minimum overvoltage is approximately 2 for the conventional quantitative electron-probe analysis. The qualitative analysis can work with values as low as 1.5 [8].

The equations (5.1.3) and (5.1.4) show that the reduction in the accelerating voltage of the current beam enables one to reach a sub-μm spatial resolution at E_0 lower than 10 keV [9]. However, the nanoanalysis must rely mostly on lower intensity peaks. For instance, an accelerating voltage $E_0 = 5\,\mathrm{keV}$ and an overvoltage $U = 2$ will be sufficient for the excitation of K lines up to the element with the atomic number $Z = 16$ (sulphur, $K_\alpha = 2.307\,\mathrm{keV}$), L lines up to $Z = 43$ (technetium, $L_{\alpha 1} = 2.4240\,\mathrm{keV}$) and M lines up to $Z = 83$ (bismuth, $M_{\alpha 1,2} = 2.4197\,\mathrm{keV}$). All these peaks, including K lines of light elements, have low fluorescence yields. In almost all possible compositions, they are likely to be closely positioned on the energy scale. In this case, the resolving power depends entirely on the energy resolution of the spectrometer, particularly if a composition of a specimen is not known in advance to apply the mathematical deconvolution technique.

A good example of the resolving power of X-ray spectrometers is shown in Figure 5.4. It demonstrates two spectra of the same $BaTiO_3$ sample measured with a state of art Si EDS (a) and a NIST microcalorimeter EDS(b). The sample was excited in the low voltage region with $E_0 = 3\,\mathrm{keV}$. The conventional liquid-nitrogen-cooled Si EDS had the energy resolution of 129 eV at MnK_α (5890 eV) and 61 eV at CK_α (282 eV), whereas, for the NIST microcalorimeter, the figures were better than 10 eV and 2 eV respectively. The mass fractions in the sample were divided between Ba (0.589), Ti (0.205) and O (0.206). Carbon is present as a contaminant.

Problems associated with low peak intensities and the peak overlapping are obvious. In Figure 5.4(a), Ba M-peaks are barely visible above the background despite the large mass fraction of Ba in the sample.

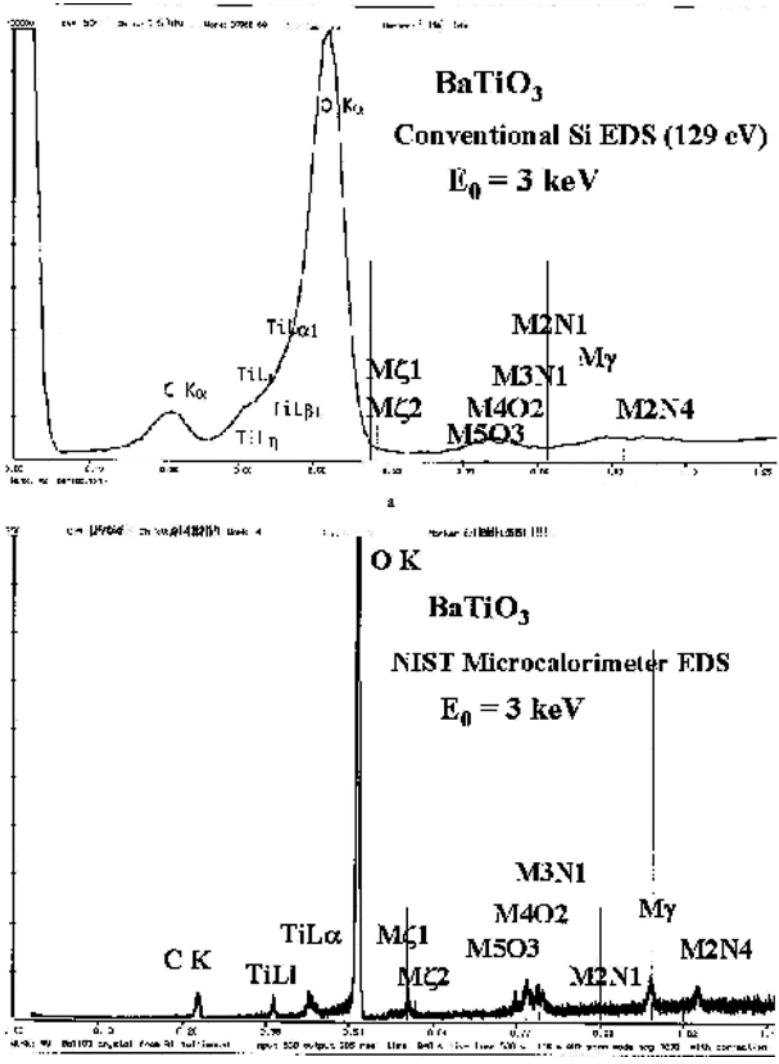


Figure 5.4 It demonstrates two spectra of the same $BaTiO_3$ sample measured with a state of art Si EDS (a) and a NIST microcalorimeter EDS (b). (Reproduced with permission from D. Newbury, 2002.) NIST The sample was excited in the low voltage region with $E_0 = 3$ keV. The conventional liquid-nitrogen-cooled Si-EDS had the energy resolution of 129 eV at MnK_α (5890 eV) and 61 eV at CK_α (282 eV), whereas, for the NIST microcalorimeter, the figures were better than 10 eV and 2 eV respectively. The mass fractions in the sample were divided between Ba (0.589), Ti (0.205) and O (0.206). Carbon is present as a contaminant

The Ti L-peaks remained unresolved forming a shoulder on the O K peak. Following [10], we can estimate the concentration limit of detection as

$$C_{DL} = \frac{3.29a}{\sqrt{n\tau \frac{P}{C_i}\eta}} \tag{5.1.5}$$

where a is the constant in Ziebold–Ogilvie hyperbolic equation [11], n is the number of measurements, τ is the measurement time, P is the peak counting rate, C_i is the concentration (mass fraction) of an element of interest in the compound and η is the peak to background ratio on the standard.

Newbury calculated the detection limit for both spectra shown in Figure 5.4 as a function of the measurement time. The calculations were based on the most prominent Ba M-peak resulted from the M_3N_1 transitions. For a reasonable measurement time $\tau = 1000$ s, the detection limit of the Ba-like element in $BaTiO_3$ with Si EDS ($E_0 = 3$ keV and 20% dead time) is $C_{DL} = 0.098$ [9].

The spectrum of the sample presented in Figure 5.4(b) is much clearer. The microcalorimeter EDS provides complete resolution of Ti L-peaks from oxygen, and even separation between Ba M-family-peaks. The Newbury evaluation showed a fivefold improvement in the concentration detection limit. Its value is approximately $C_{DL} = 0.0179$ for the same measurement time $\tau = 1000$ s. The calculations assumed a maximum count rate of 800 c/s which can be achieved with the transition edge sensor microcalorimeter. The detection limit can further be reduced by a factor of 3 by employing an STJ detector spectrometer capable to count with a rate of 10 kcps.

5.1.2 Solid Angle

The count rate capability is an important figure of merit of the X-ray spectrometer. It often defines a scope of applications, which can be handled with the instrument. Yet, as we mentioned before, a specific feature of the electron-probe nanoanalysis is a relatively low photon generation rate (PGR). Thus, the measurement time is governed by the active area of the detector intercepting these photons and the distance between the detector and the excitation volume in the specimen. Figure 5.5 illustrates the statement. Provided that the X-ray intensity is uniform in all directions of the semisphere above the specimen plane, the input count rate is given by

$$ICR = PGR\frac{\Omega}{2\pi} \tag{5.1.6}$$

where Ω is the solid angle of the detector.

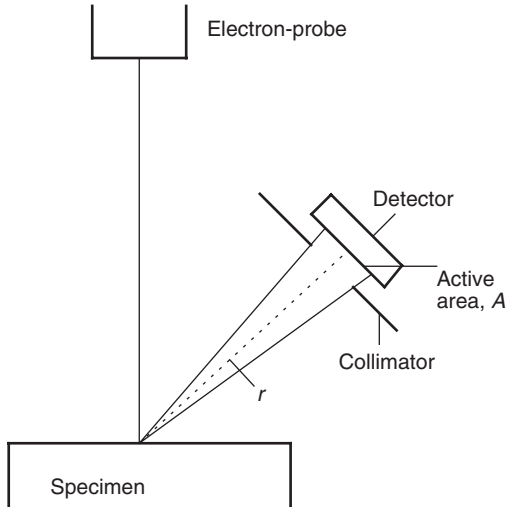


Figure 5.5 Schematic representation of the electron-probe specimen-detector arrangement in the nanoanalysis

The solid angle is a three-dimensional analogue of the ordinary angle. A solid angle, Ω, subtended by a surface S is defined as the surface area, Ω, of a unit-sphere covered by the projection of the surface S onto the unit-sphere [12], that is,

$$\Omega = \iint_s \frac{\hat{\boldsymbol{n}}}{r^2} \mathrm{d}a \tag{5.1.7}$$

where $\hat{\boldsymbol{n}}$ is the unit-vector from the origin (middle of the excitation volume), da is the differential area of a surface patch and r is the distance from the origin to the patch.

The solid angle is measured in steradians, sr. A solid angle of 1 sr represents a cone that covers an area of 1 m^2 on the surface of a 1-m radius sphere. A full sphere has a solid angle of 4π sr. If the solid angle is below 0.2 sr, the equation (5.1.7) can be written down in the simplified form as

$$\Omega = \frac{A}{r^2} \tag{5.1.8}$$

where A is the active area of the detector after the collimation and r is the distance between the excitation volume and the central point of the absorber surface.

5.1.3 X-ray Optics

A typical area of a superconducting absorber attached to the STJ is below 1 mm^2, as opposite to tens of mm^2 of active area in semiconductor detectors. The active area is normally further reduced by the collimation to ensure that X-rays are incident only on the absorber and do not hit sensors themselves. This, according to equation (5.1.8) yields a solid angle as low as

$$\Omega = \frac{A}{r^2} \leq \frac{\pi(0.5\,\text{mm})^2}{(40\,\text{mm})^2} = 4.9 \times 10^{-4}\,\text{sr}$$

Using the X-ray optics increases the effective solid angle of the spectrometer. An X-ray lens collects characteristic X-rays over a large solid angle and focuses them onto a small area absorber of the superconductor detector, as shown in Figure 5.6.

The major figures of merit characterising performance of the focusing optics include an achievable spot size and an X-ray flux-density gain. It is desirable to maximise both of them. The research and development is currently concentrated on two types of lenses based on:

1. tapered monocapillaries [13,14]; and
2. polycapillary design.

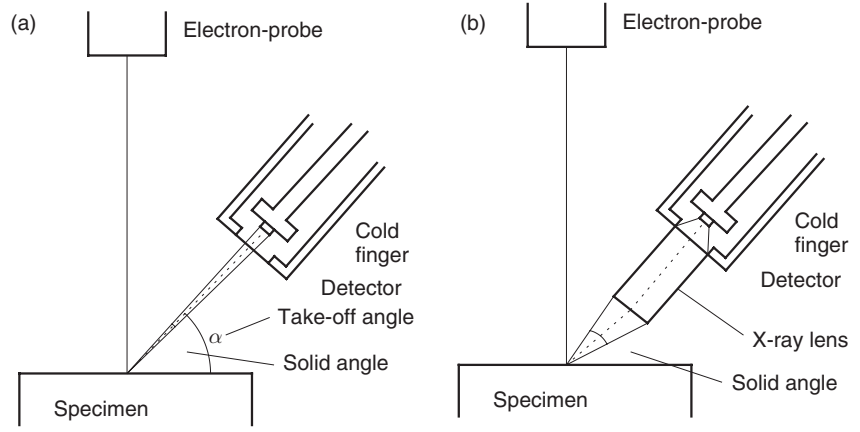


Figure 5.6 The solid angles of detectors without and with X-ray optics (a) and (b) respectively. Using the X-ray optics increases the effective solid angle of the spectrometer. An X-ray lens collects characteristic X-rays over a larger solid angle and focuses them onto a small area absorber of the superconductor detector

Monocapillaries must have a very smooth internal glass surface, which prevents the divergence of the X-rays by total repeated reflections. Monocapillary lenses are able to provide a modest flux-density gain and, therefore, X-ray spectrometers tend to utilise more advanced polycapillary counterparts.

A polycapillary lens can contain up to several thousands of individual glass fibres. The fibres apart from those right in the centre of the lens are bent so that they all are directed towards one focusing point coinciding with the surface of the detector absorber. The lens must be carefully selected to fit an intended measurement. The selection criterion is primarily concerned with the energy range of X-ray photons that have to be detected. The matter is that the outer surface microfibres have the maximum bending angle. They are least efficient in transporting X-rays. The efficiency decreases with the increase in the photon energy owing to the energy dependence of the critical angle for the total reflection. Thus, lenses are characterised by the cut-off energy, E_c.

The diameter of the spot on the absorber achieved with a polycapillary lens is given by [15]

$$D_{(\mathrm{FWHM})} \approx \sqrt{d_{\mathrm{ch}}^2 + (F \cdot \Theta_c)^2} \tag{5.1.9}$$

where d_{ch} is the output diameter of the overall capillary channel, F is the focal distance and Θ is the critical angle for the total reflection. The critical angle, in radians, can be derived approximated from the following equation

$$\Theta_c(E) \approx 2.04 \times 10^{-2} \frac{\sqrt{\rho}}{E} \tag{5.1.10}$$

with E being the energy of the characteristic X-ray photons and ρ is the density of the capillary material.

The commercially available polycapillary optic enables the reduction of the spot size to a value of 0.3 mm at energies below 1 keV (see, for instance, [16]). An optical gain of an X-ray lens as a function of the energy was evaluated experimentally in [2]. A gain factor was reported to be approximately 300 that ensured a solid angle of 4 mst. The solid angle was referred to a typical specimen-detector distance of 35 mm without the optic in place.

The effective solid angle and the take-off angle have a strong effect on the quantitative analysis corrections for the self-absorption and other matrix effects. The take-off angle is defined as an angle between the surface of the specimen which is emitting the X-rays and a line connecting

the excitation volume with the middle of the detector absorber or the lens, if present. When an X-ray energy analyser is intended for use with a scanning electron microscope, the solid angle, the take-off angle and the distance between the excitation volume to the detector should be included in the main specification list.

5.1.4 Cryogenics

As we mentioned before, superconductor detectors operate at ultralow temperatures (ULT) below 0.3 K. For those, who are not familiar with the ULT cryogenics, it seems to be overcomplicated and is a serious obstacle for the commercialisation of new advanced spectrometers. In reality, refrigeration techniques are simple and commercially available systems are quite compact and reliable [17]. Since superconductor detectors (this is equally true for high-performance semiconductor detectors) must be surrounded by vacuum to prevent icing, the contamination of the active area and the X-ray absorption, the cryogenics, in fact, becomes a natural integral part of the system.

Any cooling method is always based on the process of entropy reduction controlled by means of an external influence. Two of these methods are most widely utilised in the practical cryogenics. The first one is based on controlling the internal energy of gases, U, by means of the pressure, p, and the volume, V, in the form

$$\mathrm{d}U = T \cdot \mathrm{d}S - p \cdot \mathrm{d}V \tag{5.1.11}$$

The second one relates to the entropy in paramagnetic materials governed by the applied magnetic field, B, and the magnetisation, M, so that

$$\mathrm{d}U = T \cdot \mathrm{d}S + B \cdot \mathrm{d}M \tag{5.1.12}$$

The equation (5.1.11) represents the first law of the thermodynamics derived for an ideal gas, whereas the equation (5.1.12) is its analogue describing the thermodynamics of an ideal system of magnetic domains. In both the equations, T is the temperature of the lattice and S is the entropy.

5.1.4.1 The ^{3}He refrigerator

Figure 5.7 illustrates a concept of the cryogenic arrangement based on the evaporation of ^{3}He gas. The whole system is pre-cooled to approximately

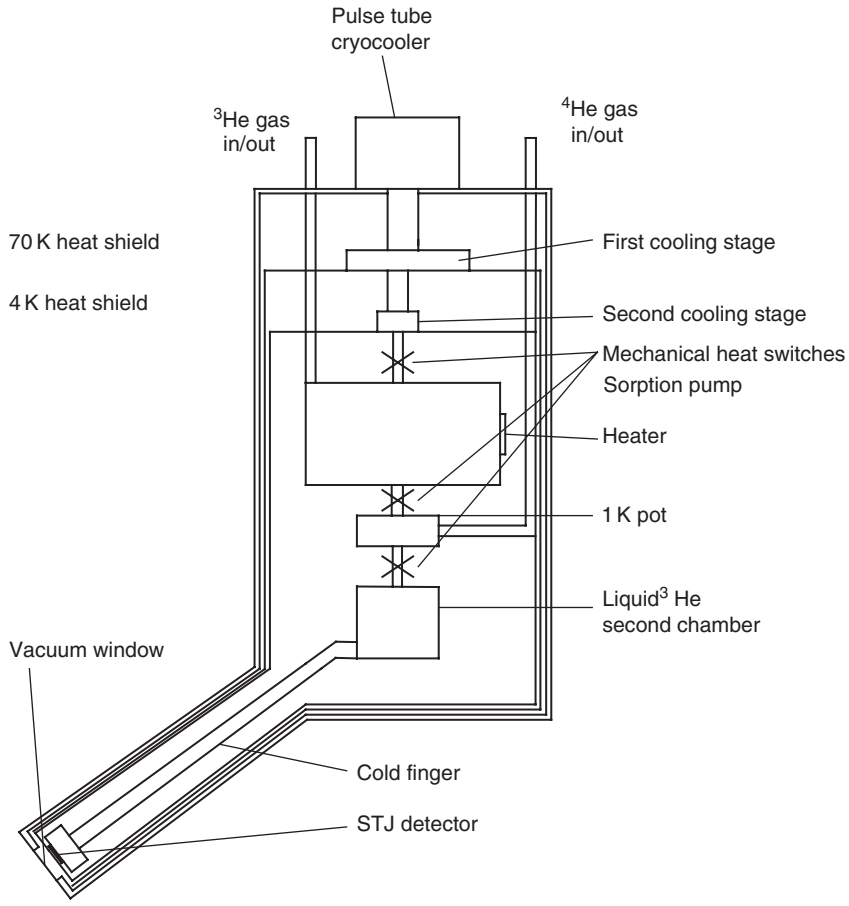


Figure 5.7 A concept of the cryogenic arrangement based on the evaporation of ^{3}He gas. The whole system is pre-cooled to approximately 70 and 4 K with a two-stage pulse-tube mechanical cooler. The first 70 K stage provides cooling for copper heat shield that surrounds the inner volume of the cryostat. The shield together with the aluminised super-insulation film intercepts the heat load from the room temperature to the 4 K shield. The ^{3}He refrigerator consists of three major parts. These are: the sorption pump, a first ^{4}He liquefying chamber, called 1 K pot and a second ^{3}He liquefying chamber to which the cold finger with detector is attached. ^{4}He gas is condensed in the 1 K pot. An outlet of the 1 K pot is connected to a vacuum pump to reduce the evaporation pressure and, thereby, to pre-cool the fist chamber to approximately 1 K. This temperature is sufficient to condense the ^{3}He gas from the room temperature storage. The liquid drips down into the second chamber at the base of the refrigerator. Once the gas was condensed, the sorption pump is used to reduce the vapour pressure of the ^{3}He cooling liquid and attached to it cold finger to a temperature approximately 0.24 K within a few minutes

70 and 4 K with a two-stage pulse-tube mechanical cooler. This particular type of the cooler is normally chosen due to the omission of displacers moving up and down with each cycle inside vacuum seals. The feature reduces substantially the vibration level compared to other competitive technologies (which is very important for the electron-probe nanoanalysis) and lengthens the mean time between maintenance up to 3–10 years of exploitation [18].

The first 70 K stage provides cooling for copper heat shield that surrounds the inner volume of the cryostat. The shield together with the aluminised super-insulation film intercepts the heat load from the room temperature to the 4 K shield. The cooling power of the 70 K stage is as large as 70 W. Therefore, it can also accommodate input-FETs of preamplifiers to reduce their noise. The 4 K shield is attached to the second cooling stage of the cryocooler. Its cooling power varies between 0.5 and 1 W being a subject of a size, a design and the input power.

The ^{3}He refrigerator consists of three major parts. These are: the sorption pump, a first ^{4}He liquefying chamber, called 1-K pot, and a second ^{3}He liquefying chamber to which the cold finger with detector is attached. ^{4}He gas is condensed in the 1-K pot. An outlet of the 1-K pot is connected to a vacuum pump to reduce the evaporation pressure and, thereby, to pre-cool the fist chamber to approximately 1 K. This temperature is sufficient to condense the ^{3}He gas from the room temperature storage. The liquid drips down into the second chamber at the base of the refrigerator. Once the gas was condensed, the sorption pump is used to reduce the vapour pressure of the ^{3}He cooling liquid and attached to it cold finger to a temperature approximately 0.24 K within a few minutes.

The temperature of the sorb has strong effect on its gas absorption rate. This, in turn, defines the vapour pressure of the ^{3}He vapour, and, thus, the temperature of the liquid ^{3}He. A temperature stability as low as 1 mK can be achieved on the cold finger with a temperature controller that is set up to measure the sample temperature and control the power supplied to the sorb heater.

A detailed description of the principle and design issues of the ^{3}He refrigerator can be found in [19] and [20].

5.1.4.2 The Adiabatic Demagnetisation Refrigerator (ADR)

Transition Edge Sensor (TES) microcalorimeters demonstrate their best performance at a base temperature below 100 mK [2]. ^{3}He refrigerators are incapable to cover this temperature range. Therefore, microcalorimeters are normally mounted on cold fingers of ADRs. The principle of the ADR is

based on the second technique, described by the equation (5.1.12). It employs a volume filled with a paramagnetic material whose entropy is controlled by an external magnetic field, B. The external magnetic field interacts with magnetic moments of domains in such a way that they tend to align along the vector $\bar{B}$, thereby lowering the magnetic entropy of the system. The entropy reaches its minimum when $\bar{B}$ is strong enough to perfectly align all domains in its direction. The process of the alignment is followed by the generation of heat in each domain, which is transferred to the lattice rising the physical temperature of the volume. The heat can be absorbed from the lattice through a specially designed thermal conduction link. When the material regains its starting temperature, the thermolink and the magnetic field are removed. The process of the randomisation of magnetic moments in adiabatic conditions results in an increase of the magnetic entropy and a subsequent decrease in the lattice temperature.

Figure 5.8 illustrates the process of the adiabatic refrigeration. It shows the entropy of the paramagnetic material as a function of temperature. The upper curve represents the entropy at a zero magnetic field, whereas the lower one corresponds to $B = B_0 > 0$. The process starts at point 1 ($T = T_1$, $B = 0$). With the heat link activated and the magnetic field set up to a value of B_0, the material becomes isothermally magnetised reducing the entropy to point 2 ($T = T_1$, $B = B_0$). The next step is removing the heat link and the external magnetic field. During the process of the adiabatic demagnetisation the operating point returns onto the upper curve, but at a lower temperature, T_2 (point 3 with $T = T_2$, $B = 0$). The heat load will eventually warm up the volume. The entropy will follow the upper curve until it has reached the starting point 1.

An example of the ADR design is shown schematically in Figure 5.9. It consists of three major components. The first one is the paramagnetic material inside a container, usually called the salt pill. The second component is the superconducting magnet. The salt pill is suspended inside the bore of the magnet with a carefully designed support structure. This is made usually of tensioned strings having a very low thermal conductivity to provide the adiabatic condition. The third major component is the mechanical heat switch. The mechanical heat switch helps to maintain temperature of the guard salt pill close to the temperature of the second stage of the pulse tube (approximately 4 K) during the isothermal magnetisation. The TES microcalorimeter is mounted on the end surface of the cold finger which, in turn, is attached to the salt pill.

The ADR is not capable to reduce a temperature from 4 K to below 100 mK in a single step. Therefore, it normally employs two different paramagnetic materials, such as gadolinium gallium garnet (GGG) as a first

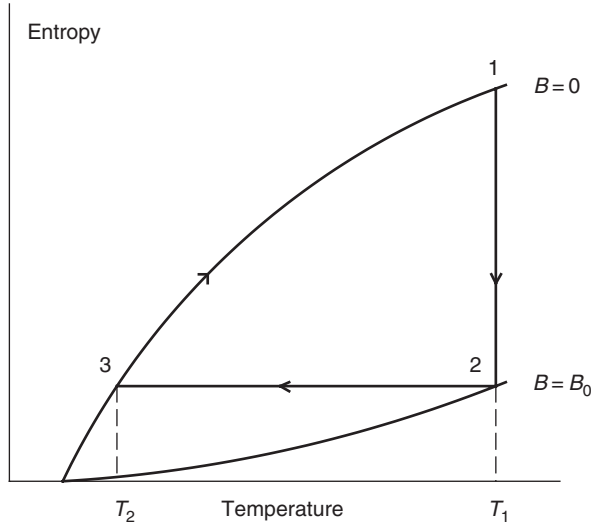


Figure 5.8 The entropy of the paramagnetic material as a function of temperature. The upper curve represents the entropy at a zero magnetic field, whereas the lower one corresponds to $B = B_0 > 0$. The process starts at point 1 ($T = T_1$, $B = 0$). With the heat link activated and the magnetic field set up to a value of B_0, the material becomes isothermally magnetised reducing the entropy to point 2 ($T = T_1$, $B = B_0$). The next step is removing the heat link and the external magnetic field. During the process of the adiabatic demagnetisation, the operating point returns onto the upper curve, but at a lower temperature, T_2 (point 3 with $T = T_2$, $B = 0$). The heat load will eventually warm up the volume. The entropy will follow the upper curve until it has reached the starting point 1

stage guard pill (the upper pill in Figure 5.9) and ferric ammonium alum (FAA), as a second stage (the lower pill in Figure 5.9). An example of a two-stage ADR was described in [21]. The NIST group, incidentally, used a refrigerator of this design to house the microcalorimeter EDS reported in [2].

Two-stage ADRs require a rather high magnetic field, B_0, of between 4 and 9 T to bring to minimum the entropy of both the paramagnetic materials. The magnetic field of this strength generates a lot of thermal power in pills, which a compact mechanical cooler will struggle to absorb over a reasonably short period of time. In order to avoid this drawback, there is an upgraded design of the refrigerator. It is built as a truly two-stage ADR, with separate magnets and heat switches for each pill [22]. Such a design is called the double ADR. The double ADR needs only half of the magnetic field strength to achieve performance similar to those of the two-stage ADR with the guard pill. Their peak thermal load is compatible with existing commercial pulse tube coolers. Details of the design and technical specifications of the two-stage ADR can be found, for instance, in [22].

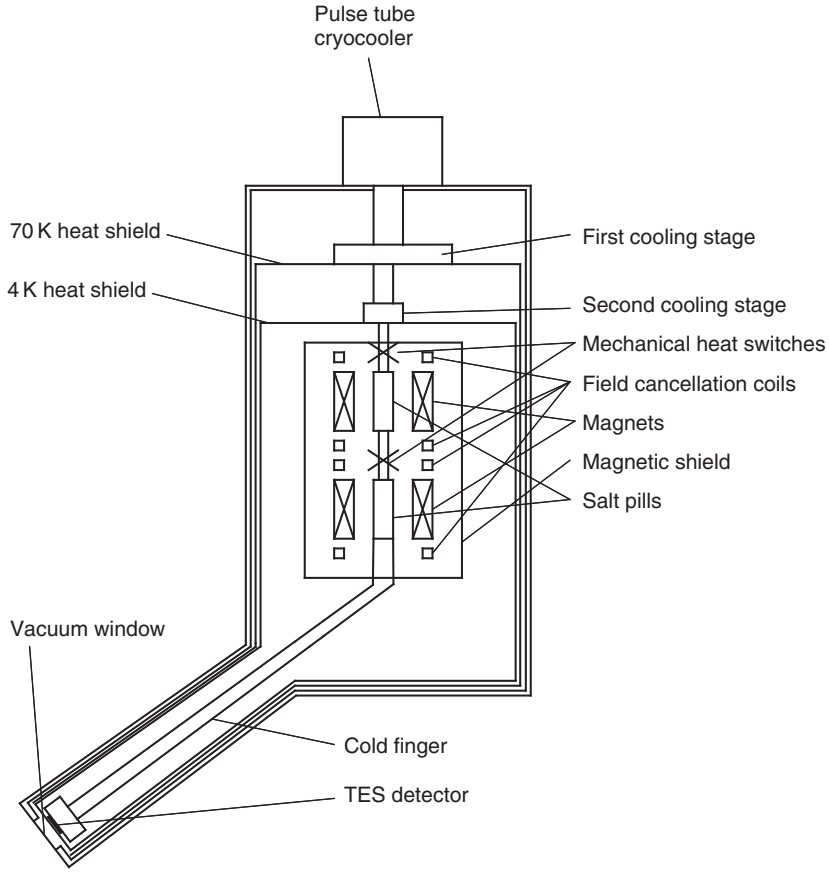


Figure 5.9 An example of the ADR design. It consists of three major components. The first one is the paramagnetic material inside a container called the salt pill. The second component is the superconducting magnet. The third major component is the mechanical heat switch. The mechanical heat switch helps to maintain temperature of the guard salt pill close to the temperature of the second stage of the pulse tube (approximately 4 K) during the isothermal magnetisation. The TES microcalorimeter is mounted on the end surface of the cold finger which, in turn, is attached to the salt pill

The stability of the base temperature is a crucial factor in maintaining the constant gain of the spectrometer. Practical ADRs operate at a temperature, T_b, slightly higher than T_2 (Figure 5.8). For that purpose, the demagnetisation process is put to hold, once T_b has been reached by the initial reduction of the magnetic field to an intermediary value B_1 rather than zero. Figure 5.10 illustrates the case. The power supply of the magnet is controlled with an intelligent temperature controller that

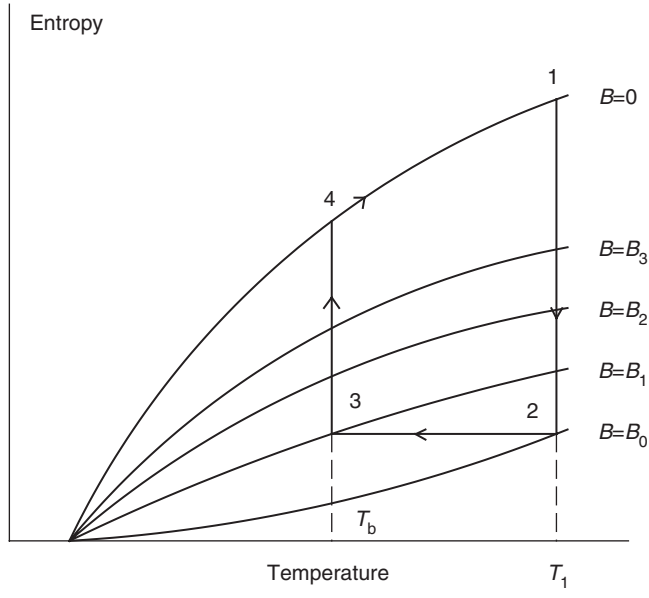


Figure 5.10 The entropy of the paramagnetic material as a function of temperature in a practical ADR. The magnetic field of the salt pill solenoid is gradually reduced to maintain a stable base temperature T_b of the cold finger

monitors the temperature of the cold finger. A constant temperature is maintained via a slow reduction of the field so that the cooling effect of the demagnetisation process compensates the heat load from the detector and wiring, heat leaks into the cold finger and the pills, and, finally, the heat radiation from the 4 K shield. The algorithm of the servo-control originates from the following expression [22]

$$\frac{dT}{dt} = \frac{dQ_{leak}}{dt}\frac{1}{C} + \frac{\partial T}{\partial B}\frac{dB}{dt} \tag{5.1.13}$$

In this equation, Q_{leak} is the heat load, C is the specific capacitance of the cold stage and t is the time.

5.1.4.3 Thermal conduction through solids

Perhaps the largest thermal load propagates through the wiring connecting the room temperature electronics with devices mounted on the cold stage. These include a detector or an array of detectors, a SQUID preamplifier, thermometers and so on. In the normal metal wires, the

thermal conductivity, κ, is proportional to the electric conductivity, σ. This follows from the Wiedemann–Franz law

$$\kappa = \frac{\pi^2}{3}\left(\frac{k_b}{e}\right)^2 T_\sigma \tag{5.1.14}$$

where T is the average temperature of the wire and e is the charge of the electron. A current flowing through the wire also generates the Joule heating due to the resistance. In order to eliminate it, the 4 K stage and devices on the cold finger are connected by superconductor wires. Below the transition temperature superconductors have a small amount of conducting normal electrons. Therefore, superconductor wires have very low thermal conductivity and a virtually zero Joule heating.

A specific feature of superconductor wiring in the vacuum is the fact that leads must be properly heat sinked on cold metallic surfaces to maintain their superconductivity. In places where it cannot be done, copper-clad wires are used. The copper surface layer keeps the wire cold, whereas the current flows through the superconductor core without heat dissipation.

Non-metallic joints and support structures also conduct heat into the ULT stage. The heat flow through a bar of cross-sectional area, A, under a temperature gradient, $\frac{\partial T}{\partial x}$, is given by

$$Q = \lambda(T) A \frac{\partial T}{\partial x} \tag{5.1.15}$$

where $\lambda(T)$ is the thermal conductivity as a function of temperature. The thermal conductivity for the majority of materials that are utilised at low temperatures can be found in [12, 19, 23].

5.1.4.4 The thermal conduction through gas

If two parallel surfaces have different temperatures, T_1 and T_2, the heat conducted by a low-pressure gas is given by [23]

$$Q = \frac{a_0}{4}\frac{\gamma+1}{\gamma-1}\sqrt{\frac{2R}{\pi M}}\, p \frac{T_2 - T_1}{\sqrt{T_1}} \tag{5.1.16}$$

where M is the molecular weight, γ is the ratio of specific heats of the gas, R is the universal gas constant, p is the pressure and a_0 is a factor characterising surfaces. The factor is defined as

$$a_0 = \frac{a_1 a_2}{a_2 + \left(\frac{A_1}{A_2}\right)(1 - a_2)a_1} \tag{5.1.17}$$

It can vary between 0 and 1. To be on the safe side one can assume a_0 equal to 1. However, for clean polished surfaces and helium gas, a_0 may be as low as 0.025 [23].

5.1.4.5 The thermal radiation

The heat between two parallel surfaces can also be transferred through the thermal radiation. White [23] proposed to evaluate it as follows

$$Q = \sigma A(T_1^4 - T_2^4)\frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 + \varepsilon_2 - \varepsilon_1 \varepsilon_2} \tag{5.1.18}$$

where σ is the Stephan constant, and ε_1 and ε_2 are emissivities of surfaces. The emissivity is a strong function of the material type and the quality of its mechanical finishing. For instance, for rough non-metallic surfaces the emissivity is close to 1, whereas clean polished metallic surfaces have ε as low as 0.01.

Vibrations in the combination with the presence of the magnetic field generate electromagnetic field and eddy currents which can be picked up by sensitive superconductor detectors and the read-out electronics through the inductive coupling or/and the heat propagation. The phenomenon is often called "microphonic noise". It is potentially very damaging. To avoid that superconducting spectrometers employ the magnetic shield and compensation coils around salt pills and combined cryoperm/superconductor shields protecting the primary detector and SQUIDs.

5.1.5 Magnetic Shielding at Low Temperature

Magnetic field is usually characterised using two units. These are the magnetic flux density, B, and the magnetic field strength, H. The International System of Units, SI, quantifies B in teslas (T) and H in A/m. The old CGS system is still very popular among scientists especially in expressing moderate magnetic fields. For convenience, we will give here the conversion factors. These are:

– oersted: $1\,\mathrm{Oe} = 79.57\,\mathrm{A/m}$;
– gauss: $1\,\mathrm{G} = 10^{-4}\,\mathrm{T}$; and
– $1\,\gamma = 1\,\mathrm{nT} = 10^{-5}\,\mathrm{G}$.

The functional dependence between the magnetic flux density and the magnetic field strength in a soft magnetic material is given by

$$B = \mu_0 \mu_\mathrm{r} H \tag{5.1.19}$$

where $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H\,m^{-1}}$ is the magnetic constant, μ_r is the relative permeability of the material.

The shielding factor of any screening configuration is generally defined as a ratio between an increment in the external magnetic field strength, H_e, and an increment in the magnetic field strength inside the shield, H_i, caused by H_e, that is,

$$S = \frac{H_\mathrm{e}}{H_\mathrm{i}} \tag{5.1.20}$$

The analytical computation of S as a function of position in a screen of an arbitrary form is quite complex. The complexity can be avoided by using the numerical methods based on the finite element analysis. A number of numerical simulation programs are now commercially available. Many of them have a user-friendly interface, which merely requires setting up the geometry and boundary conditions.

In practice, the shields are often being manufactured in three simple forms. These are: cylinder with closed ends, cylinder with open ends and, finally, cylinder with one open end. Assuming that the external field is perfectly homogeneous we can roughly evaluate the shielding effectiveness of each configuration.

5.1.5.1 Cylinder with closed ends

Figure 5.11 illustrates the configuration. The static shielding effect with respect to the transversal field, H_t, is given by

$$S_\mathrm{t} = \frac{\mu_\mathrm{r} d}{D} + 1 \tag{5.1.21}$$

where d is the wall thickness of the magnetic material and D is the internal diameter of the cylinder. The equation (5.1.21) works well for

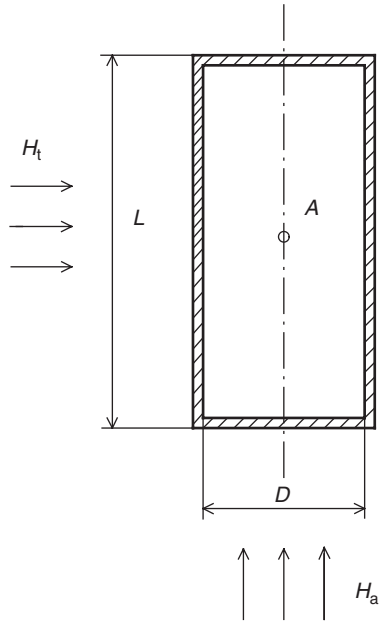


Figure 5.11 The shield configuration as a cylinder with both ends closed

a volume around a central point A with a radius $r < \frac{D}{2}$ provided that a condition $L > D$ is met.

When the external magnetic field is parallel to the axis, the shielding factor takes the form

$$S_a = \frac{4N(S_t - 1)}{1 + \frac{D}{2L}} \tag{5.1.22}$$

where N is the demagnetisation factor. The numerical values of N can be found in [24]. Here we will give 3 values for it: $N = 0.9$ at $L/D = 0.1$; $N = 0.5$ at $L/D = 1$; and $N = 0.2$ at $L/D = 10$.

5.1.5.2 Cylinder with open ends

The analysis of this configuration will refer to Figure 5.12. The attenuation of the transversal external field, H_t, at the centre point of the tube, A, is given by

$$S_{t2} = \frac{S_t S_{tc2}}{S_t + S_{tc2}} \tag{5.1.23}$$

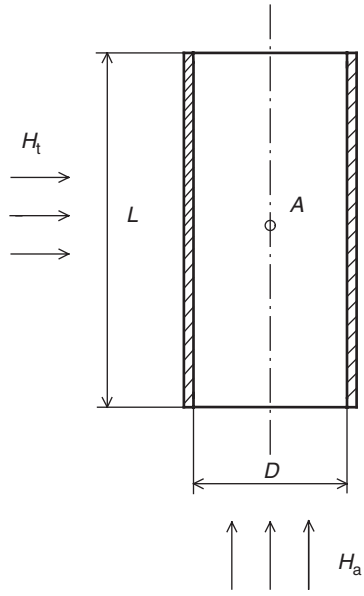


Figure 5.12 The shield configuration as a cylinder with both ends open

whereas that with respect to the axial field has the form

$$S_{a2} = \frac{S_a S_{ac2}}{S_a + S_{ac2}} \tag{5.1.24}$$

In the equations (5.1.23) and (5.1.24), S_{tc2} and S_{ac2} represent correction factors that take into account the field penetration through the open ends. They are defined as

$$S_{tc2} \approx 1.6 \times e^{3.443\frac{L}{D}} \tag{5.1.25}$$

$$S_{ac2} \approx 0.45 \times e^{2\frac{L}{D}} \quad \text{at} \quad L/D > 0.8 \tag{5.1.26}$$

The expressions (5.1.25) and (5.1.26) were restored in the analytical form from the experimental data published in [25].

5.1.5.3 Cylinder with one open end

Shields with one open end are the most widespread in the low-temperature equipment as they allow an easy access to the screened volume. The configuration is shown in Figure 5.13. The total shielding factors are given by

$$\frac{1}{S_{t3}} = \frac{1}{S_t} + \frac{1}{S_{tc3}} \tag{5.1.27}$$

for the transversal component of the field and

$$\frac{1}{S_{a3}} = \frac{1}{S_a} + \frac{1}{S_{ac3}} \tag{5.1.28}$$

According to [24], the correction factors S_{tc3} and S_{ac3} have the form

$$S_{tc3} = 31 \times e^{2.425\frac{x}{D}} \tag{5.1.29}$$

$$S_{ac3} = 0.43 \times e^{4.5\frac{x}{D}} \tag{5.1.30}$$

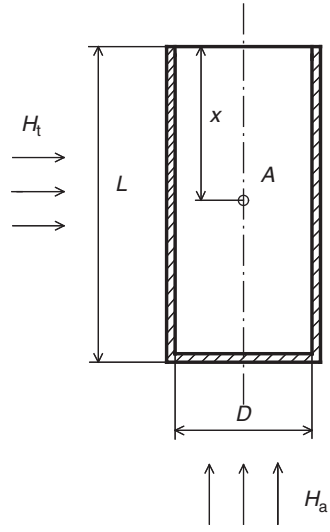


Figure 5.13 The shield configuration as a cylinder with one end closed

The transversal component of the magnetic field has a stronger impact on the overall shielding effectiveness. This fact is sometimes not appreciated by designers.

The most popular magnetic material used at cryogenic temperatures is Cryoperm 10, which was developed by Vacuumschmelze. It is an alloy with a high nickel concentration and has a large initial permeability. Its maximum permeability, however, is achieved over a certain low temperature range in which the shield should operate. The temperature range can be controlled and optimised by appropriate heat treatment and annealing procedures. The material optimised for operation at a liquid nitrogen temperature (78 K) and a liquid helium temperature (4 K) is available as standard options. Cryoperm 10 optimised for operation at 4.2 K has the following major characteristics [25]:

- the maximum permeability 70 000;
- the maximum flux density 0.65 T;
- coercivity 0.03 A/cm;
- saturation flux density 0.8 T;
- Curie temperature 400 °C;
- electrical resistivity 0.55 $\Omega mm^2/m$; and
- density 0.87 g/cm^3.

Cryoperm 10 with the saturation flux density of 0.8 T cannot be used directly for screening the salt pill magnet, which in the two-stage ADR supplies as much as 6 T or more. In order to circumvent the problem the magnet must include dedicated field cancellation coils. The coils can attenuate the stray field to a level well below 0.8 T so that the shield is able to practically eliminate the influence of the ADR magnet on the detector performance.

5.1.5.4 Superconducting shielding

Magnetic shielding with the μ-metal material only is often not sufficient for sensitive superconducting electronics. Even if the shielding factor is adequate, magnetic domains generate magnetic noise that can be picked up by a SQUID preamplifier, thereby, reducing the signal-to-noise ratio. Superconductor shields are free from this drawback. Due to the Meissner effect the superconductor *I* excludes the external magnetic field from the interior as long as its strength is below a certain value, called the first critical magnetic field, H_{c1}.

Unfortunately, a superconducting cylinder cooled down at the presence of an external flux Φ_e, will essentially trap this flux inside completing it to a nearest quantum value $n\Phi_0$. Here n is the integer number. In the superconducting state, the cylinder maintains the trapped flux to a very high accuracy by developing screening currents, I. Fluctuations of the trapped magnetic flux are below a measurable level. What causes a concern is a parasitic signal generated by mechanical vibrations of a sensor in the gradient of the trapped field. In the majority of applications incorporating SQUIDs, it is desirable to minimise the magnetic field inside the shield. This partially can be achieved by special cooling techniques. But in practice, shields combine two cylinders, as shown in Figure 5.14. An internal superconducting cylinder is immersed inside of the μ-metal shield. The μ-metal shield reduces the DC component of the external field, whereas the superconducting cylinder works against time varying fields. Superconducting materials have superior shielding efficiency compared to normal metals until the normal skin depth becomes

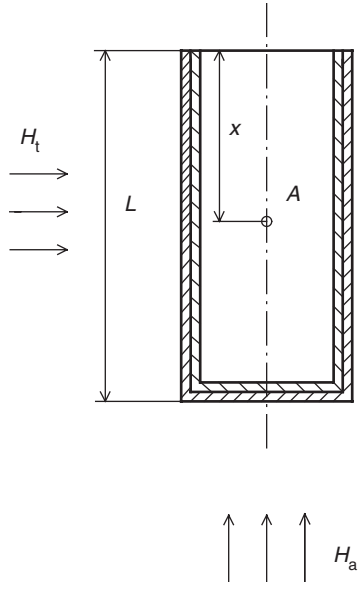


Figure 5.14 A combined μ-metal (outer cylinder) superconducting (inner cylinder) shield. The μ-metal shield reduces the DC component of the external field, whereas the superconducting cylinder works against time varying fields. Superconducting materials have superior shielding efficiency compared to normal metals until the normal skin depth becomes comparable to the field penetration depth. This happens above microwave frequencies. Beyond the cross-over they are as efficient as normal metals at high frequencies and superior at low frequencies down to zero

comparable to the field penetration depth. This happens above microwave frequencies. Beyond the cross-over they are as efficient as normal metals at high frequencies and superior at low frequencies down to zero.

A superconductor cylinder with closed ends represents an ideal shield with a virtually infinite shielding factor. With one end open, attenuations of the axial and transversal components of the magnetic field follow the following equations [26]:

$$H_{\mathrm{ax}} = H_{\mathrm{ae}} \mathrm{e}^{-\frac{3.5x}{r}} \tag{5.1.31}$$

and

$$H_{\mathrm{tx}} = H_{\mathrm{te}} \mathrm{e}^{-\frac{1.8x}{r}} \tag{5.1.32}$$

respectively. The overall shielding factor of the combined shield can be found as a product of individual shielding factors derived from equations (5.1.23), (5.1.27), (5.1.31) and (5.1.32).

5.1.6 Solenoid Inside Cylindrical Superconducting Shield

The principle of STJ detectors requires the presence of a moderate magnetic field to suppress the Josephson supercurrent. Some reports suggest that the magnetic field also helps to reduce extra noise caused by the flux-flow of Abrikosov vortices in transition edge sensors. If a solenoid and a shield have comparable dimensions, the latter has a significant effect on the strength and the uniformity of the field. The uniformity is particularly important at operating detectors involving STJ arrays. As an example, we will consider a configuration shown in Figure 5.15. In this figure, a superconducting cylinder with closed ends surrounds a thin solenoid generating the field. According to [27], the distribution of the flux density inside the solenoid can be approximated by the following equation:

$$B_x(x,y) = -\frac{16\pi NIr}{cL} \sum \left[\cos k_{2n}x \sin \frac{k_{2n}l}{2} I_0(k_{2n}y) K_1(k_{2n}r) \left(\frac{I_1(k_{2n}r) K_1\left(\frac{k_{2n}D}{2}\right)}{K_1(k_{2n}r) I_1\left(\frac{k_{2n}D}{2}\right)} - 1 \right) \right] \tag{5.1.33}$$

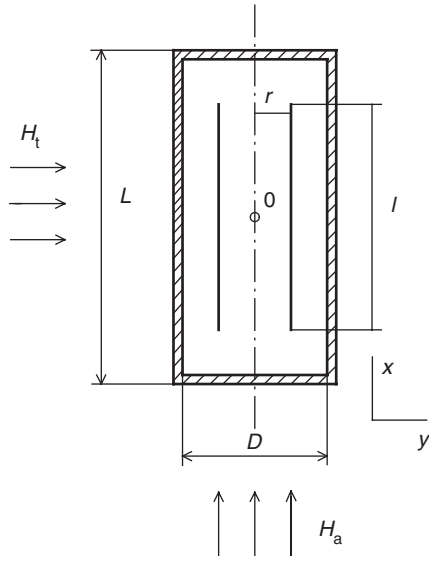


Figure 5.15 A schematic representation of a thin solenoid inside a superconducting shield with both ends closed

where r and I are the radius and the length of the solenoid respectively, D and L are the diameter and the length of the superconducting shield respectively, N is the number of turns in the solenoid, I is the current flowing through the solenoid windings, I_0 and K_0 are modified Bessel functions of order 0, I_1 and K_1 are the modified Bessel functions of order 1, and $k_{2n} = \frac{(2n-1)\pi}{L}$.

The equation (5.1.33) was derived under an assumption that the superconducting cylinder has a perfect shape. Any deviations from this shape will affect $B_x(x, y)$. The Figure 5.16 illustrates this statement. Magnetic field generated by a solenoid near a superconducting surface can also be calculated as a superposition of fields produced by the solenoid itself and its mirror image. For instance, the tilted superconducting surface, shown in Figure 5.16 (b), modifies a value of B_x as well as the pointing vector. An increase in the field component normal to the surface of the detector absorber and/or the STJ itself may lead to trapping vortices, which will modify the response of the detector as the whole.

The annealed lead is normally utilised as the type-I superconducting material for the shield manufacture. Due to its softness such a shield must include a rigid non-magnetic former around which lead is

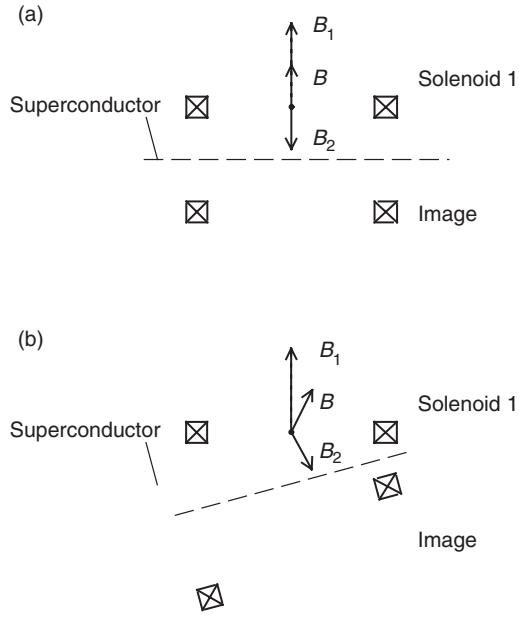


Figure 5.16 Magnetic field of a solenoid perfectly aligned with the shield axis (a) and tilted with respect to the axis (b)

shaped. Sometimes, shields are made of niobium. Niobium is the superconductor of the type-II. Therefore, it traps Abrikosov vortices during the cool down in its pinning centres. The vortices further increase the non-uniformity of the magnetic field inside the solenoid. In addition, they generate magnetic noise that can degrade performance of SQUID amplifiers. Combining μ-metal and superconducting shields together, however, attenuates the ambient field during the cool down and reduces dramatically the number of trapped vortices. It is a good practice to use such an arrangement.

5.1.7 Cryogenic Packaging for Superconducting Electronics

There are three types of major interconnects, which are employed in cryogenic systems. They include superconductor–superconductor, superconductor–normal metal wirings and superconductor–semiconductor interfaces. For instance, a superconductor–superconductor interface is used in arrangements operating with persistent currents, such as

superconducting matching transformers. Other circuits can also employ it in order to eliminate the Johnson noise or the thermoelectric effect. Superconductor–semiconductor interconnects are necessarily present in multipixel detector arrays to accommodate address lines of MOSFET switches and high data rate input–output.

An ULT cryogenic packaging has to meet a number of specific requirements. These are

– reliability at thermocycling (thermal shocks);
– no microphonic noise;
– low inductance;
– low thermal conductivity and so on.

They are not always easily achievable in the practical instruments. For instance, covering the wire with a thicker insulation helps to prevent developing short circuits, but reduces the heat flow from wire to a heat sink. As a result, the wire may stay above its transition temperature. Similar considerations must be taken into account at the design stage of chips.

In general, the on-chip deposition and processing techniques of superconductors and semiconductors are often incompatible. Therefore, the cryogenic electronics bounds to implement packaging concepts that involve some form of the flip-chip bonding. Bonded dies are organised into functional modules. The base chip of each module normally has larger dimensions and contains rows of deposited contact pads along its perimeter. The interconnection scheme between each module and a PCB depends on an application and a required serviceability. For not demanding ground-based research, modules can be repeatedly wire-bonded on the PCB. Alternatively, flipped modules can be electrically connected through beryllium–copper flat springs, as shown in Figure 5.17. Signals from the PCB are brought out of the cryogenic environment via specialised coax cables or soft ribbon cables made of wire looms.

Specialised coax cables employ low thermal conductivity materials for the central lead as well as for the shield braid to reduce the overall thermal load onto the cold stage. Most popular materials are silver-plated steel or/and beryllium–copper. There is a large variety of cable diameters available starting from as small as 20 mils.

Low frequency systems with a bandwidth below several MHz work well with conventional wire looms. They also can work at higher frequencies if covered by a metallisation layer serving as a ground plane. A right choice of dielectric material and its thickness can provide 50 Ω transmission. A similar cable arrangement with deposited conducting

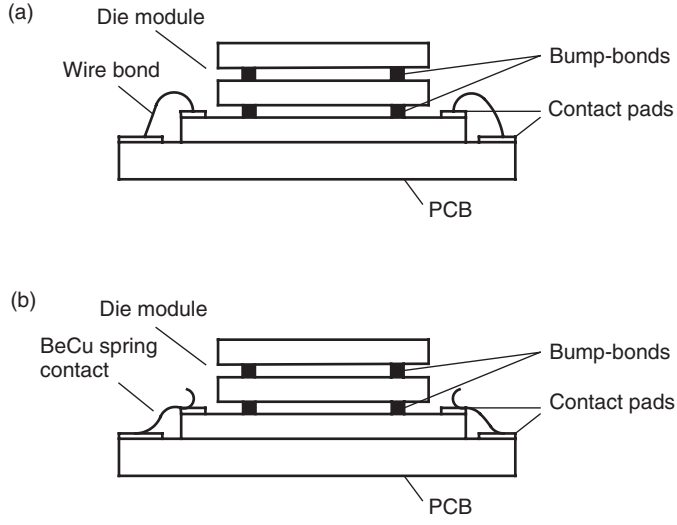


Figure 5.17 Packaging of bump-bonding flip-chips on PCB: with wire-bonding (a), with beryllium–copper flat springs (b)

tracks instead of wires is described in [28]. The cable has extremely low thermal conductivity and operates adequately in multi-GHz I/O connections. A width as low as 1 in. accommodated 127 signal lines. This can be further increased by a factor 9 [28].

The robust die attachment is often based on the conventional bump-bonding technique. For this, chips with contact pads of wettable metal are dipped into a melted solder. If that is not possible, solder spheres of an appropriate diameter covered with wetting adhesive are placed on pads. Then dies get aligned under the microscope to bring their surfaces together into contact. Tacking is achieved under pressure by elevating temperature to a value just below the melting point of the solder. The reflow stage concludes the attachment process. Superconducting interconnects may require two or more reflow processes, as found in [29]. A second reflow without pressure allows the surfaces to self-align avoiding the formation of a possible cantilever effect at die edges.

The bump bonding technique utilises a solder with a low-temperature melting point. This can be InSn (<140 °C) [29], the Wood alloy (65.5 °C) and so on. Some chips, however, cannot withstand even this moderate heating. In this case, the solder bumps are replaced by the conductive epoxy which can be cured at a low temperature. The proximity effect does not work in normal metal pastes. Therefore, interconnects have a finite

resistance at all temperatures. This is not always a disadvantage. Superconductor interconnects have quite high thermal resistance due to the absence of conducting electrons. Normal metal adhesives provide a better heat sinking of upper dies, if required. A superconducting paste has not been developed yet. An average resistivity of In–Sn bumps is approximately 0.1 mΩ at 4.5 K for 0.1 mm square pads. This can be more than one order of magnitude higher for the low-temperature silver loaded epoxy.

5.1.8 Competitive Techniques in the Electron-Probe Nanoanalysis

There are some other techniques apart from X-ray electron-probe analysis and different types of X-ray spectrometers capable of delivering information about a chemical composition of the specimen with a sub-micrometer resolution. In this section, we will discuss briefly the Auger spectroscopy as the closest competitive technique and the wave dispersive spectrometer, whose energy resolution can be comparable with that provided by superconductor detectors.

5.1.8.1 Auger spectroscopy

Similar to X-ray electron-probe analysis the Auger spectroscopy is based on the interaction of the electron beam with the specimen. Figure 5.18 shows a simplified shell structure of the atom, which will help us to explain the principle of the method. An incident electron from the excitation beam with kinetic energy E_p ejects a core electron from, say, K shell in the atom (a). An internal transition in the atom fills the core hole. In Figure 5.18(b), an electron from the L shell takes the vacancy releasing the energy $E_k - E_L$. There is a probability that the photon $E_k - E_L$ will transfer its energy to another electron, for instance, to one on the M shell. This electron, called the Auger electron, is then ejected from the atom available for detection. The energy of the Auger electron is given by

$$E_a = E_k - E_L - E_M^* \tag{5.1.34}$$

where E_M^* is the binding energy of the M shell electron in the presence of a hole in the L shell. Figure 5.18(c) shows the atom in the double ionised state, which is final from a point of view of the Auger spectroscopy. The equation (5.1.34) suggests that the Auger energy is unique

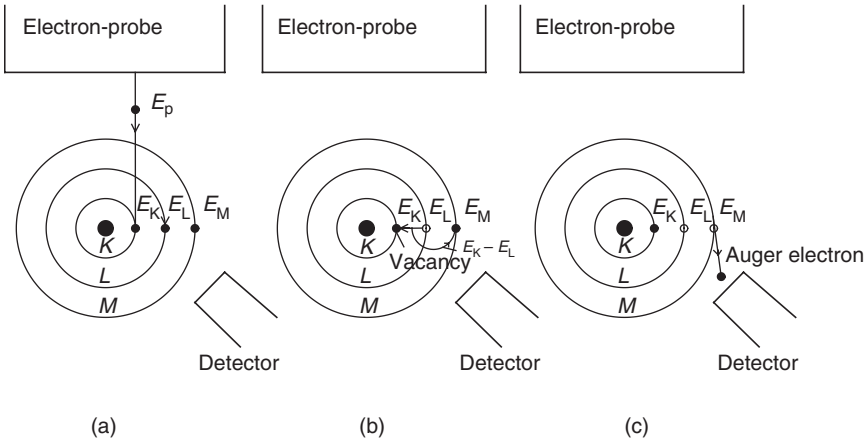


Figure 5.18 A simplified shell structure of the atom, which helps to explain the principle of the Auger spectroscopy. An incident electron from the excitation beam with kinetic energy E_p ejects a core electron from K shell in the atom (a). An internal transition in the atom fills the core hole (b). An electron from the L shell takes the vacancy releasing the energy $E_k - E_L$. There is a probability that the photon $E_k - E_L$ will transfer its energy to another electron, for instance, to one on the M shell. This electron, called the Auger electron, is then ejected from the atom available for the detection

to each atom and the method itself is capable to identify all chemical elements containing more than two shell electrons (i.e. except hydrogen and helium).

The Auger spectroscopy is often called a surface-sensitive technique. This is due to the fact that it provides information to a depth of only 5–10 Å [30]. The Auger electrons from deeper layers do not reach the detector as a result of intense inelastic scattering that occurs in solids with low energy electrons.

The method has a clear advantage if the surface analysis is actually required. However, if the surface is contaminated or oxidised even in this application its capabilities become very limited. The probability of ejecting the Auger electron rather than X-ray photon is quite low. Therefore, the accumulation of the Auger spectrum takes a longer time in the nanoanalysis. The peak to background is worse compared to X-ray electron-probe analysis as well.

The fact that Auger electrons arrive from shallow depths only makes the Auger spectroscopy complementary to the conventional electron-probe analysis when complex multilayer structures are to be studied. It helps, for instance, to identify chemical composition from upper oxidised or contaminated layers. This information can further be used

in interpreting X-ray spectra obtained at different energies of the electron beam.

The technique is most suitable for specimens consisting of light elements, which have a higher yield of Auger electrons and where characteristic X-rays are prone to be masked by the bremstrahlung.

5.1.8.2 Wave dispersive spectrometers

Wave dispersive spectrometers, based on the Braggs invention in 1913, is one of the oldest instruments in the nuclear science. It has been used since the introduction of the first combination of the SEM and the electron-probe X-ray analysis [31]. Before superconductor detectors have appeared, the WDS was the only system capable to resolve between the characteristic peaks in the energy spectrum with accuracy as good as 10 eV FWHM.

Figure 5.19 illustrates the principle of the Bragg spectrometer. As we discussed earlier, the specimen emanates X-ray photons characteristic of its chemical composition as a result of the interaction with the electron-probe beam. Characteristic X-rays hit the Bragg crystal acting essentially as a monochromator, that is it reflects only X-rays with the energy given by

$$E_{\text{Bragg}} = \frac{hc}{2d \sin\theta} \tag{5.1.35}$$

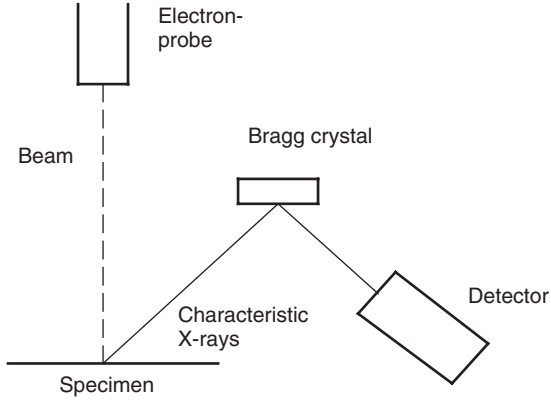


Figure 5.19 The principle of X-ray detection with the Bragg spectrometer

where d is the interplanar spacing of the crystal and θ is the diffraction angle. The intensity of the reflected X-ray is measured with a fast, gas-filled proportional detector. The peak position on the energy spectrum is derived from the diffraction angle, θ, so that the energy resolution of the detector itself becomes largely irrelevant.

According to the equation (5.1.35), the energy resolution of the spectrometer rests on the mechanical precision with which separate components and the assembly, as whole, is manufactured and maintained over time. Oriented crystals must be carefully polished to avoid any mechanical defects on their working surfaces on the atomic scale. Since the take-off angle of characteristic X-rays does not change during the entire measurement at a selected x, y position on the specimen, the diffraction crystal and the gas proportional counter have to move along a focusing curve to be able to satisfy the Braggs condition at different X-ray energies.

The focusing volume where the specimen can be positioned is obviously limited. The actual dimensions of the volume depend on the focusing geometry of a WDS chosen. For instance, using the fully focusing Johansson geometry, the focusing area is represented by an ellipsoid with axes of several mm parallel to the width of the diffraction crystal, but only several μm normal to the diffracting planes. This suggests that the co-ordinates of the measurement spot on the sample relative to spectrometer optics need to be known with the sub-μm accuracy. The knowledge should be maintained throughout the entire analysis.

According to [32], the efficiency of WDS varies with the solid angle subtended by the Braggs crystal, and the reflection efficiency of the crystal and the detection efficiency of the proportional counter. As the Bragg angle, θ, increases, the efficiency shows downward trend, whereas the energy resolution is improving due to the fact that the effect of aberrations of the curved crystal geometry becomes less pronounced. This means that a type of the crystal and a focusing geometry optimised, say, for measurements of well-separated peaks with a much greater intensity than the background will not be optimal for the overlapped low intensity peaks. A general purpose instrument must bear a compromise between the efficiency and the energy resolution. The intercomparison between different techniques and instruments should also involve compromised performance rather than the best major figures of merit ever achieved with one type of the specimen.

Once the required mechanical precision is achieved, an optimum focusing geometry and crystals are selected, and the whole set of instruments operates under the full computer automation all mentioned and some other technical problems become largely irrelevant for a user (not

for a manufacturer though!). A usefulness of a spectrometer is defined by such factors as the quality of the spectrum and its acquisition time. From this point of view, the WDS has never been and hardly will ever become a self-sufficient instrument for the electron-probe analysis. This conclusion results from the time serial nature of its spectral collection. It takes n times as long for the WDS to derive an energy spectrum of the same quality as that obtained with a Superconductor Energy Dispersive Spectrometer (SEDS), where n is the number of bins in the energy spectrum. In the electron-probe microanalysis, the low efficiency of WDS can be partially compensated by operating them at high count rates with large beam currents of several hundreds nA. The gas filled proportional counter can easily operate up to 50 kcps. This, however, is not acceptable in the nanoanalysis: such strong currents would simply damage the nanostructure of interest.

In the electron-probe nanoanalysis with its limited photon generation rate and ultimately low-intensity overlapped peaks, the quantum, geometrical and time efficiencies of the spectrometer become ultimately important. The WDS is good at tracing known elements. It works well in combination with semiconductor EDS providing initial qualitative information about a specimen. The SEDS alone replaces both the instruments. In other words, it is capable to provide a similar or better quality of information about the specimen as both of them at a much shorter time.

5.2 BIOPOLYMER MASS SPECTROMETER

Matrix Assisted Laser Desorption/Ionization (MALDI) and Time-of-Flight mass spectrometers (TOFMS) are probably the most widely utilised tools in the modern biochemistry to analyse the mass distribution in large biological molecules such as proteins and DNA fragments as well as the whole bacteria cells and microorganisms [33]. The principle of the instrument is illustrated in Figure 5.20. The specimen is a single-crystal silicon substrate covered with a film containing molecules of interest embedded in a light absorbing matrix. For instance, in experiments, described in [34], the MALDI-probe consisted of a lysozyme protein solution with a molecular mass $M = 14\,300$ Da mixed in a sinapic acid matrix with $M = 224$ Da (1 Da = 1 amu). After drying in air under the infrared lamp, the matrix takes the form of small crystals with a low melting point. Thus, when the laser pulse with sufficient energy hits the surface of the specimen, affected matrix crystals melt and evaporate,

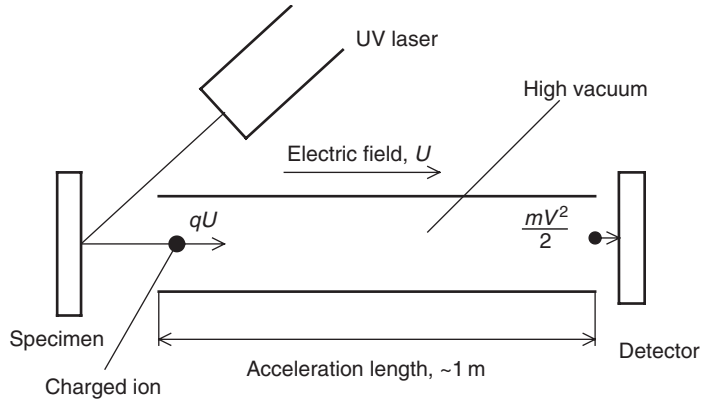


Figure 5.20 The principle of the time-of-flight mass-spectrometer. The specimen is a single-crystal silicon substrate covered with a film containing molecules of interest embedded in a light absorbing matrix. When the laser pulse with sufficient energy hits the surface of the specimen, affected matrix crystals melt and evaporate, thereby freeing unfragmented proteins. Charged molecules are then accelerated in the electric field and fly ballistically over the distance between the sample and the detector

thereby freeing unfragmented proteins. Charged molecules are then accelerated in the electric field and fly ballistically over the distance between the sample and the detector. During the flight a molecule is exposed to a potential energy of the electric field given by

$$E_{\mathrm{p}} = qU \tag{5.2.1}$$

where q is the charge of the molecule, U is the acceleration voltage (usually around 30 kV). Just before hitting the surface of a detector, the molecule gains a kinetic energy

$$E_{\mathrm{k}} = M\frac{\upsilon^2}{2} \tag{5.2.2}$$

where M is the mass of the molecule, υ is its velocity at the end of the flight. The equations (5.2.1) and (5.2.2) suggest that the mass of the molecule is a function of the time-of-flight, t, which can be expressed in a simplified form as

$$M \propto \frac{21qU}{l^2}t^2 \tag{5.2.3}$$

where l is the length of the acceleration tube.

In this application, the energy resolution of the detector is largely irrelevant. It needs to perform only one function, which is the accurate determination of the arrival time of each bio-molecule at its surface. The differentiation of the equation (5.2.3) yields that the mass resolution is directly related to the time-of-flight resolution [35] as follows

$$\frac{\Delta M}{M} \equiv 2\frac{\Delta t}{t} \tag{5.2.4}$$

Conventional MALDI–TOF spectrometers employ micro-channel plates (MCPs) to detect the arrival times of molecular ions. Their principle is based on the secondary electron emission effect. An incident charge molecule interacts with the surface of the detector. There is a probability that the interaction will result in the ejection from an atom of a high energy photoelectron. Scattering of the photoelectron on other atoms of the detector sensitive volume generates an avalanche of secondary electrons, which are collected by a charge sensitive amplifier. An example of the output signal of the preamplifier is shown in Figure 5.21. Micro-channel plates, like any other ionization detectors, are intrinsically very fast. The rise time of pulses can be as short as 10 ns. Therefore, a digitised amplifier output practically does not need additional data

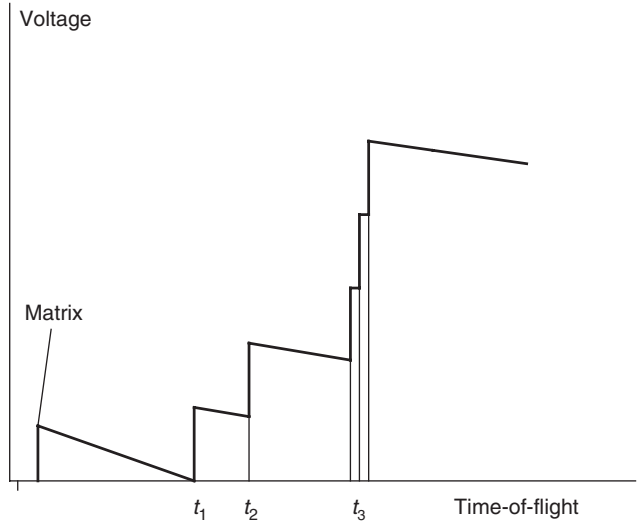


Figure 5.21 A schematic example of the output signal of a charge sensitive preamplifier in response to molecules hitting the active surface of the detector

processing to derive the final mass histogram illustrated schematically by Figure 5.22.

All equally charged molecules interacting with the detector have the same kinetic energy, E_k. As the mass of a molecule increases its velocity and together with it the probability to eject the primary photoelectron have downward trends. There is a characteristic critical mass for each MCP, above which its efficiency becomes too low for the practical analysis. D. Twerenbold recognised the problem and was the first to propose cryogenic detectors as an effective solution [36].

Unlike MCPs, the cryogenic detectors are energy sensitive devices. On impact, molecules transfer their kinetic energies gained in the electrostatic acceleration field to the absorber lattice creating relatively high energy phonons. In the superconductor absorber, the process of phonon relaxation breaks Cooper pairs creating quasiparticles that diffuse towards the tunnel junction. In the normal metal absorber, phonons raise temperature of the electron system measured by TES. The calorimetric sensitivity of the cryogenic detectors makes their response independent of the molecule mass. A practically 100% efficiency can be maintained up to virtually any size of a charged/neutral particle as long as it does not incur mechanical damage to the detector.

The fact that cryogenic detectors provide information on particle arrival time and its kinetic energy adds the whole new dimension in

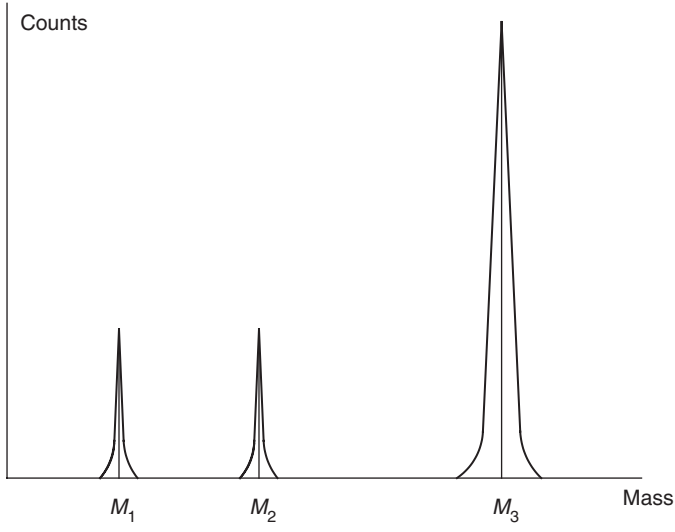


Figure 5.22 A schematic representation of the mass histogram produced by a time-of-flight mass-spectrometer

the mass-spectrometry. It opens a new exciting opportunity to analyse samples involving complex substances. The matter is that after melting, the matrix molecules do not exist just as single-charged monomers producing a clean detector response. Ions can consist of single-double-charged monomers, dimers and even more complex combinations. For instance, MCPs would see a double-charged monomer purely as a half-mass molecule. The cryogenic detector is capable to recognise the difference because the ion would deposit a doubled kinetic energy into the absorber compared to single-charged monomer. This is illustrated in Figure 5.23. The figure presents an energy spectrum obtained from monomers of BSA molecules. The left peak corresponds to the double-charged monomers. The Figure 5.24 shows the pulse height histogram of cytochrom C (12 kDa). The large peak corresponds to the single-charged monomers whereas the second peak is produced by single-charged dimers.

The development of a procedure for the preparation of a new specimen is probably the most responsible and difficult part of TOF analysis. Laser pulses with non-optimum energy, a non-optimal matrix may lead to either the presence of a large number of multimolecule conglomerats, or, alternatively, the fragmentation of molecules themselves. The ability of cryogenic detectors to measure the kinetic energy of molecules resolves the ambiguity regarding the mass-to-charge ratio. This immediately

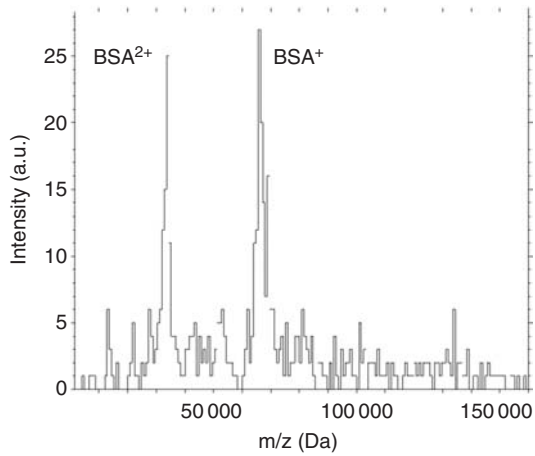


Figure 5.23 A spectrum obtained from monomers of BSA molecules. The left peak corresponds to the double-charged monomers (reproduced by permission from S. Uchaikin (Max-Planck Institute, Munich))

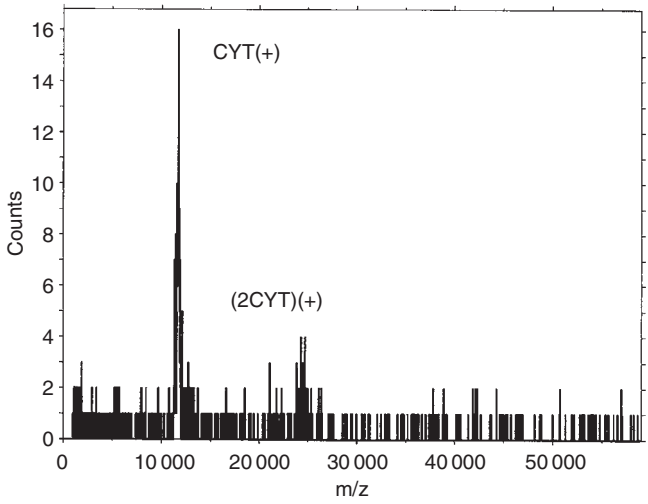


Figure 5.24 A pulse height histogram of cytochrom C (12 kDa) (reproduced by permission from S. Uchaikin (Max-Planck Institute, Munich)). The large peak corresponds to the single-charged monomers whereas the second peak is produced by single-charged dimers

helps to restore the true picture and speed up the sample preparation process.

A diameter of the molecular beam can be as large as several millimeters. In order to be efficient the cryogenic detectors must cover the whole cross-section of the beam. There are two ways to address the matter: either to use a large sapphire absorber [37] or to employ an array of smaller size detectors. The second approach is particularly attractive to combat the pile-up problem when dealing with large ion fluxes. A cryogenic detector will not be able to distinguish, for instance, on event-by-event basis the arrival of charged ions within the rise time of the pulse at the output of a preamplifier. Therefore, there is a much larger probability that each separate pixel of the array interacts with a single individual ion at a time.

Several groups developed mass spectrometers based on transition edge sensors and superconducting tunnel junctions. The tunnel junction is perhaps a more meritable option due to a higher temperature of operation and a faster response. For instance, the applicability of Nb–Al_2O_3–Nb tunnel junction detector operating at 1.3 K was evaluated in [38]. Output pulses of the detector had a rise time of 500 ns and a decay time of approximately 1.5 μs. The signal-to-noise ratio of the pulses enabled author to reach a time resolution as low as 100 ns. The figure yields an ultimate mass resolution $M/\Delta M$ of up to 1000 at 66 kDa and a 25 kV acceleration field.

More recent experiments demonstrated that the rise time of STJ detectors can be further reduced to a value of 20 ns [39]. Still the rise time was found to be limited by the bandwidth of the fast current amplifier rather than the detector itself.

A new detection scheme for STJs was proposed in [40] using an additional Josephson junction as a fast current discriminator. In this scheme, the time resolution is ultimately limited by the intrinsic non-equilibrium characteristics of the primary sensor, which can be as low as few ns.

A large mass sensitivity of cryogenic detector has been proven up to 750 kDa so far where it maintained a 100% efficiency [41]. The figure will surely be increased in the future.

Further information on the application can be found in the review paper [42] and references therein.

REFERENCES

[1] R. Fitzgerald, K. Keil and K. Heinrich *Science* **159**, 528 (1968).
[2] D. Wollman, K. Irwin, G. Hilton *et al. J. of Microscopy* **188**, 196 (1997).
[3] W. H. Tait. *Radiation Detection*. Butterworth, 1980, p. 20.
[4] H. Kramers. *Phil. Mag*. **46**, 836 (1923).
[5] S. Reed. *Electron Microprobe Analysis*. Cambridge University Press, 1975, p. 12.
[6] P. Statham. *J. Res. Natl. Inst. Stand. Technol*. **107**, 531 (2002).
[7] K. Kanaya and S. Okayama. *J. Phys. D.: Appl. Phys*. **5**, 43 (1972).
[8] D.E. Newbury. *X-Ray Microanalysis* **22**, 345 (2000).
[9] D. Newbury. *J. Res. Natl. Inst. Stand. Technol*. **107**, 605 (2002).
[10] T. Ziebold. *Anal. Chem*. **39**, 858 (1967).
[11] T. Zielbold and R. Ogielvie. *Anal. Chem*. **36**, 322 (1964).
[12] G. Kaye and T. Laby. Tables of Physical and Chemical Constants, Longman, 1995.
[13] S. Larsson, P. Engstrom, A. Rindly *et al. Adv. X-Ray Anal*. **33**, 623 (1990).
[14] K. Janssen, B. Vekemans, L. Vincze *et al. Spectrochim Acta Part B* **51**, 1661 (1996).
[15] M. Haschke, W. Scholz, U. Theis *et al. J. Phys. IV France* **12**, 83 (2002).
[16] S. Bichlmeier, K. Janssens, J. Hechel *et al. X-Ray Spectrometry* **30**, 8 (2201).
[17] D. Redfern, J. Nicolosi, J. Hoehne *et al. J. Res. Natl. Inst. Stand. Technol*. **107**, 621 (2002).
[18] C. Wang and P. Gifford. Performance Characteristics of 4 K Pulse Tube in Current Applications, in: *Cryocoolers II*, Kluwer Academic/Plenum Publishers, New York, pp. 803–808 (2001).
[19] F. Pobel. Matters and methods at low temperatures, Springer-Verlag, Heidelberg, 1992.
[20] R.C. Richardson, E.N. Smith, eds, Experimental techniques in condensed matter physics, Addison-Wesley, New York, 1988.
[21] C. Hagmann and P. Richards. *Cryogenics* **34**, 221 (1994).

[22] P. Bromiley. PhD Thesis, University of London (1999).
[23] G.K. White. Experimental techniques in low temperature physics, Oxford University Press, London, 1958.
[24] A. Mayer. Entwicklung von weichmagnetische Legierungen mit besonderen Eigenschaften; BMFT-Forschungbericht, 314 (1977).
[25] Magnetic shielding. Vacuumschmelze GmbH, Hanau, FS-M9 (1989).
[26] B. Deaver and W. Goree. *Rev. Sci. Instrum.* **38**, 311 (1967).
[27] K. Rigby. *Rev. Sci. Instrum.* **59**, 156 (1988).
[28] T. Tighe, G. Akerling and A. Smith. *Proceedings of ASC'98* (1998).
[29] K. Yokoyama, G. Akerling, A. Smith and M. Wire. *IEEE Trans. Appl. Supercond.* **7**, 2631 (1997).
[30] P. Palmberg and T. Rhodin, *J. Appl. Phys.* **39**, 2425 (1968).
[31] R. Castaing. Application of Electron Probes to local Chemical and Crystallographic Analysis. PhD Thesis, University of Paris, 1951.
[32] S. Reed. *J. Res. Natl. Inst. Stand. Technol.* **107**, 497 (2002).
[33] T.L. Williams, D. Andrzejwski, J. Lay and S. Musser. *J. of American Society for Mass Spectrometry* **14**, 342 (2003).
[34] D. Twerenbold, J. Vuilleumier, D. Gerber *et al. Appl. Phys. Lett.* **68**, 3503 (1996).
[35] E. Esposito, R. Cristiano, S. Pagano *et al. Physica C* **372–376**, 423 (2002).
[36] D. Twerenbold, US patent no. 5,460,010 (1997).
[37] S.V. Uchaikin, P. Christ, F. Probst, S. Rutzinger and W. Seidel. *Physica C: Superconductivity* **367**, 295 (2002).
[38] M. Frank *et al. Rapid Commun. Mass Spectrom.* **10**, 1946 (1996).
[39] M. Kishimoto *et al. Nucl. Instr. Meth.* **A444**, 124 (2000).
[40] E. Esposito, M. Eirnaes, S. Pagano *et al.* Appl. Phys. Lett. **82**, 2109 (2003).
[41] S. Labov *et al.* Proceedings of LTD-7, Max-Planck Institute for Physics, Munich, **82**, 242 (1997).
[42] M. Frank. *Nuclear Instruments and Methods in Physics Research* **A444**, 375 (2000).

6

Selected Topics of Analysis and Synthesis of Detector Systems

INTRODUCTION

In previous chapters, we assumed that readers were familiar with the network analysis methods, the theory of control systems as well as numerical calculation techniques of superconducting thin-film structures. This, obviously, does not have to be the case. Therefore, in the concluding chapter, we will brief on methods and techniques relevant to the subjects discussed in this chapter. The review will also give us an opportunity to introduce examples of some other nuclear electronic circuits, which did not fit in the scopes of any of previous chapters but play important roles in detector systems. Here, we have in mind components of the pulse recognition channels, analogue conditioning units, digital filtration, the timing circuitry, SQUID amplifiers employing the modulation–demodulation principle and so on.

Originally, we did not plan to include this chapter in the book. Our perception was that powerful SPICE-based computer simulation software had long ago made obsolete manual engineering network analysis routines. Recently, I have communicated this thought to a colleague of mine. He agreed that this was partially true with regard to the direct analysis problems when our goal would be to derive properties of a designed electronic network. However, the standard software is often

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin
 ISBN: 0-470-85759-5

powerless in solving the inverse problem. Under the inverse problems we mean the synthesis or the optimisation of a system or a network from a known transfer function or a desired time response and so on. Moreover, the modern intelligent adaptive systems must optimise their structures automatically depending on various stimuli and operating conditions. Therefore, it is still left up to a designer to outline a set of suitable networks and to derive their (as complete as possible) analytical descriptions, which can be embedded into processor algorithms or hard-wired logic. This must be preceded by precise instructions on which and when a certain network configuration or set of network component values, or program routines, and so on must be utilised.

These general remarks are relevant for both the analogue and digital hardware designs. Digital hardware designers learned to circumvent the problem by applying iterative algorithms. This works quite well in the absence of time constraints. Such algorithms are obviously of no use if it has completed by, for instance, a time a missile had missed its target or was interrupted by the fact that the missile had been destroyed by a faster weapon. Thus, if a system must react quickly, the pre-defined direct algorithms, or the hardwired logic or analogue networks remain the best option.

Besides, the modern computer simulation software has its limits even in dealing with the direct analysis of some systems. Mixed signal analogue/digital networks and synchronous modulation–demodulation systems are the two examples that would require creative analytical skills of an electronic system designer. The matter is that the performance of such systems can often be evaluated in a series of indirect experiments only. These experiments may require the development of non-standard test rigs and procedures so that the model could be verified experimentally. The detailed knowledge of analytical tools helps to perform these tasks efficiently.

The chapter consists of three major sections. Section 6.1 deals with the analogue continuous systems. We will discuss transfer functions, signal flow graphs, design of low-pass filters with the optimised time and frequency response, the analysis and synthesis methods by Nyquist and Bode. Then we will show how to derive a describing function of non-linear elements present in the system. The latter is particularly relevant to the superconductor electronics, since all superconductor detectors and the SQUID read-out electronics represent non-linear systems. In Section 6.2, we will discuss the z-transformation, the w-plane and the method of equivalent transfer matrices intended to analyse the analogue/digital circuitry and systems with periodically variable parameters. Analysis

and design of digital filters, the SQUID amplifiers employing the modulation–demodulation technique and the digital SQUIDs cannot be done without these techniques. Section 6.3 is devoted to the variational numerical method developed by W. Chang [1]. The method is based on the finite element analysis, with the number of elements sufficient to assume the current distribution within the each element to be constant. Chang showed that the momentary distribution of currents in the superconductors provide the minimum sum of the total magnetostatic energy and the total kinetic energy of the structure. Therefore, the current distribution and the inductance can be recovered from the minimised total energy distribution via a solution of the inverse problem. The method is implemented numerically and proved to be very accurate. It is widely used to calculate complex structures of analogue and digital superconducting components.

We will conclude the chapter by presenting a practical algorithm to calculate all major parameters of the superconducting quantum interference device.

6.1 ANALOGUE ELECTRONIC CIRCUITRY ANALYSIS AND DESIGN PRINCIPLES

6.1.1 Laplace transform: Transfer Functions of Electronic Networks

As we saw in the previous chapters, all electronic circuits of any complexity can be represented by a block diagram. The Figure 6.1 shows a block diagram of a simple resistive divider. It includes an input voltage

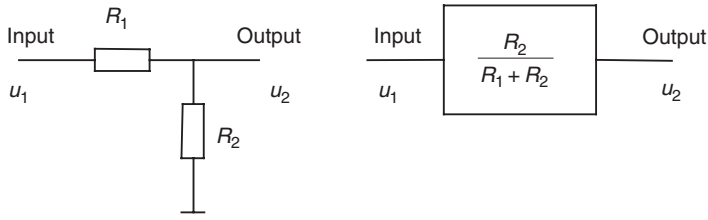


Figure 6.1 A circuit diagram and an equivalent block diagram of a simple resistive divider. The block diagram includes an input voltage terminal, an output voltage terminal and shows the relationship between the input and the output, $T = \frac{u_2}{u_1} = \frac{R_2}{R_1 + R_2}$

terminal, an output voltage terminal and shows the relationship between the input and the output:

$$T = \frac{u_2}{u_1} = \frac{R_2}{R_1 + R_2}$$

In general, the input can be any stimulus applied to the system. For instance, the input of a logic device is normally a command code that causes the device to perform some logical operation. In this case, the block diagram would give a pictorial representation of the logical cause–effect action.

All electronic devices can be classified into two types of systems:

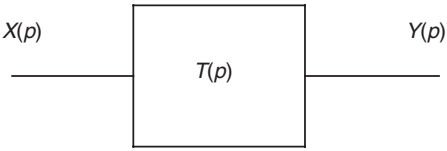
1. the open loop (or direct action) systems; and
2. the systems containing one or several feedback loops.

The resistive divider shown in Figure 6.1 represents a simple direct action system, while Figure 2.43 illustrates a closed loop control system of a SQUID amplifier. The latter includes a forward path, $F(\phi)K(p)$, a feedback path, β, and a summing point indicating that its output is the algebraic sum of the input flux, ϕ_e, and feedback signal, ϕ_β.

The cause–effect functional dependence, $T(p)$, between the input and the output is referred to as the transfer function of the block diagram. For the time-invariant linear systems with zero initial conditions, the transfer function can be defined as the ratio of the Laplace transform of the response to the Laplace transform of the input signal, that is,

$$T(p) = \frac{Y(p)}{X(p)} = \frac{\sum_{i=0}^{m} b_i p^i}{\sum_{i=0}^{n} a_i p^i} \tag{6.1.1}$$

where a_i and b_i are the coefficients of $X(p)$ and $Y(p)$ polynomials (Figure 6.2). Poles and zeroes of the Laplace transfer function carry important information on the stability and dynamic properties of a system. Therefore, it is advisable to factorise the polynomials so



$$T(p) = \frac{Y(p)}{X(p)} = \frac{\sum_{i=0}^{m} b_i p^j}{\sum_{i=0}^{n} a_i p^j}$$

Figure 6.2 The block diagram of a time-invariant linear system/network with a Laplace transfer function $T(p)$

that their poles and zeroes are presented in an explicit form as follows

$$T(p) = \frac{\prod_{i=0}^{m} b_i(p - \nu_i)}{\prod_{i=0}^{m} a_i(p - \lambda_i)} \tag{6.1.2}$$

where λ_i and ν_i are poles and zeroes of the network $T(p)$, respectively.

The Laplace transforms of commonly used input signals in the nuclear electronics are summarised in Table 6.1. The Laplace transforms of other functions can be found either in more detailed tables [2] or conservatively from the Laplace integral

$$X(p) = L[x(t)] = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-pt}\mathrm{d}t \tag{6.1.3}$$

with $p = c + j\omega$ being the complex frequency.

Table 6.1 Laplace and Fourier transforms of input signals commonly used in nuclear electronics

Time function	Laplace transform	Fourier transform
Dirac function, $\delta(t)$	1	1
Unit step, 1(t)	$\frac{1}{p}$	–
Exponent, e^{-at}	$\frac{1}{p+a}$	$\frac{1}{j\omega+a}$
Sine wave, $\sin \Omega t$	$\frac{\Omega}{p^2+\Omega^2}$	$\frac{\Omega}{\Omega^2-\omega^2}$
Cosine wave, $\cos \Omega t$	$\frac{p}{p^2+\Omega^2}$	$\frac{j\omega}{\Omega^2-\omega}$

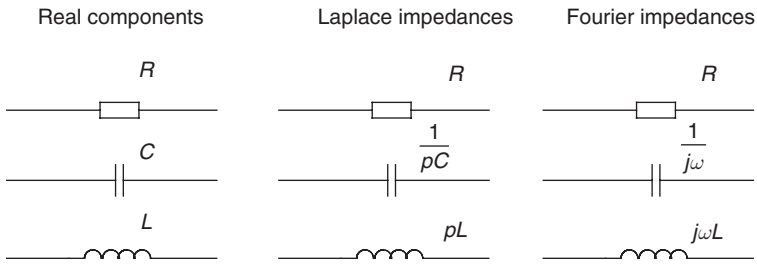


Figure 6.3 The passive electronic components and their equivalent representations in the Laplace domain and the frequency (Fourier) domain

The time response of a network is derived from the inverse Laplace transform of the product $X(p)T(p)$ as follows:

$$y(t) = L^{-1}[X(p)T(p)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(p)T(p)e^{pt}dp \qquad (6.1.4)$$

Authors of [2] also compiled the inverse Laplace transform pairs of a large variety of polynomials, which resulted from analysis of practical electronic networks published elsewhere in scientific papers.

The direct Laplace transform representation of a transfer function of any circuit containing only lumped components can be derived straight from the circuit by replacing resistances, R, capacitances, C, and inductances, L, by impedances $Z_R(p) = R$, $Z_c(p) = \frac{1}{pC}$ and $Z_L(p) = pL$, respectively (Figure 6.3). Main properties of the Laplace transform itself are summarised in Table 6.2. They suggest that the Ohm and Kirchhoff laws and other techniques of the DC current analysis are applicable for modified circuits. This will be demonstrated in the following exercises.

Exercise 6.1

In this exercise, we will find a transfer function of an L, C and R passive network, shown in Figure 6.4(a).

First step is to draw an equivalent circuit diagram of the network in the Laplace domain. To achieve that, we replace components by their impedances $Z_L(p)$, $Z_C(p)$ and $Z_R(p)$ (Figure 6.4b). $Z_C(p)$ and $Z_R(p)$ are connected in parallel. They can be replaced by a single component in

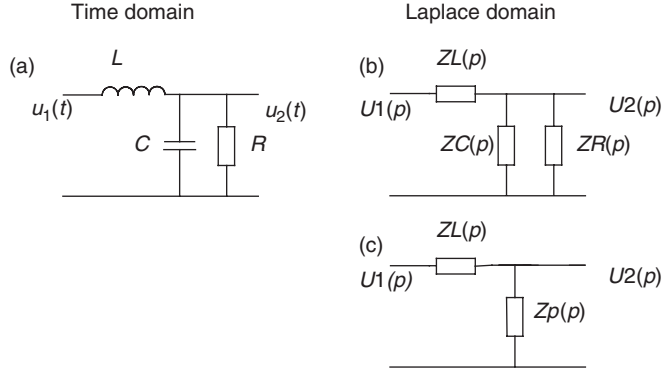


Figure 6.4 An L–C–R network and its equivalent circuit diagrams in the Laplace domain. Methods developed for the analysis of DC circuits are applicable in the Laplace domain for networks exposed to varying input signals

exactly the same way, as we would do with two parallel resistors in DC current analysis, that is,

$$Zp(p) = \frac{Z_c(p)Z_R(p)}{Z_c + Z_R(p)}$$

Figure 6.4(c) illustrates the minimised-circuit diagram of the network in the Laplace domain. This is an ordinary divider, which transfer function is given by

$$T(p) = \frac{Zp(p)}{Z_L(p) + Z(p)(p)} = \frac{1}{LC}\frac{1}{(p-\alpha_1)(p-\alpha_2)}$$

where $\alpha_{1,2} = -\frac{1}{2CR} \pm \left(\frac{1}{4C^2R^2} - \frac{1}{C}\right)^{1/2}$ are its poles.

Exercise 6.2

In this example, we will find a step response of an RC-differentiating circuit using the Laplace transform. The circuit is shown in Figure 6.5.

According to Table 6.1, the Laplace transform of the unit step signal $1(t)$ is $U_1(p) = \frac{1}{p}$. The transfer function of the circuit is given by

$$T(p) = \frac{Z_R(p)}{Z_R(p) + Z_c(p)} = \frac{p}{1 + p\tau}$$

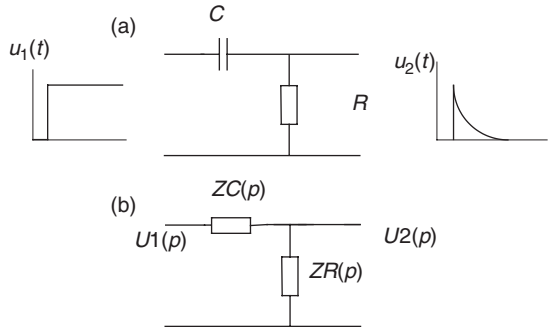


Figure 6.5 The step response of the passive C–R-differentiating network

where $\tau = RC$. The output signal in the time domain is derived from the inverse Laplace transform of the product $U_1(p)T(p)$, that is,

$$u_2(t) = L^{-1}\left[\frac{p}{p(1+p\tau)}\right] = \frac{1}{\tau}\mathrm{e}^{-\frac{t}{\tau}}$$

6.1.2 Signal Flow Graphs

A signal flow graph is a schematic representation of the signal propagation through a system of interest. Under the system, we imply any electronic arrangement ranging from a simple passive network to complex logic sequences. Flow graphs essentially embed a full set of equations describing the dynamics of a system. However, unlike dealing with pure mathematical symbols, grapho-analytical methods carry the information, which is clear and meaningful for the electronic design purposes. We will see that graphs are not just more illustrative, but also easier to manipulate. They enable us to derive transfer functions of quite complex systems without resorting to mathematical equations, just following few simple formal rules.

A signal flow graph gives a pictorial representation to a set of simple equations.

$$X_i = T_{ij}Y_j \tag{6.1.5}$$

X_i, Y_j are the input and output variables, T_{ij} is the mathematical operator mapping the relationship between X_i and Y_j in the equation (6.1.5).

In the signal flow graph, the operator T_{ij} is termed the transmission function. In this respect, the transfer function of an electronic network is a special case of the transmission function.

Let us introduce some other important terminologies and properties associated with signal flow graphs using a simple divider network shown in Figure 6.6. The Laplace network (Figure 6.6a) is represented first as a block diagram (Figure 6.6b) with the transfer function $T(p)$ and then as a signal flow graph (Figure 6.6c) with variables $U1(p)$, $U2(p)$ and a transmission function $T(p)$. This, in fact, is the simplest form of a signal flow graph. Each variable is designated by a dot called a node. The line with the arrow corresponds to the transmission function. It is referred to as a branch of the signal flow graph. The arrow points in the direction of the signal propagation.

Figure 6.6 demonstrates that the signal flow graph has a close relationship with the block diagram. In the majority of cases, they can be directly derived from each other.

More complex signal flow graphs contain a number of other elements. We give their explanations in Figure 6.7. Figure 6.7(a) shows a general example of the signal flow graph. It contains nine branches and a number of *paths*. Each path includes one-directional branches that have

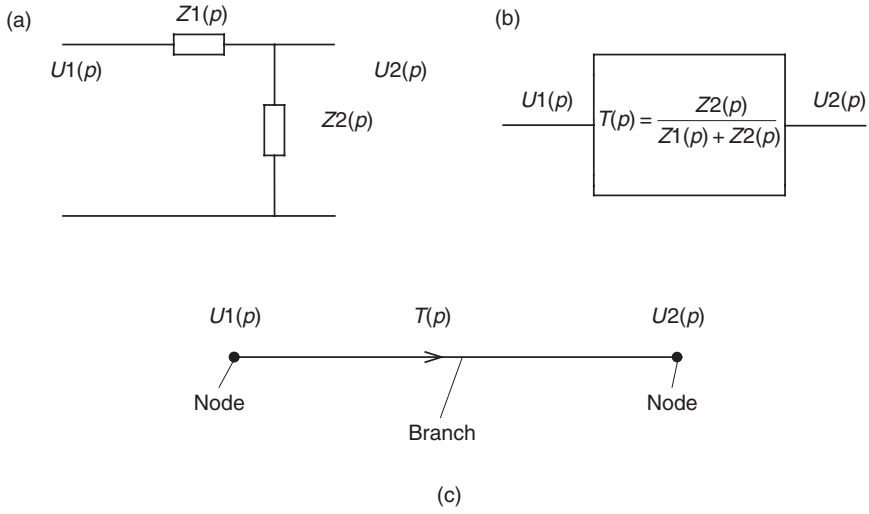


Figure 6.6 The Laplace network (a) represented as a block diagram (b) with the transfer function $T(p)$ and as a signal flow graph (c) with variables $U1(p)$, $U2(p)$ and a transmission function $T(p)$. Each variable is designated by a dot called node. The line with the arrow corresponds to the transmission function. It is referred to as a branch. The branch arrow points in the direction of the signal propagation

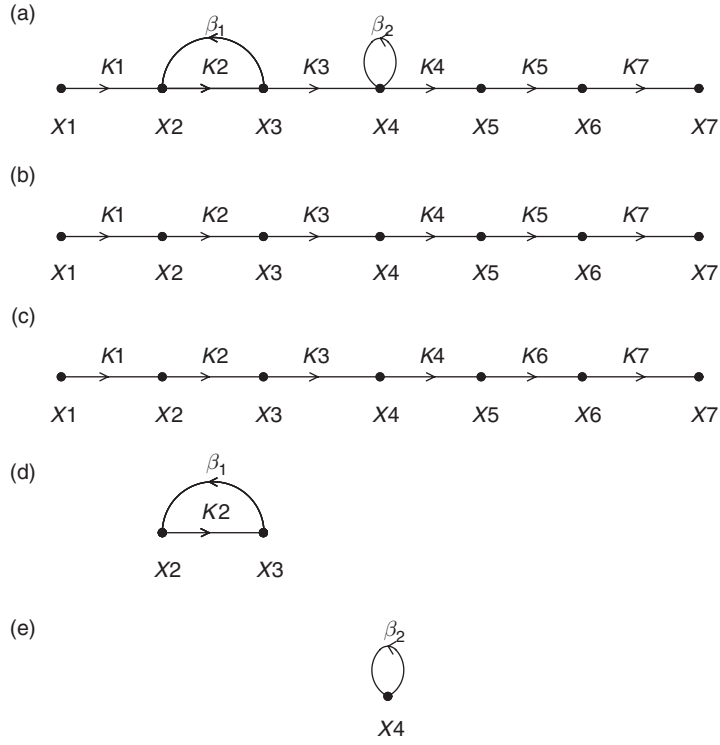


Figure 6.7 A general example of the signal flow graph (a). It contains nine branches and two complete paths depicted as separate diagrams (b, c). Each path includes one-directional branches that have common nodes. In a path, a signal should not pass any of these branches more than once. Long paths (b, c) can be broken into a number of partial paths. (d, e) illustrate the feedback loop and self-loop of the graph, respectively

common nodes. In a path, a signal should not pass any of these branches more than once. The signal flow graph in Figure 6.7(a) has, in fact, four complete paths depicted separately in Figure 6.7(b,c,d,e). Obviously, long paths (Figure 6.7b,c) can be broken into a number of partial paths. The shortest path would be one that consists of a single branch.

The node $X1$ in Figure 6.7a–c represents an input node of the graph. It is called *a source*. By definition, source can have only outgoing branches. Sometimes, an additional unity branch needs to be added to the graph so that its input node would satisfy the definition of the source. We will demonstrate that in the forthcoming examples.

The node $X7$ in the same graph is called an outgoing node or *a sink*. The sink is connected only to incoming branches. The sources and sinks

contain between each other forward paths directed from the source to the sink. The *path gain* of a forward path is a product of the transmission functions of individual branches involved. For instance, a path gain of the forward path (Figure 6.7b) is given by

$$K_{F1} = K1 \cdot K2 \cdot K3 \cdot K4 \cdot K5 \cdot K7 \tag{6.1.6}$$

whereas that of the forward path shown in Figure 6.7(c) is

$$K_{F2} = K1 \cdot K2 \cdot K3 \cdot K4 \cdot K6 \cdot K7 \tag{6.1.7}$$

A bi-directional path (Figure 6.7d) originates and terminates on the same node $X2$. Such paths form the so-called *feedback loops*. The loop gain is a product of transmission functions of branches that constitute the loop. As an example, a loop gain of the feedback loop (Figure 6.7d) is given by

$$F_1 = K2 \cdot \beta_1 \tag{6.1.8}$$

The smallest feedback loop may consist of a single branch that starts and ends on the same node. It is termed *a self-loop*. Figure 6.7(e) shows an extracted self-loop, which is formed by a node $X4$ and a branch β_2. The definition for the self-loop gain is the same as for any feedback loop. In the example, it is $F_2 = \beta_2$.

One of the objectives of the electronic network analysis is to derive transfer functions between various nodes. As we mentioned before, there is a number of other methods and techniques to do that. In what follows, we will show that signal flow graphs enable us to write down the input–output relationship of a network of any complexity by applying relatively simple formal rules. The overall transfer function of any signal flow graph can be derived from the following equation:

$$T = \frac{\sum_{i=1}^{n} K_i \Delta_i}{\Delta} \tag{6.1.9}$$

where K_i is the gain of ith forward path, Δ is the signal flow graph determinant. The determinant is defined as

$$\Delta = 1 - (-1)^{k+1} \sum_{j=1}^{l} \sum_{k=1}^{m} F_{jk} \tag{6.1.10}$$

where F_{jk} is the jth possible product of k loop gains of possible pair combinations of loops that do not touch each other or, in other words, loops that do not share the same nodes.

For instance, let us assume that a network is described by a signal flow graph with two non-touching loops, that is, $m = 2$. Then the determinant will take the following form

$$\Delta = 1 - \sum_{j=1}^{l} F_{j1} + \sum_{j=1}^{l} F_{j2} \tag{6.1.11}$$

where $\sum_{j=1}^{l} F_{j1}$ is the sum of all existing loop gains and $\sum_{j=1}^{l} F_{j2}$ is the sum of loop gains of all possible pair combinations of loops that do not share common nodes.

The last unidentified term in the equation (6.1.9) is Δ_i. It is given by

$$\Delta_i = \Delta - \sum_{i=1}^{n} F_i \tag{6.1.12}$$

where $\sum_{i=1}^{n} F_i$ is a sum of loop gains of feedback loops that share nodes with forward paths K_i.

Further, we will give several useful examples, which illustrate the applicability of the signal flow graph technique.

Exercise 6.3

In this exercise, we will try to draw a signal flow graph for a divider network presented in Figure 6.8(a) and derive its transfer function using equations (6.1.9–6.1.12).

The network is described with the following pair of equations:

$$I(p) = \frac{1}{Z1(p)} U1(p) - \frac{1}{Z1(p)} U2(p) = K1(p) \cdot U1(p) - \beta 1(p) U2(p)$$

$$U2(p) = I(p) Z2(p) = K2(p) I(p)$$

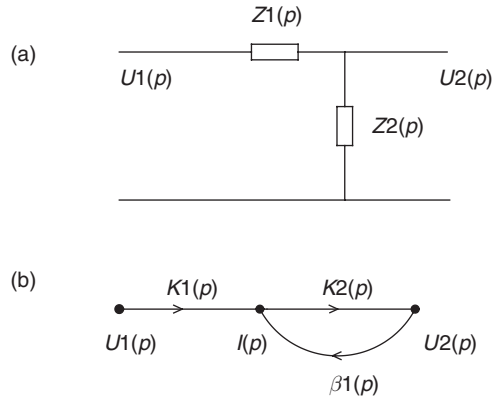


Figure 6.8 A single-stage divider network (a) and its signal flow graph in the Laplace domain (b)

These equations lead to the signal flow graph presented in Figure 6.8(b). There is only one forward path with a gain $K_1 = K1(p)K2(p)$ and a single feedback loop that touches the forward path. Here we immediately can write down: $\Delta_1 = 1$, $F_{11} = -\beta 1K2$. The signal flow graph determinant is $\Delta = 1 - (-\beta 1K2)$. Thus, the transfer function takes the well-known final form

$$T(p) = \frac{K1}{1 - F_{11}} = \frac{K1(p)K2(p)}{1 + \beta 1(p)K2(p)} = \frac{Z2(p)}{Z1(p) + Z2(p)}$$

Exercise 6.4

In this exercise, we will analyse two divider networks connected in series, as shown in Figure 6.9(a).

Following the Kirchhoff laws, we write down four equations describing the electric dynamics of the network. These are:

$$\begin{aligned} I1(p) &= \left(\frac{1}{Z1}\right) U1(p) - \left(\frac{1}{Z1}\right) U2(p) \\ U2(p) &= I1(p)Z2 - I2(p)Z2 \\ I2(p) &= \left(\frac{1}{Z3}\right) U2(p) - \left(\frac{1}{Z3}\right) U3(p) \\ U3(p) &= I2(p)Z4 \end{aligned}$$

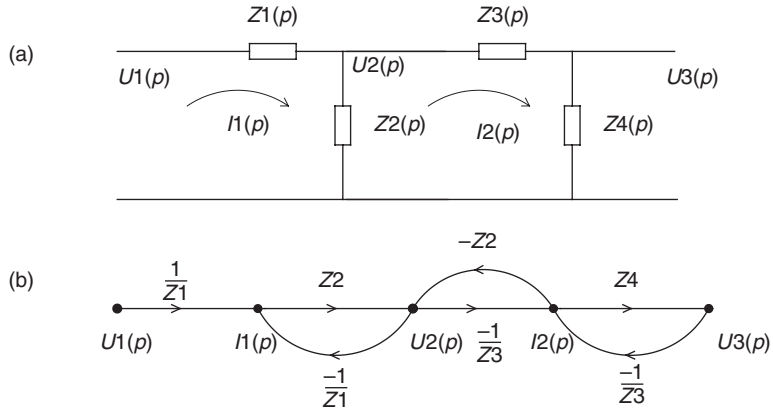


Figure 6.9 A double-stage divider network (a) and its signal flow graph in the Laplace domain (b)

The equations have five variables: $U1$, $I1$, $U2$, $I2$ and $U3$, which become the nodes of the signal flow graph. The transmission functions of branches are derived from appropriate coefficients of the equations. The signal flow graph is shown in Figure 6.9(b).

Again, there is one forward path in the graph with a gain given by

$$K_1 = \frac{Z2Z4}{Z1Z3}$$

but three feedback loops, two of which do not share common nodes. Their loop gains are

$$F_{11} = -\frac{Z2}{Z1}; \qquad F_{21} = -\frac{Z2}{Z3}; \qquad F_{31} = -\frac{Z4}{Z3}$$
$$F_{12} = F_{11}F_{31} = \frac{Z2Z4}{Z1Z3}$$

Substituting loop gains into equation (6.1.11), we derive the graph determinant in the following form

$$\Delta = 1 - (F_{11} + F_{21} + F_{31}) + F_{12} = 1 + \frac{Z2}{Z1} + \frac{Z2}{Z3} + \frac{Z4}{Z3} + \frac{Z2Z4}{Z1Z3}$$

Since all loops share nodes with the forward path, the factor $\Delta_1 = 1$. Using the equation (6.1.9), we derive the transfer function as

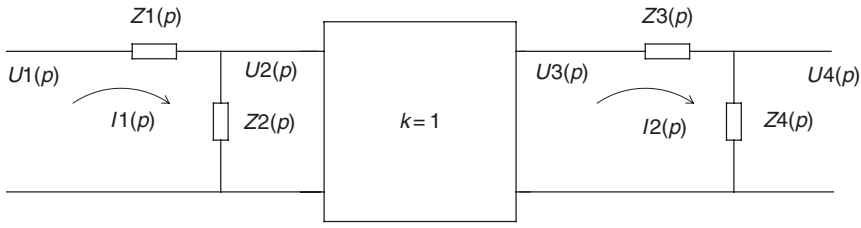


Figure 6.10 A double-stage divider network with a buffer eliminating the loading effect. The overall transfer function of the network is given by a product of transfer functions of the individual stages

$$T(p) = \frac{K_1}{\Delta} = \frac{Z2Z4}{Z1Z3\left(1 + \frac{Z2}{Z1} + \frac{Z2}{Z3} + \frac{Z4}{Z3} + \frac{Z2Z4}{Z1Z3}\right)}$$

Note that the transfer function of two connected in-series dividers is not equal to the product of their individual transfer functions, $T1(p)T2(p)$. This results from the loading effect of $(Z3 + Z4)$ on the impedance $Z2$. In order to achieve $T(p) = T1(p)T2(p)$ the network must employ a decoupling unit with the unity gain, as well as high input and low output impedances to eliminate the loading effect, as shown in Figure 6.10.

Exercise 6.5

In this exercise, we will derive a transfer function of an active network represented by an operational amplifier with a negative parallel voltage feedback, as shown in Figure 6.11.

We assume that input and output impedances of the preamplifier are $Z_{\text{in}} \to \infty$, $Z_{\text{out}} \to 0$ respectively. If no current flows into the amplifier input terminal, the direct $U1/U2$ and feedback $U3/U2$ branches take the following form:

$$\gamma = \frac{Z2}{Z1 + Z2}, \qquad \beta = \frac{Z1}{Z1 + Z2}.$$

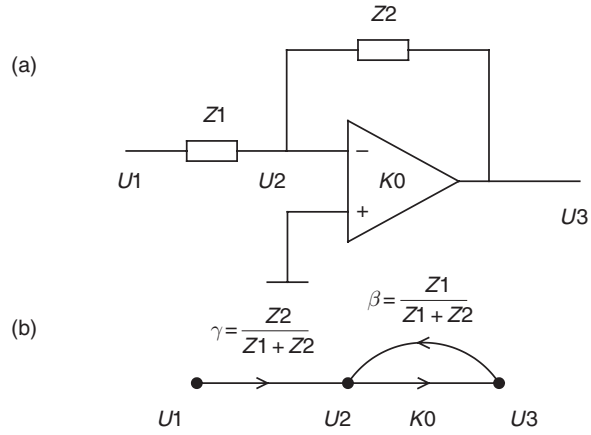


Figure 6.11 An active network based on an operational amplifier with a negative parallel voltage feedback (a) and its signal flow graph (b)

The whole signal flow graph of the amplifier with the feedback is shown in Figure 6.11(b). According to the equation (6.1.9), a transfer function is given by

$$K = \frac{K_0 \gamma}{1 + K_0 \beta}$$

Assuming $K0 \to \infty$, it reduces to $K = \dfrac{Z2}{Z1}$.

Exercise 6.6

In this example, we will discuss a mutual relation between a block diagram and a signal flow graph representing the same system.

Let us write down a transfer function of a SQUID amplifier with a block diagram presented in Figure 2.43. The block diagram already contains units (blocks) with defined transfer functions. These can be directly translated into transmission functions of branches of a signal flow graph, as shown in Figure 6.12. In order to accommodate the summing point, one has to introduce a unity gain branch between the input node ϕ_e and the summing point output node ϕ.

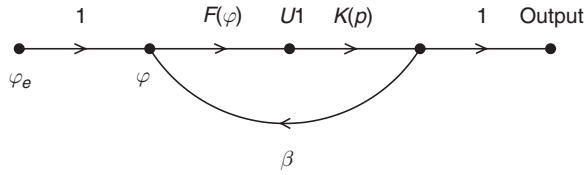


Figure 6.12 A signal flow graph of the SQUID amplifier with the block diagram presented in Figure 2.43. In order to accommodate the summing point, one has to introduce a unity gain branch between the input node ϕ_e and the summing point output node ϕ

The input–output transfer function is given by

$$T(p) = \frac{F(\phi)K(p)}{1 + F(\phi)K(p)\beta}$$

In practice, real networks will be described by a large amount of characteristic equations. In this case, an original signal flow graph takes a rather complex form comprising tens to hundreds of nodes. Before attempting to derive the transfer function it is advisable at first to try to simplify the graph. One can state that it is possible to reduce a majority of single input–single output systems to a canonical form presented in Figure 6.13. This can be done by using simplification rules. Figure 6.14 summarises the rules.

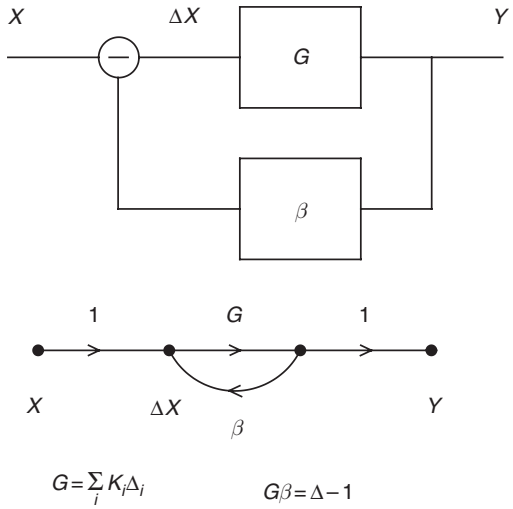


Figure 6.13 A canonical form of a feedback system. The majority of linear single input-single output systems can be reduced to this form

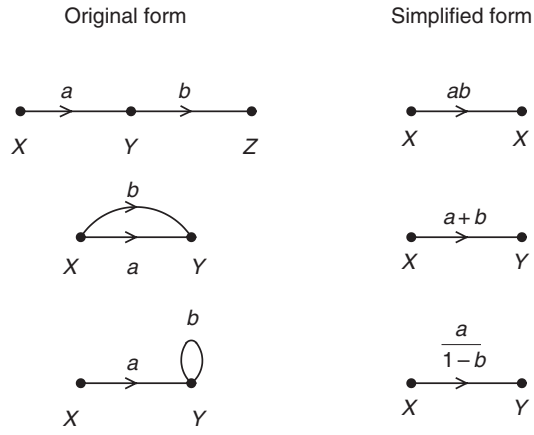


Figure 6.14 A summary of rules that enable the reduction of complex signal flow graphs to the canonical form shown in Figure 6.13

6.1.3 Low-Pass Filters

Low-pass analogue filters are utilised in the main pulse-processing channel prior the pulse digitisation. Their frequency characteristics are designed to minimise the aliasing effect (see Chapter 4 for details). In the analogue recognition channels, filters reduce the time jitter of triggers caused by high-frequency noise.

Conceptually, an analogue filter represents a frequency-selective electrical/electronic network that is intended to attenuate a portion of the signal frequency spectrum. The range of frequencies in which an input signal passes without considerable attenuation is called the passband, whereas the rest of the frequency range is called the stopband. Figure 6.15(a) represents a frequency response of an ideal low-pass filter with an abrupt roll-off of the amplitude-frequency characteristic between the passband and the stopband. A practical realisation of such a filter is not possible, as it would require involvement of an infinite number of electronic components. Therefore, real filters must always have some compromised specifications. For instance, in the frequency domain, compromised specifications would primarily be concerned with tolerances for the attenuation factor in the passband and the stopband and with the steepness of the roll-off of the frequency characteristics between the passband and the stopband, as shown in Figure 6.15(b). In this figure, $A^p_{\min}$, $A^p_{\max}$ and $A^s_{\min}$, $A^s_{\max}$ represent the minimum and maximum attenuation factors in the passband and the stopband, respectively.

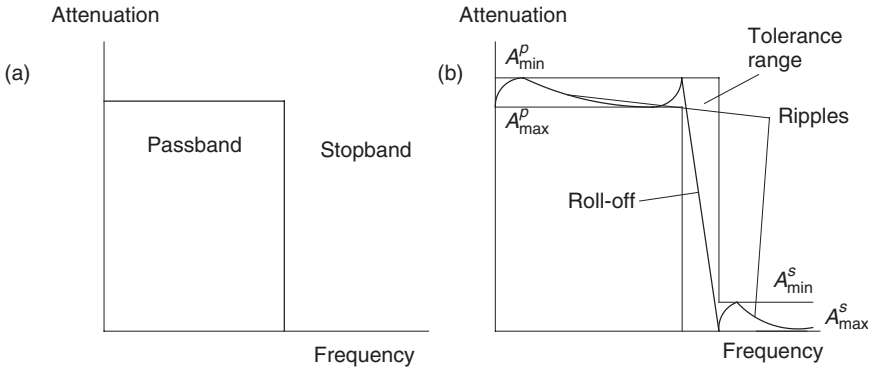


Figure 6.15 A frequency response of an ideal low-pass filter with an abrupt roll-off of the amplitude–frequency characteristics between the passband and the stopband (a). A practical realisation of such a filter is not possible, as it would require involvement of an infinite number of electronic components. (b) shows a frequency response of a low-pass filter with compromised specifications. In this figure, A^p_{min}, A^p_{max}, and A^s_{min}, A^s_{max} represent the minimum and maximum attenuation factors in the passband and the stopband respectively

The synthesis of an analogue filter is divided into two phases:

Phase 1 The derivation of an approximate transfer function of a filter that is expected to be within a predefined tolerance range.

Phase 2 The circuit synthesis involving the process of converting the transfer function approximated earlier at the Phase 1 into an electrical network.

6.1.3.1 Approximation problem (Phase 1)

The approximation procedure depends on requirement specifications and operating conditions of a designed filter as well as a class of input signals to be processed. In general, these can be classified into four major categories:

1. when specifications are defined for the attenuation factors in the passband and the stopband only. This is the easiest case, since the whole design can be performed entirely in the frequency domain using tabulated pole positions published in numerous specialised handbooks;

2. when a filter must operate as a delay line without specified (or broadly specified) attenuation factors;
3. when a filter must meet stringent requirements on both the attenuation factors and the frequency-dependent delay or the phase linearity; and
4. when requirement specifications are set for the attenuation factors and the impulse or step response. Filters with simultaneous requirements on the frequency and time response are rarely treated in the literature. This case, however, is the most relevant to filter networks implemented in the detector spectrometers.

In the both applications of the analogue filters, whether in the main channel in the recognition unit of a spectrometer, we are certainly looking for the best achievable rejection of high-frequency noise. At the same time, the filter should have a little effect on the original pulse shape and maintain a stable baseline in the whole range of the input count rates. The problem has already been discussed elsewhere in Chapter 4 with regard to the anti-aliasing pre-filter. Similarly, such factors as the presence of overshoots, ringing effect, a relatively long decay in the combination with a finite (variable) rise time of the input pulses and so on, alter the dynamics of the recognition channels. Figure 6.16 illustrates the statement. It shows three pulses propagating from the output of a filter

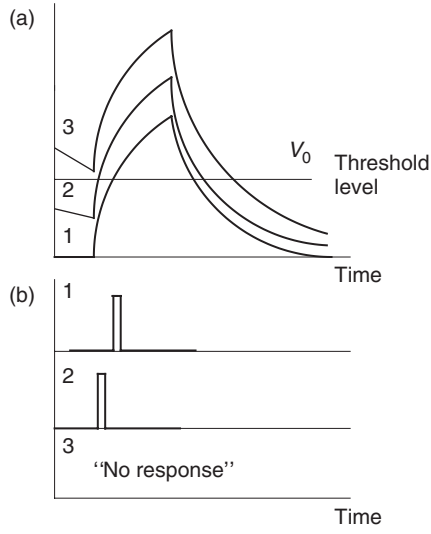


Figure 6.16 Pulse shapes at the input of recognition comparator. Timing of the comparator is affected by the instability of the base line

to the input of a trigger indicating the pulse arrival. Pulse 1 starts from "zero" baseline, whereas pulses 2 and 3 represent pile-upped pulses starting below and above the threshold level of the recognition channel. The pulse 1 illustrates the reference case. A moment of time t_0 when the pulse signal magnitude intersects a pre-defined threshold level of the discriminator, the latter delivers the output as a logic pulse for a system to start the processing. The pulse 2 starts from a finite level above zero, but below V_0. Its arrival is registered but with a different timing. This phenomenon is known as time "slewing" or time "walk". The way in which this may be of importance is best illustrated by considering distributed arrays. Distributed arrays deliver two or more output signals generated by the same nuclear event into designated electrodes. The ratio of amplitudes of the pulses depends on a position where an incoming photon or a particle interacts with the absorber. This magnifies the pile-up effect on the time jitter. Each event yields pulses of what will be termed as true coincidences. However, a certain number of accidental coincidnces can go through the discriminator gate in Figure 6.16. If we assume the coincidence gate has a resolving time of $\pm\tau$, the processor can pair wrong pulses that pass through the channel within a time τ of one another.

In the ultimate case (pulse 3 in Figure 6.16), the channel does not react on the pulse arrival. Therefore, it will be ignored by the main processing channel.

If a filter to be designed contains only lumped components, its transfer function can always be represented by a ratio of two polynomials given by equations (6.1.1) and (6.1.2). Some poles and zeroes are complex, but in practical stable filters, they appear in complex conjugate pairs. Thus, we can re-write the equations in the following form:

$$T(p) = \frac{\prod_{i=1}^{m} k_i(p - \upsilon_i)(p - \upsilon_i^*)}{\prod_{i=1}^{n}(p - \lambda_i)(p - \lambda_i^*)} = \frac{N(p)}{D(p)} \tag{6.1.13}$$

where p is a complex frequency defined as $p = \sigma + j\omega$. The real part of p, σ, can be interpreted as a damping factor, whereas the quantity ω in the imaginary part of p is the angular frequency. $\upsilon_i, \upsilon_i^*, \lambda_i, \lambda_i^*$ are complex variables, zeroes and poles, occurring in complex conjugate pairs, that is,

$$\upsilon_i = \mathrm{Re}(\upsilon_i) + j\mathrm{Im}(\upsilon_i); \qquad \upsilon_i^* = \mathrm{Re}(\upsilon_i) - j\mathrm{Im}(\upsilon_i)$$
$$\lambda_i = \mathrm{Re}(\lambda_i) + j\mathrm{Im}(\lambda_i); \qquad \lambda_i^* = \mathrm{Re}(\lambda_i) - j\mathrm{Im}(\lambda_i)$$

All poles and zeroes are often shown on a plot of the p-plane. Figure 6.17 gives an example of the positioning of a complex conjugate zero pair υ_i, υ_i^* and a real pole λ_i.

The position of poles on the p-plane characterises the stability of a system described by the transfer factor $T(p)$. We distinguish three states in which systems can operate: absolutely stable, conditionally stable and unstable. Therefore, it is sensible to divide the p-plane into three characteristic regions, as shown in Figure 6.18:

Region 1 Points to the left of $j\omega$-axis, but not including the $j\omega$-axis itself. It will be referred to as the left-hand plane.

Region 2 Points to the right of $j\omega$-axis, but not including the $j\omega$-axis itself and is referred to as the left-hand plane.

Region 3 Points on the $j\omega$-axis including the zero co-ordinates, $p = 0$.

Example 1 If a filter transfer function contains a single real pole situated on the negative real axis, the filter is absolutely stable. Moreover, it returns to its stable state without ripples (oscillatory behaviour) as shown in Figure 6.19.

Example 2 If a system contains a pair of conjugate poles in the left-hand p-plane it remains to be stable, but acquires an oscillatory behaviour, as shown in Figure 6.20.

Example 3 Poles situated on the right-hand plane characterise an unstable system, as shown in Figure 6.21 with responses $y(t) = Ae^{\alpha t}$ (a real positive pole), and $y(t) = Ae^{\alpha t}\cos(\omega t + \theta)$ (a pair of conjugate poles).

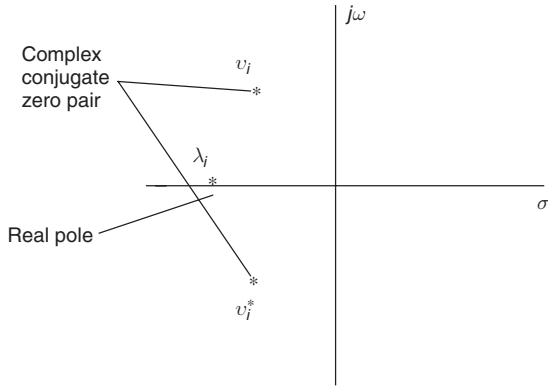


Figure 6.17 A generic example illustrating the positions of a complex conjugate zero pair υ_i, υ_i^* and a real pole λ_i on the p-plane

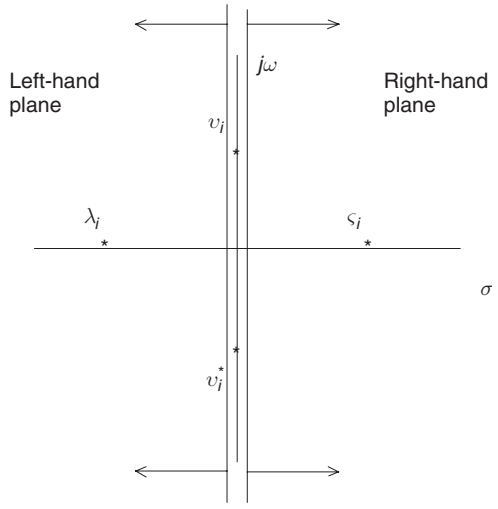


Figure 6.18 The *p*-plane divided into three characteristic regions: Region 1 embraces points to the left of $j\omega$-axis, but not including the $j\omega$-axis itself. It will be referred to as the left-hand plane; Region 2 with points to the right of $j\omega$-axis, but not including the $j\omega$-axis itself and is referred to as the right-hand plane; Region 3 restricted to the $j\omega$-axis including the zero co-ordinates, $p = 0$

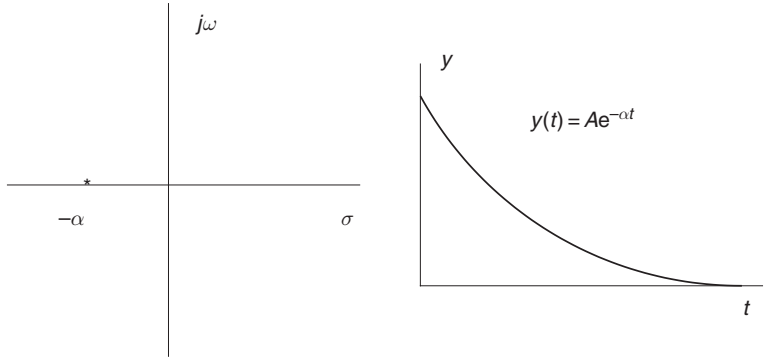


Figure 6.19 The time response of a filter network with its transfer function contains a single real pole situated on the negative real axis. The filter is absolutely stable returning to its original state without ripples

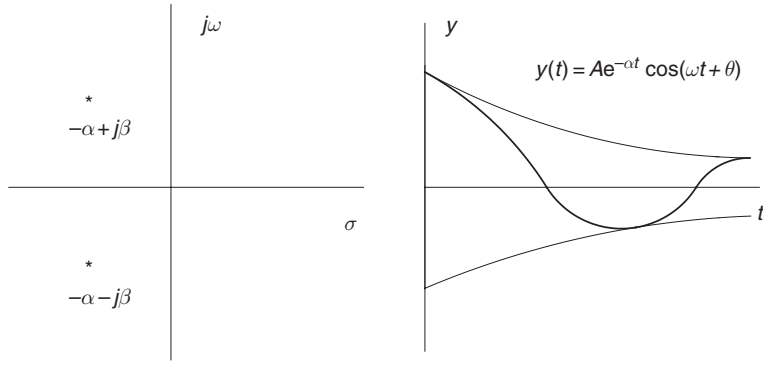


Figure 6.20 The time response of a system with a pair of conjugate poles located in the left-hand of the *p*-plane. It remains to be stable, but acquires an oscillatory behaviour

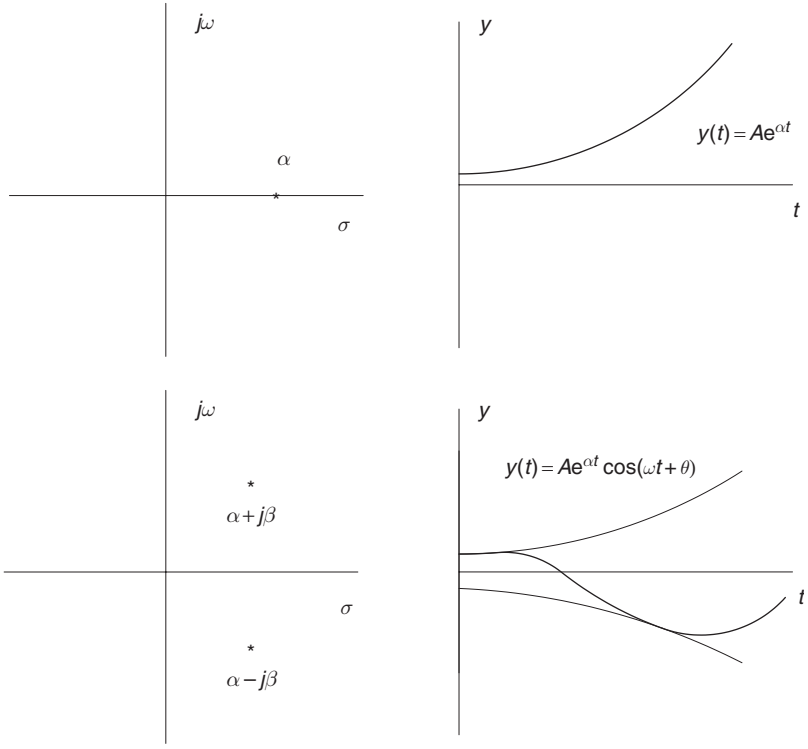


Figure 6.21 The poles situated on the right-hand plane characterise an unstable system, with responses $y(t) = Ae^{\alpha t}$ (a real positive pole) and $y(t) = Ae^{\alpha t}\cos(\omega t + \theta)$ (a pair of conjugate poles)

6.1.3.2 Circuit synthesis (Phase 2)

In order to simplify the synthesis process, filters are usually realised by the cascading of filter units (blocks). Therefore, the overall transfer function of a filter can be expressed as a product of transfer functions of the individual units:

$$T(p) = \prod_{i=1}^{N} T_i(p) \tag{6.1.14}$$

Base filter units can comprise a first-order filter, a second-order filter and, more rarely, a third-order filter. Thus, an all-pole filter of the fifth order can be broken either into two units of the second and third orders or, alternatively, into three units including two second-order filters and one first-order filter and so on. In this respect, the filter synthesis is similar to all other inverse problems. A filter that must meet certain requirement specifications can be realised in a number of circuit solutions.

In what follows, we will briefly introduce all the three base filter units starting from the first-order unit.

First-order filter unit

There are two commonly used basic configurations of the first-order active filters: one with a non-inverting gain, shown in Figure 6.22(a), and another one with an inverting gain, shown in Figure 6.22(b).

First-order filters have a single real pole situated in the left-hand p-plane. In both the configurations the capacitor is defined as

$$C = \frac{1}{\lambda_0} \tag{6.1.15}$$

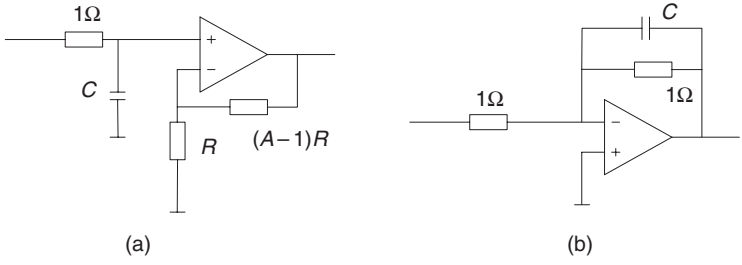


Figure 6.22 Two commonly used basic configurations of the first-order active filters: one with a non-inverting gain (a) and another one with an inverting gain (b)

where λ_0 is the normalised real axis pole.

A normalised low-pass filter is a filter with a 3-dB cut-off frequency 1 rad/s. The normalisation enables us to use existing curves and tables published in specialised manuscripts. A given filter response can be scaled to a different frequency range by multiplying values of all reactive components in the filter unit by a frequency scaling factor. This statement remains true for the second- and third-order filter units as well. The frequency-scaling factor is defined as

$$\eta = \frac{1\,\text{rad/s}}{\omega_{3\text{dB}}} \tag{6.1.16}$$

where $\omega_{3\text{dB}}$ is the real 3-dB cut-off frequency of the designed filter.

In a more general case, when we have designed a filter for a certain 3-dB cut-off frequency point $(\omega_{3\text{dB}})_1$, but would like to shift it to a point $(\omega_{3\text{dB}})_2$, the definition for the frequency scaling factor becomes as follows

$$\eta = \frac{(\omega_{3\text{dB}})_1}{(\omega_{3\text{dB}})_2} \tag{6.1.17}$$

Tuning the cut-off frequency requires adjusting component values in the circuit. In practice, this is easier to perform with resistors rather than with reactive elements. Besides, using a $1\,\Omega$ resistor in the feedback loop is not practical since an operational amplifier would have to deliver a 1-A current for a 1-V output voltage. In order to deal with these and other similar issues, we apply the impedance-scaling theorem. According to the theorem, a linear RC-network maintains its original transfer function if all its resistors are multiplied and all its capacitors are divided by the same impedance scaling factor, ζ.

Thus, a de-normalised value of capacitor in the first-order filter block is given by

$$C_{\text{dn}} = C\frac{\eta}{\zeta} \tag{6.1.18}$$

Second-order filter units

A two-pole low-pass filter unit is a building block whose transfer function is given by

$$T_i(p) = k_i \frac{1}{(p - \lambda_i)(p - \lambda_i^*)} \tag{6.1.19}$$

Sometimes these filters are called biquads (bi-quadratic filter cells). Biquads based on a single operational amplifier are also often referred to as "Sallen–Key" filters acknowledging the earliest publication on this subject by R. Sallen and E. Key [3]. Modern biquads, however, are not the same as early networks calculated by these authors.

The most commonly used biquads are those based on the positive-feedback configurations with a gain factor k. Figure 6.23 illustrates a generic block diagram and a signal flow graph of the configuration.

In the figure, the transfer functions are defined as follows:

$$D(p) = \frac{k\omega_{\mathrm{p}}}{p^2 + 2\gamma p + \omega_{\mathrm{p}}^2} \tag{6.1.20}$$

$$F(p) = \frac{p}{p^2 + 2\gamma p + \omega_{\mathrm{p}}^2} \tag{6.1.21}$$

where $\gamma = \frac{\omega_{\mathrm{p}}}{2Q_{\mathrm{p}}}$, ω_{p} is the resonance frequency (or the frequency of forced oscillations) and Q is the resonance quality factor of the filter.

The transfer function of the block diagram depicted in Figure 6.23 is given by

$$T(p) = \frac{1}{k} D(p) \frac{k}{1 - kF(p)} = \frac{D(p)}{1 - kF(p)} \tag{6.1.22}$$

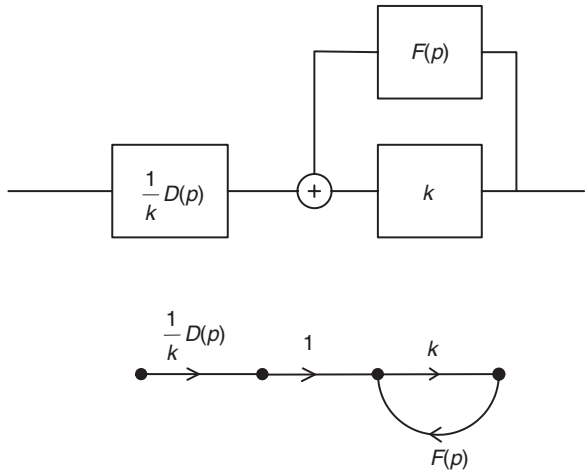


Figure 6.23 A generic block diagram and a signal flow graph of the second-order filter unit (biquad) based on the positive feedback

The amplification k is a quantity by what the factor 2γ is reduced. In this respect, k defines the sensitivity of the biquad to variations in the component values. The larger the k the tighter tolerances on the component values must be envisaged. A large positive feedback is known to make a network unstable. The biquad remains stable if k lies in the range from 0 to 2γ.

If we re-write the transfer function in the following general form

$$T(p) = \frac{A_0}{1 + ap + bp^2} \tag{6.1.23}$$

the pole-pair quality factor can be found from the following expression

$$Q = \frac{b}{a} \tag{6.1.24}$$

The equations (6.1.23) and (6.1.24) suggest that filters with real poles only will have Q-factors less than 0.5. They are always stable by definition.

It is obvious, that there are numerous circuit realisations of the biquad with a block diagram presented in Figure 6.23. An example of a low-pass all-pole biquad based on a single operational amplifier is shown in Figure 6.24. Its transfer function can be written in the following form:

$$T(p) = k\frac{\omega_p^2}{p^2 + \psi p + \omega_p^2} \tag{6.1.25}$$

where $\Psi = \frac{\omega_p}{Q_p}$, amplification factor $k = R1\frac{R2}{R1+R2}\left(1 + \frac{R6}{R5}\right)$, the characteristic normalisation frequency of the passband edge

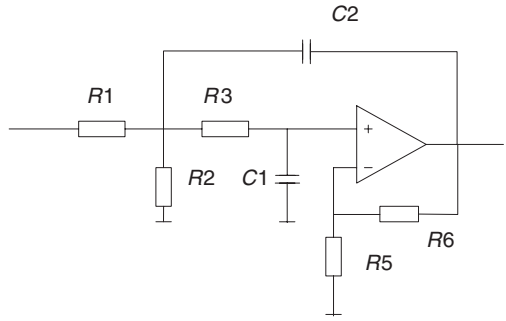


Figure 6.24 An example of a low-pass all-pole biquad based on a single operational amplifier

$$\omega_{\mathrm{p}}^2 = \frac{1}{(R1||R2)C1R3C2} \tag{6.1.26}$$

the quality factor

$$Q_{\mathrm{p}} = \frac{\sqrt{\frac{R3C1}{(R1||R2)C4}}}{1 + \frac{R3}{R1||R2} - \frac{R6C1}{R5C4}} \tag{6.1.27}$$

and, finally,

$$R1||R2 = \frac{R1 \cdot R2}{R1 + R2}$$

Figure 6.25 illustrates a special case of a normalised biquad with the unity gain and the lowest sensitivity. Following [4], its Laplace transfer function is given by

$$T(p) = \frac{1}{C1C2p^2 + 2C2p + 1} \tag{6.1.28}$$

In terms of pole locations on the p-plane, it takes the following form

$$T(p) = \frac{1}{\frac{1}{\alpha^2+\beta^2}p^2 + \frac{2\alpha}{\alpha^2+\beta^2}p + 1} \tag{6.1.29}$$

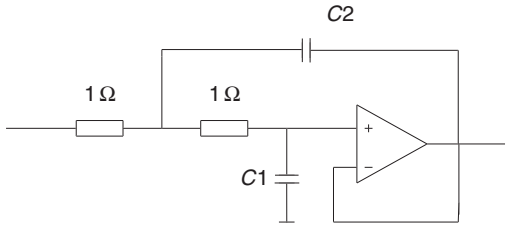


Figure 6.25 A special case of a normalised biquad with the unity gain and the lowest sensitivity. In this configuration, a fine adjustment of the pole position on the p-plane is achieved by selecting appropriate capacitors

Equating coefficients of p^2 and p in the equations (6.1.28) and (6.1.29), we derive values of capacitances expressed directly via pole co-ordinates:

$$C1 = \frac{1}{\alpha} \tag{6.1.30}$$

$$C2 = \frac{\alpha}{\alpha^2 + \beta^2} \tag{6.1.31}$$

where α and β are the real and imaginary co-ordinates of the biquad pole pair on the p-plane.

Figure 6.26 shows another example of the two-pole filter unit, which features two equal capacitors. In this configuration, a fine adjustment of the pole position on the p-plane is achieved by tuning resistor values rather than by selecting appropriate capacitors as in the previous biquad (Figure 6.25). This solution is more convenient and cost-effective, but comes at a price of doubling the filter gain, $k = 2$, and as a result the sensitivity of the circuit to tolerances of component values. An increased gain also means that the operational amplifier may need to have at least a doubled gain-bandwidth product in some technically challenging applications to maintain designed frequency characteristics.

Resistor values in the circuit shown in Figure 6.26 can be derived from the following equations [4]:

$$R1 = \frac{1}{2\alpha' C} \tag{6.1.32}$$

$$R2 = \frac{2\alpha'}{C[(\alpha')^2 + (\beta')^2]} \tag{6.1.33}$$

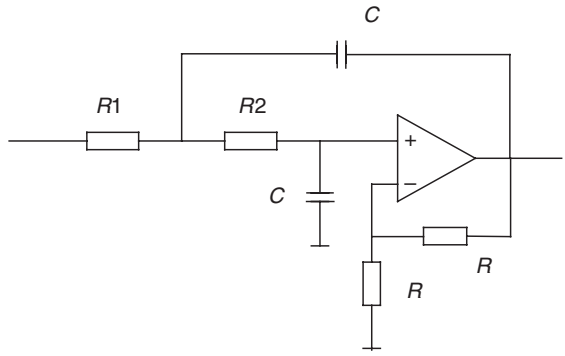


Figure 6.26 An example of the two-pole filter unit, which features two equal capacitors. In this configuration, a fine adjustment of the pole position on the p-plane is achieved by tuning resistor values rather than by selecting appropriate capacitors

where α' and β' are the actual (denormalised) real and imaginary pole co-ordinates, respectively.

Absolute values of resistors R lying within a reasonable range centred at approximately 1 kΩ do not affect the pole positions.

6.1.3.3 Three-pole filter section

An example of a normalised three-pole building block with the unity gain is shown in Figure 6.27. It is in fact a series connection of the passive low-pass filter (C2) and a unity gain low Q biquad. Using other modifications of the biquad will change the overall filter block gain and its sensitivity. In any case, the unit is expected to produce a single real pole and a pair of complex conjugate poles. Since two filters are not separated by a matching buffer, original pole positions of both filters are shifted due to the loading effect. According to [4], a transfer function of the normalised network is given by

$$T(p) = \frac{1}{p^3A + p^2B + pD + 1} \tag{6.1.34}$$

where $A = C1C2C3$, $B = 2C3(C1 + C2)$, $D = C2 + 3C3$.

6.1.3.4 Low-pass filter design

There are three types of all-pole low-pass filters commonly utilised in the precision instrumentation. These are:

1. Butterworth filters;
2. Chebyshev filters; and
3. Bessel filters.

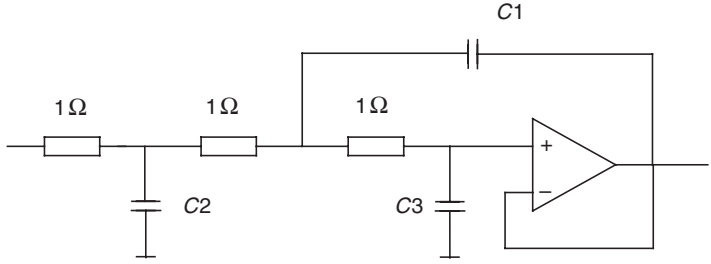


Figure 6.27 An example of a normalised three-pole building block with the unity gain. It includes the passive low-pass filter (C2) and a unity gain low Q-biquad. A type of the biquad defines the overall filter block gain and its sensitivity

In order to approach the attenuation approximation problem in each of these filters, we will use the network transfer function in a slightly modified form following the concept of Webster [5]. Assuming that $T(p) = \frac{N(p)}{D(p)}$ is the transfer function of a linear network, its absolute, value can be written as

$$|T(p)|^2_{p=j\omega} = T(p)T^*(p)|_{p=j\omega} = \frac{1}{1 + R(p)R^*(p)} \tag{6.1.35}$$

where $R(p)$ is the characteristic function of a network represented by the ratio of two polynomials as follows

$$R(p) = \frac{F(p)}{N(p)} \tag{6.1.36}$$

Here $F(p)$ is an arbitrary polynomial of the same order as $D(p)$, but without sharing common roots. The relationship between all three polynomials is analytically described by the well-known Feldtkeller equation

$$|D(p)|^2 = |F(p)|^2 + |N(p)|^2 \tag{6.1.37}$$

Butterworth filters

The Butterworth low-pass filters represent networks with the following characteristic function

$$R(p) = E\left(\frac{p}{\omega_\mathrm{p}}\right)^n \tag{6.1.38}$$

In this equation, ω_p again is responsible for the normalisation frequency of the passband edge. The Butterworth filter has maximally flat passband with the attenuation factor monotonically increasing with the frequency. The roll-off steepness of the frequency response between ω_p and a certain specified angular frequency $\omega_s(\omega_s > \omega_p)$ is essentially defined by the filter order n in the following form

$$n \geq \frac{\log(L)}{2\log\left(\frac{\omega_\mathrm{p}}{\omega_\mathrm{s}}\right)} \tag{6.1.39}$$

where $L = \frac{10^{0.1Apass}-1}{10^{0.1Astop}-1}$; $E^2 = 10^{0.1Apass} - 1$.

Here Apass corresponds to the maximum loss in the passband. In the Butterworth filter, the maximum loss is observed at an angular frequency $\omega = \omega_p$ (this is not necessarily true in other types of filters). Astop is the maximum attenuation factor in the stopband, that is, from ω_s to the infinity.

According to the equation (6.1.35), the pole positions of the Butterworth filter can be computed analytically by zeroing the following expression

$$1 + |R(p)|^2 = 1 + (-1)_n \, E^2 \left(\frac{p}{\omega_\mathrm{p}}\right)^{2n} = 0 \tag{6.1.40}$$

A solution of the equation for each individual k-pole can be presented in the following form [5]:

$$\left(\frac{p_\mathrm{k}}{\omega_\mathrm{p}}\right)^{2n} = \frac{(-1)^{n+1}}{E^2} = \frac{\mathrm{e}^{j\pi(n+1+2k)}}{E^2} \tag{6.1.41}$$

Assuming $E = 1$, we finally obtain all pole positions on the p-plane

$$\frac{p_\mathrm{k}}{\omega_\mathrm{p}} = \mathrm{e}^{\frac{j\pi(n=1+2k)}{2n}} \tag{6.1.42}$$

The equation (6.1.42) shows that all normalised poles of the Butterworth filter lie on the unit circle. The real and imaginary parts of pole positions can be expressed in an explicit form as follows

$$\frac{p_\mathrm{k}}{\omega_\mathrm{p}} = -\sin\left[\frac{\pi(n+1+2k)}{2n}\right] + j\cos\left[\frac{\pi(n+1+2k)}{2n}\right] \tag{6.1.43}$$

Therefore, all poles on the circle must be separated from each other by the same angle $\frac{\pi}{n}$. Poles with the maximum Q-value are at an angle of $\frac{\pi}{2n}$ from the imaginary axis $j\omega$. Figure 6.28 illustrates the pole positions for the Butterworth filters of the third and fourth orders. An odd-order filter includes necessarily at least one first-order filter unit with a real pole. Therefore, odd-order Butterworth filters always contain one pole situated on the negative real axis at -1. For an even order n, all poles are complex conjugates and two poles with the lowest Q-factor are at angles $\pm\frac{\pi}{n}$ to the negative real axis, as shown in Figure 6.28(b).

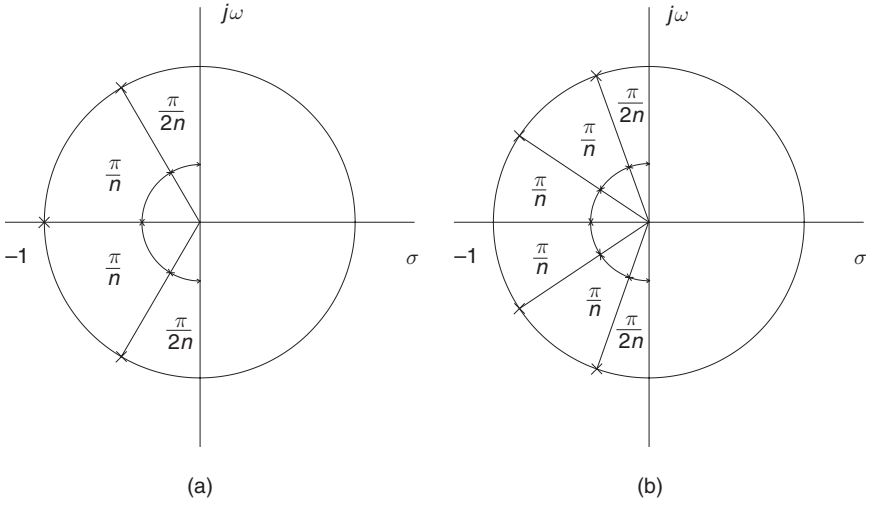


Figure 6.28 Pole positions for the Butterworth filters of the third (a) and fourth (b) orders. Odd-order Butterworth filters always contain one pole situated on the negative real axis at -1. For an even-order n, all poles are complex conjugate and two poles with the lowest Q-factor are at angles $\pm\frac{\pi}{n}$ to the negative real axis

6.1.3.5 Chebyshev low-pass filters

The Chebyshev low-pass filters represent linear networks with the characteristic function given by

$$R(p) = EG_n\left(\frac{p}{\omega_p}\right) \tag{6.1.44}$$

where G_n is a polynomial that is ranging between ± 1 in the passband. It is given by

$$G_n\left(\frac{p}{\omega_p}\right) = \cosh\left[n\cosh^{-1}\left(\frac{p}{\omega_p}\right)\right] \tag{6.1.45}$$

A factor E denotes the tolerance margin that determines the maximum attenuation in the passband, Apass, in the following form

$$Apass = 10\log_{10}(1 + E^2) \tag{6.1.46}$$

Note that the frequency response of the filter is no longer maximally flat. The maximum attenuation Apass can be reached several times over the

passband causing the presence of ripples in the amplitude-to-frequency characteristics of a filter.

For a specified attenuation factor Astop at a frequency $\omega_s > \omega_p$, the required filter order can be found from the following equation [5]:

$$n \geq \frac{\cosh^{-1}(L^{-1})}{\cosh^{-1}\left(\frac{\omega_s}{\omega_p}\right)} \tag{6.1.47}$$

where $L = \frac{10^{0.1Apass}-1}{10^{0.1Astop}-1}$.

The pole positions on the p-plane can be found in an analytical form solving the equation involving the characteristic function, that is,

$$1 + |R(p)|^2 = 0 \tag{6.1.48}$$

Taking into account the equations (6.1.44) and (6.1.45), we derive

$$\cosh\left[n\cosh^{-1}\left(\frac{p}{\omega_p}\right)\right] = \pm\frac{j}{E} \tag{6.1.49}$$

which yields the pole co-ordinates [6]

$$\frac{p_k}{\omega_p} = \cosh\left\{\left(\frac{1}{n}\right)\left[\sinh^{-1}\left(\frac{1}{E}\right) + j\pi\frac{1+2k}{2}\right]\right\} \tag{6.1.50}$$

or with the explicit real and imaginary parts:

$$\begin{aligned}\frac{p_k}{\omega_p} = &-\sin\left[\frac{\pi(1+2k)}{2n}\right]\cosh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{E}\right)\right] \\ &+ j\cos\left[\frac{\pi(1+2k)}{2n}\right]\sinh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{E}\right)\right]\end{aligned} \tag{6.1.51}$$

Properties of the Chebyshev filter are such that in the passband from 0 to ω_p it has an equiripple behaviour between the limits 1 and $\frac{1}{1+E^2}$. Above the angular frequency ω_p the transfer function falls monotonically to zero. Variations in E generate a family of frequency responses for a filter of the same order n. If we admit larger ripples in the passband (a larger E) the roll-off of the frequency characteristics between the passband and the stopband becomes steeper.

The comparison of a Butterworth filter and a Chebyshev filter with the same 3-dB cut-off frequency ω_c yields that the latter has always

a better frequency selectivity (or, in other words, a larger gradient of the frequency response above the end of the passband). A closer examination of the equations (6.1.43) and (6.1.51) indicates that the poles of the Chebyshev filter can also be obtained by multiplying the real and the imaginary parts of the Butterworth pole co-ordinates by $\sinh\left(\frac{1}{n}\sinh^{-1}\frac{1}{E}\right)$ and $\cosh\left(\frac{1}{n}\sinh^{-1}\frac{1}{E}\right)$ respectively. As a result, all poles of the Chebyshev filter lie on an ellipse rather than a circle, which intercepts the real axis at $\pm\sinh\left(\frac{1}{n}\sin^{-1}\frac{1}{E}\right)$ and the imaginary axis at $\pm\cosh\left(\frac{1}{n}\sin^{-1}\frac{1}{E}\right)$.

Following [7], we will normalise the pole positions of the Chebyshev filter in such a way that the pole ellipse intercepts the j-axis at $\omega = \omega_p$ using a frequency scaling factor $\gamma = \sec h\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{E}\right)\right]$. The normalised pole positions can now be expressed in the following form:

$$\frac{p_k}{\omega_p} = -\sin\left[\frac{\pi(1+2k)}{2n}\right]\tanh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{E}\right)\right] + j\cos\left[\frac{\pi(1+2k)}{2n}\right] \tag{6.1.52}$$

From a comparative analysis of the equations (6.1.43) and (6.1.51), we can conclude that the normalised Chebyshev pole co-ordinates follow from the Butterworth pole co-ordinates by multiplying their real parts by a factor $\tanh\left[\frac{1}{n}\sinh^{-1}\left(\frac{1}{E}\right)\right]$ and leaving their imaginary parts unchanged. Thus, the Butterworth filter can be considered as a special case of the normalised Chebyshev filter with the maximum flatness (no ripples) in the passband that is achieved at $E = 0$. As the attenuation tolerance factor E increases from zero onwards, all poles of the normalised Chebyshev filter move parallel to the real axis towards the imaginary $j\omega$-axis, thereby increasing their Q-factor. The phenomenon is illustrated in Figure 6.29.

The design of the normalised Chebyshev filter is straightforward using the grapho-analytical method involving the plot in Figure 6.29. The following denormalisation is achieved by performing reverse frequency scaling or, in other words, simply by substituting a product $\gamma \cdot p$ instead of p.

6.1.3.6 Bessel low-pass filters (case of maximally flat delay)

An all-pole low-pass Bessel filter is a network with a transfer function given by the ratio:

$$T(p) = \frac{B_n(0)}{B_n(p)} \tag{6.1.53}$$

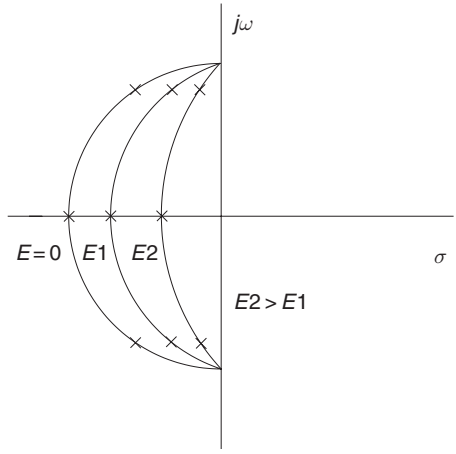


Figure 6.29 Pole positions of the normalised third-order Chebyshev filter. Poles move towards the imaginary axis as the ripples in the passband increase

where $B_n(p)$ is the Bessel polynomial defined by the recursion formula

$$B_n(p) = (2n-1)B_{n-1}(p) + p^2 B_{n-2}(p) = b_0 + b_1 p + b_2 p^2 + \cdots + b_n p^n \tag{6.1.54}$$

with an initial term $B_0(p) = 1$. Table 6.3 presents a sequence of the Bessel polynomials up to the fifth order. More complete tables of the polynomials can be found in many texts [8].

Whereas the Chebyshev and Butterworth filters are designed on the basis of tuning their amplitude–frequency characteristics, the synthesis of the Bessel filters primarily optimises their phase–frequency response.

Table 6.2 Properties of the Laplace transform

Function $f(t)$	Laplace transform	Fourier transform
$f_1(t) + f_2(t)$	$L_1(p) + L_2(p)$	$F_1(j\omega) + F_2(j\omega)$
$af(t)$	$aL(p)$	$aF(j\omega)$
$\int_0^t f(t)\mathrm{d}t$	$\frac{L(p)}{p}$	$\frac{F(j\omega)}{j\omega}$
$\frac{\mathrm{d}f(t)}{\mathrm{d}t}$	$pL(p)$	$j\omega F(j\omega)$
$f_1(t)^* f_2(t)$	$L_1(p)L_2(p)$	$F_1(j\omega)F_2(j\omega)$
$f(t - t_0)$	$L(p)\mathrm{e}^{-pt_0}$	$F(j\omega)\mathrm{e}^{-j\omega t_0}$
$f(at)$	$\frac{1}{a}L(\frac{p}{a})$	$\frac{1}{a}F(\frac{j\omega}{a})$

Table 6.3 Bessel polynomials up to the 5th order

$B_1(p) = 1 + p;$
$B_2(p) = p^2 + 3p + 1;$
$B_3(p) = p^3 + 6p^2 + 15p + 15;$
$B_4(p) = p^4 + 10p^3 + 45p^2 + 105p + 105;$
$B_5(p) = p^5 + 15p^4 + 105p^3 + 420p^2 + 945p + 945 \cdots$

An ultimate goal of the design procedure is to provide no overshoots and ringing in the step response of a network.

It is possible to show that an ideal square wave response is achieved by a network having a frequency-independent group delay, that is, with a phase shift proportional to frequency. If ϕ is the phase shift of a network, the definition for the group delay takes the following form

$$t_g = -\frac{d\phi}{d\omega} \tag{6.1.55}$$

In order to compare the quality of the time response of different filters, designers operate with a normalised group delay, which can be derived from

$$t_{gn} = \frac{t_g \omega_p}{2\pi} \tag{6.1.56}$$

The equation (6.1.56) may contain other more convenient characteristic frequencies instead of the passband edge frequency ω_p, for instance, the cut-off angular frequency at an attenuation point of -3 dB, and so on.

Unlike the case with the Butterworth and Chebyshev filters, there is no simple analytical expression for the pole locations of the Bessel filters. If, however, they are designed with an aim to provide the maximally flat group delay at $\omega = 0$ and the following monotone decay, we can use a reasonably accurate approximation to the Bessel transfer function (6.1.53) given by [4]

$$T(p) = \frac{1}{\sinh p + \cosh s} \tag{6.1.57}$$

The Bessel family of polynomials results from a continued fraction expansion approximating the hyperbolic functions $[\sinh(p) + \cosh(p)]$. According to [4], poles of the normalised Bessel filter are approximately located on a circle separated by $\frac{2}{n}$ fraction of the diameter in their imaginary parts, as shown in Figure 6.30.

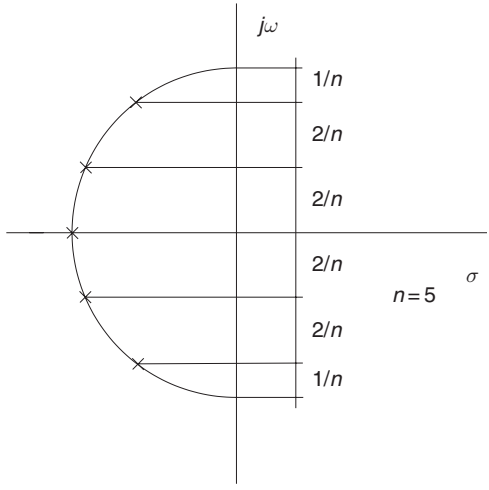


Figure 6.30 Pole positions of the normalised fifth-order Bessel filter designed with an aim to provide the maximally flat group delay at $\omega = 0$ and the following monotone decay. Poles are approximately located on a circle separated by $\frac{2}{n}$ fraction of the diameter in their imaginary parts

In order to compare the overall performance of the Butterworth, Chebyshev and Bessel filters, Figure 6.31 summarises the frequency characteristics and step time responses of all of them for $n = 4$. The design criteria of filters were the following:

- the Butterworth filter should have the maximally flat amplitude-to-frequency characteristics in the passband,
- the passband ripples of the Chebyshev filter could not exceed 0.1 dB; and
- the Bessel filter should have the maximally flat group delay.

All filters were normalised to have the same -3-dB cut-off frequency.

Comparative analysis of plots in Figure 6.31 suggests that the Butterworth and Chebyshev filters are desirable from the point of view of the amplitude-to-frequency characteristics. They provide an excellent passband selectivity rejecting efficiently the high-frequency noise. However, poorer performance in terms of the group delay results in overshoots and ringing in their step responses. The Bessel filters have the opposite behaviour. The phase linearity helps to achieve a faster rise and settling time. The step response shows virtually no overshoots or ringing whereas the impulse response lacks any oscillatory behaviour. Yet, the

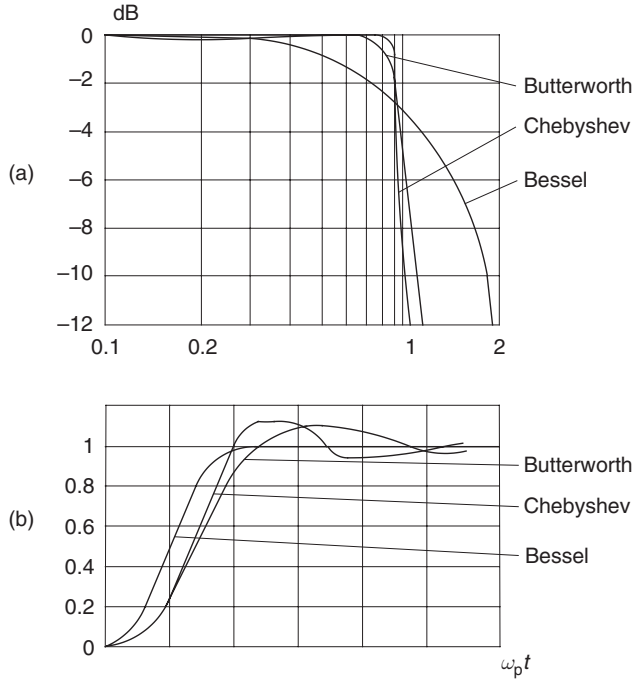


Figure 6.31 The frequency characteristics and step and time responses of the Butterworth, Chebyshev and Bessel filters for $n = 4$

roll-off of the amplitude-to-frequency characteristics is the least steep compared to that of other types of filters indicating their poor selectivity with respect to high-frequency noise.

It is obvious that electronic designers would attempt to synthesise filters that combine the advance features of the Butterworth–Chebyshev filters, on the one hand, and those of the Bessel filters, on the other hand. If a new network is constrained to the all-pole concept, these attempts end up with some compromised performance in the frequency and time responses. The Butterworth–Thomson filter is a good illustration to the statement. The poles of the filter are situated on the unity circle somewhere between positions assigned by the Butterworth and Bessel filters [8].

Filters with transfer functions containing polynomials in their numerators, $N(p)$, (having zeroes as well as poles) can, in principle, approach closer in their behaviour the ideal low-pass filter. Such filters, however, are more complex to design and more expensive to manufacture. We will briefly discuss the synthesis of this type of networks in the next paragraph when we introduce the Nyquist and Bode analysis

methods. This section will be concluded with remarks on the step response of the all-pole filters.

Once a frequency response of a hybrid non-standard filter has been approximated in the form

$$T(p) = \frac{T_0}{\prod_{i=1}^{N}(p - \upsilon_i)} \tag{6.1.58}$$

the Laplace transform of the step response is $\frac{T(p)}{p}$. Following [5], before trying to take the inverse Laplace transform it is useful to expand the function $\frac{T(p)}{p}$ into partial fraction form:

$$\frac{T(p)}{p} = \sum_{i=1}^{n+1} \frac{A_i}{p - \upsilon_i} \tag{6.1.59}$$

where $A_i = T_0 \frac{1}{\prod_{k=1, \neq i}^{N+1} (\upsilon_i - \upsilon_k)}$ is the residue at the pole υ_i.

Now, the step time response can be expressed as follows:

$$a(t) = \sum_{i=1}^{N} A_i e^{\upsilon_i t} \tag{6.1.60}$$

Some poles υ_i, as we saw earlier, are complex. In all real stable filters, they tend to appear in complex conjugate pairs. Thus, corresponding residues will also be complex conjugate yielding a real time–step response $a(t)$.

6.1.4 Graphical Methods of Analysis and Synthesis in the Frequency Domain

The purpose of the analysis of an electronic network is to examine whether it meets requirement specifications on the stability, the transient performance and the steady state error. The majority of designed systems (transducers, amplifiers, filters, etc.) are expected to be absolutely stable. We have already learned in the section 6.1.3 that poles of an absolutely stable all-pole filter must be situated in the left-hand p-plane. There are several analytical methods (for instance, by Routh, Hurwitz,

etc.) that indeed study locations of the poles and zeroes as a function of various impetuses. Practicing engineers, however, rarely use these methods due to their complexity and the necessity to deal with lengthy analytical expressions of transfer functions. The time domain methods are also rather inconvenient as, for instance, the pole positioning cannot be directly verified by a non-complicated experiment.

The analysis, obviously, does not stop on mere registering a fact of the stability of operation. In our design and calculations, we manipulate with the nominal values of components. These values have tolerance margins specified by manufactures. Moreover, they may also vary with temperature, time and changing operating conditions. Thus, a designer needs additional information on the relative stability or, in other words, the stability margin. It is vital to determine how close a system is to being unstable in the worst-case scenario. In what follows, we will show that the frequency domain representation is a much preferred option in doing that.

There is a number of grapho-analytical frequency domain methods available for the electronic network analysis and design. In this paragraph, we will concentrate on the Nyquist diagram and the Bode plots. Both methods are based on the graphical exploration of the open loop gain as a function of frequency. They are applicable for both continuous and discrete time systems. The latter will be briefly discussed in the next section.

6.1.4.1 Nyquist diagrams

The Nyquist diagram offers undoubtedly the best way to introduce the stability criteria and formulate the definition for the measure of the relative stability. Similar to complex poles and/or zeroes on the p-plane, the complex transfer function $T(p)$ can be visualised on a 2D graph with axes $\mathrm{Im}[T(p)]$ vs $\mathrm{Re}[T(p)]$. In order to be closer to experimental results obtained with a sinusoidal stimulus, we need to operate in the frequency domain. Therefore, in our further discussions, we will tend to use the Fourier transform rather than the Laplace transform. There is a little difference between these two techniques. In fact, the Fourier transform is a special case of the Laplace counterpart where the complex variable p is restricted to lay on the imaginary axis. In its classical form, it does not operate with signals of the infinite duration, for instance, the step function (Table 6.1). However, there is no need to deal with functions for which the transform does not exist along the imaginary axis only. Such functions would not characterise the networks that can be realised in practice. Following that, the Fourier transform can be derived simply by substituting $j\omega$ for p in the Laplace expressions.

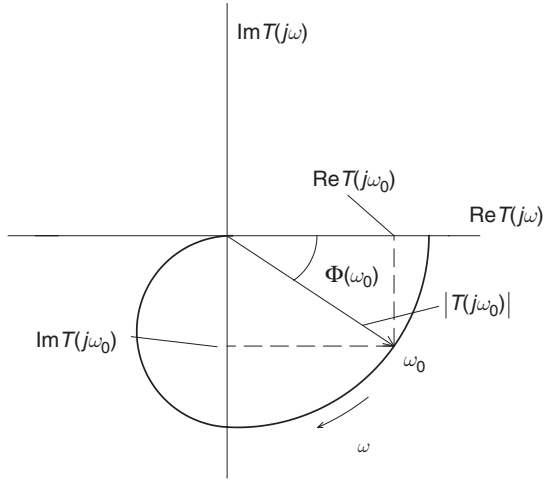


Figure 6.32 A generic example of the Nyquist polar plot illustrating the transfer function $T(\mathrm{p})$

The polar form of Fourier transfer function is given by

$$T(j\omega) = |T(j\omega)|_{\mathrm{e}}^{j\phi(\omega)} \tag{6.1.61}$$

In the complex form it is

$$T(j\omega) = Re[T(j\omega)] + j\mathrm{Im}[T(j\omega)] \tag{6.1.62}$$

A Nyquist polar plot is a 2D graph with axes $\mathrm{Im}[T(j\omega)]$, $\mathrm{Re}[T(j\omega)]$ which is mapped by varying the angular frequency ω. In principle, ω can vary from zero to infinity. However, the polar plot can contain only points from a frequency range of interest. A generic example of the Nyquist polar plot is shown in Figure 6.32. It illustrates the equations (6.1.61) and (6.1.62) at $\omega = \omega_0$.

6.1.4.2 The Nyquist criterion for the stability of a feedback amplifier

In practice, we normally deal with networks or systems whose gain falls off and the phase shift increases monotonically with frequency. From a common sense, if such a network or system is to be stable, then the loop gain must fall off to less than the unity before the feedback turns from negative to positive one, or, in other words, the phase shift reaches $-\pi(-180^\circ)$.

Let us consider a linear amplifier with a single feedback loop. Its closed loop gain is given by

$$K(j\omega) = \frac{A(j\omega)}{1 + \beta A(j\omega)} \tag{6.1.63}$$

where $A(j\omega)$ is the gain of the bare amplifier, β is the transfer function of the feedback network, $L(j\omega) = \beta A(j\omega)$ is the loop gain. The equation (6.1.63) has a singularity at a point $L(j\omega) = [-1,0]$ which is called the critical point of the Nyquist diagram.

The Nyquist stability criterion states that *the amplifier will be stable if the polar plot* L(jω) *and its conjugate do not encircle the critical point*. Figure 6.33 shows an example of Nyquist diagram of a system. The unit circle represents a loop gain 1. The polar plot does not encircle the critical point thereby meeting the Nyquist criterion of the stable system. An angular frequency corresponding to a point where the plot crosses the unit circle, ω_1, is called the unity gain angular frequency. At this frequency, we evaluate the phase margin of a network, which is

$$\phi_m = \pi + \arg[K(j\omega)] = \pi - \phi(\omega_1) \tag{6.1.64}$$

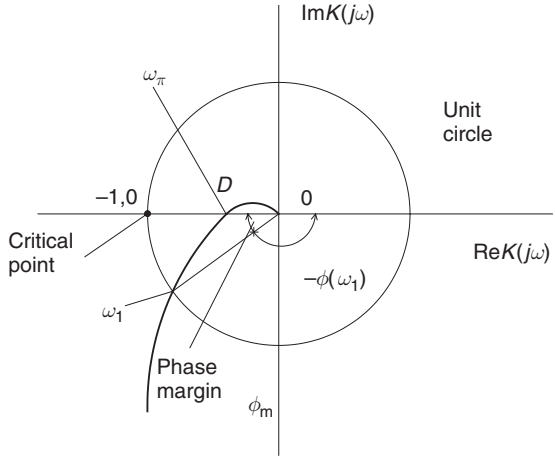


Figure 6.33 An example of Nyquist diagram of a system. The unit circle represents a loop gain 1. The polar plot does not encircle the critical point (−1, 0), thereby meeting the Nyquist criterion of the stable system

The definition of the gain margin has the following form:

$$G_{\mathrm{m}} = \frac{1}{|K(j\omega_{\pi})|} \tag{6.1.65}$$

where $K(\omega_{\pi})$ is the open loop gain at the crossover angular frequency ω_{π}. The phase shift of the open loop gain at ω_{π} is $\arg[K(\omega_{\pi})] = -\pi$.

It is convenient to express the gain margin in the logarithmic form, as follows

$$\log_{10} G_{\mathrm{m}} = -20 \log_{10} |K(\omega_{\pi})| \tag{6.1.66}$$

This corresponds to the vector length

$$\log_{10} G_{\mathrm{m}} = -20 \log_{10} |OD| \tag{6.1.67}$$

on the Nyquist diagram shown in Figure 6.33.

A polar plot depicted in Figure 6.34 does not meet the Nyquist criterion. It encircles the critical point, which is characteristic for the *unstable* system.

In terms of the stability there is a third kind of systems. Figure 6.35 presents an example of the Nyquist diagram of such a system. According to the diagram, the system is likely to become unstable if the loop gain decreases. A network that is stable, but becomes unstable if the (loop) gain decreases is said to be *conditionally stable*.

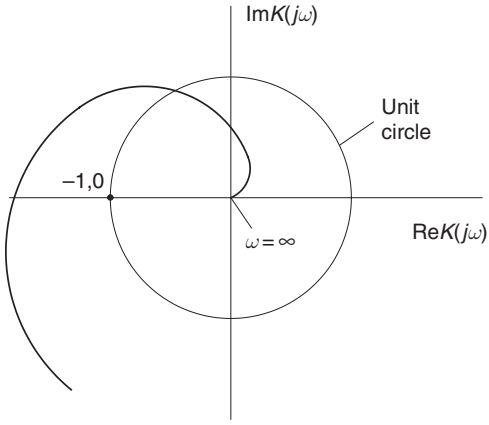


Figure 6.34 A polar plot of a transfer function that does not meet the Nyquist criterion. It encircles the critical point $(-1, 0)$, which is characteristic for the *unstable* system

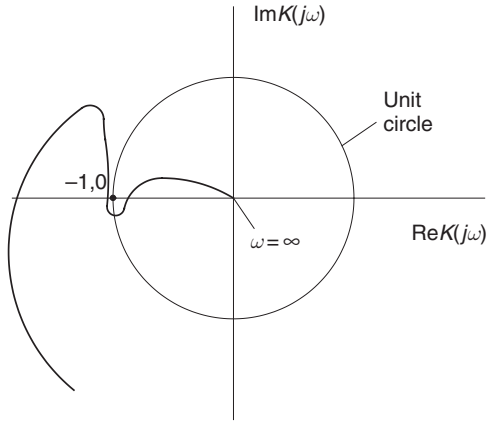


Figure 6.35 An example of the Nyquist diagram of conditionally stable system. According to the diagram, the system is likely to become unstable if the loop gain decreases

Before turning to the Bode analysis and design, we will make some remarks on the experimental procedure for obtaining the loop gain as a function of frequency. The procedure is illustrated by Figure 6.36 that depicts a SQUID amplifier with a closed feedback. In order to measure the loop gain, we break the feedback loop at some point, normally adjacent to the system output. The test signal is applied to the left terminal. The response is the signal returned to the other side of the

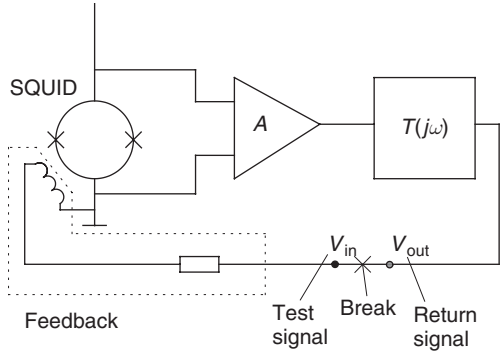


Figure 6.36 A SQUID amplifier with the broken feedback loop illustrating the experimental procedure for obtaining the loop gain as a function of frequency. The test signal is applied to the left terminal. The response is the signal returned to the other side of the break

break. Applying the sinusoidal test signal at different angular frequencies ω_i it is possible to restore the loop gain as

$$L(j\omega_i) = \left| \frac{V_{\text{out}}(\omega_i)}{V_{\text{in}}(\omega_i)} \right| e^{j\phi(\omega_i)} \tag{6.1.68}$$

where $\phi(\omega_i)$ is the phase shift between sinusoidal signals V_{in} and V_{out} at angular frequencies ω_i whereas $V_{\text{in}}(\omega_i)$ and $V_{\text{out}}(\omega_i)$ are the magnitudes of the sinusoidal signals.

6.1.4.3 Bode analysis and design

The Bode method is, in fact, very similar to the Nyquist method in that both techniques analyse the open loop gain of a network under study in the frequency domain. However, the Bode method employs a slightly different graphical representation consisting of two plots:

1. the absolute value of the loop gain vs frequency, $|L(j\omega)|$; and
2. the phase shift of the loop gain vs frequency, $\phi(j\omega)$.

Bode diagram is perhaps the most natural way to visualise the behaviour of a network in the frequency domain. This may explain the fact that the plots are most widely utilised not just by practising electronic engineers but also those who never heard about Bode theorems.

Bode formulated his two theorems for systems including only linear minimum phase blocks. The theorems essentially point at the fact that there is an unambiguous relationship between the loop gain variations as a function of frequency and the frequency dependence of the phase shift. According to the first theorem, the phase shift of a system in rad at an angular frequency ω_{d} of interest can be found from the following equation:

$$\phi(\omega_{\text{d}}) = \frac{\pi}{2} \left| \frac{\text{d}G}{\text{d}n} \right|_{\text{d}} + \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\left| \frac{\text{d}G}{\text{d}n} \right| - \left| \frac{\text{d}G}{\text{d}n} \right|_{\text{d}} \right] \ln \cot g \left| \frac{n}{2} \right| \text{d}n \tag{6.1.69}$$

where G represents the loop gain in nepers (1 neper $= \ln(e)$), $n = \ln(\frac{\omega}{\omega_{\text{d}}})$, $\left| \frac{\text{d}G}{\text{d}n} \right|$ is the slope of the amplitude-log-frequency curve in nepers per unit change of n (1 neper/unit change of n is equivalent to 20 dB/decade), $\left| \frac{\text{d}G}{\text{d}n} \right|$ is the slope of the amplitude-log-frequency curve at ω_{d}, $\ln \coth \left| \frac{n}{2} \right|$ is the Bode weighting function.

According to the equation (6.1.69), a phase shift at a selected frequency ω_d is caused by the presence of slopes in the whole magnitude–frequency response. The weighting factor with which the slopes influence the $\phi(\omega_d)$ is not uniform over the whole frequency range of $L(j\omega)$. It becomes larger as the characteristic approaches the frequency ω_d. Therefore, the system stability is primarily governed by the frequency characteristic near the unity gain frequency ω_1.

The plot of the Bode weighting function determining the phase shift contributions at ω_d from gradients of $L(j\omega)$ at other frequencies is shown in Figure 6.37. It suggests that the integral in the equation (6.1.69) becomes relatively small compared to the first term at small and large n. In this case, the slope of the amplitude-log-frequency curve should be just under -2 nepers/unit of n (-40 dB/decade) in the vicinity of the unity frequency, ω_1, in order to ensure stability. Of course, the statement is true only if a system does not contain ranges with a very steep roll-off of the magnitude–frequency response that would overpower the weighting factor.

A mathematical expression of the second Bode theorem has the following form:

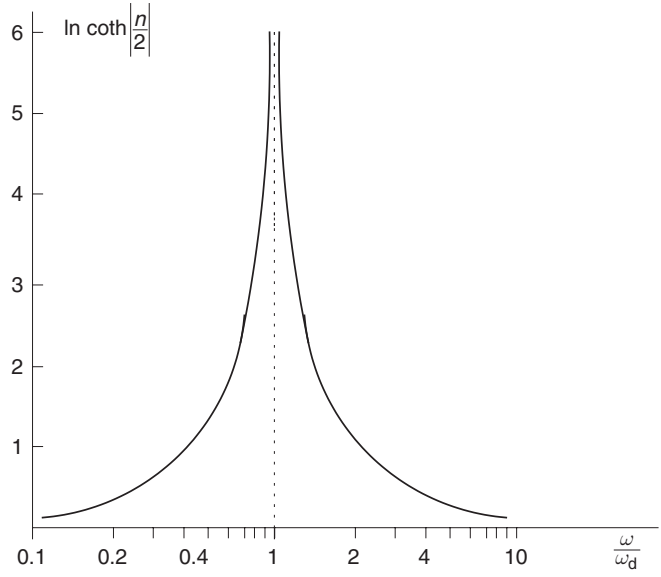


Figure 6.37 A plot of the Bode weighting function determining the phase shift contributions at ω_d from gradients of $L(j\omega)$ at other frequencies

$$\int_0^{\omega_s} \frac{G\mathrm{d}\omega}{\sqrt{\omega_s^2 - \omega^2(\omega)^2}} + \int_{\omega_s}^{\infty} \frac{\phi\mathrm{d}\omega}{\sqrt{\omega_s^2 - \omega^2(\omega)^2}} = R \tag{6.1.70}$$

where

$$R = \frac{\pi}{2} \frac{\phi(\omega_\mathrm{d})}{\omega_\mathrm{d}\sqrt{\omega_\mathrm{s}^2 - \omega_\mathrm{d}^2}} \quad \text{at} \quad \omega_\mathrm{d} < \omega_\mathrm{s} \tag{6.1.71}$$

$$R = -\frac{\pi}{2} \frac{G(\omega_\mathrm{d})}{\omega_\mathrm{d}\sqrt{\omega_\mathrm{d}^2 - \omega_\mathrm{s}^2}} \quad \text{at} \quad \omega_\mathrm{d} > \omega_\mathrm{s} \tag{6.1.72}$$

ω_s is the angular cross-over frequency.

The equation (6.1.70) suggests that we have the complete information about the frequency response of a system even in the case when the gain is specified only below ω_s and the phase is specified only above ω_s. This theorem, again, works adequately for linear minimum phase networks only. This property is very helpful at the system design stage. The matter is that, in practice, requirement specifications can often be elaborated for the amplitude or the phase solely in a certain range(s) of frequencies and the other quantity at other frequencies. The Bode theorems enable the restoration of the full frequency response (magnitude and phase) in the whole frequency range of interest from the partial information. The design of the SQUID amplifier with the maximised feedback depth discussed in Chapter 2 is a good illustration to this statement.

In what follows, we will re-formulate the Nyquist stability criterion and the relative stability factors using the Bode plots shown schematically (not to scale) in Figure 6.38. The gain margin is $\log G_m = \log\left(\frac{1}{|L(j\omega_\pi)|}\right) = -\log|L(j\omega_\pi)|$. It essentially represents a number of decibels below 0 dB at the phase cross-over frequency ω_π. Equally, the phase margin is a number of radians (or degrees) above $-\pi$ at the gain cross-over frequency ω_1 ($L(j\omega_1) = 1$). In most cases of systems involving linear minimum phase network units, the positive gain and phase margins are necessary (but not sufficient) conditions of a stable operation. Yet, the Nyquist polar plot remains to be a preferable graphical method in the verification of the absolute stability.

The loop amplification of some active networks including, say, operational amplifiers may well exceed several millions at low frequencies. The frequency range itself can extend from zero into the GHz region. In order to accommodate the whole magnitude–frequency range on the

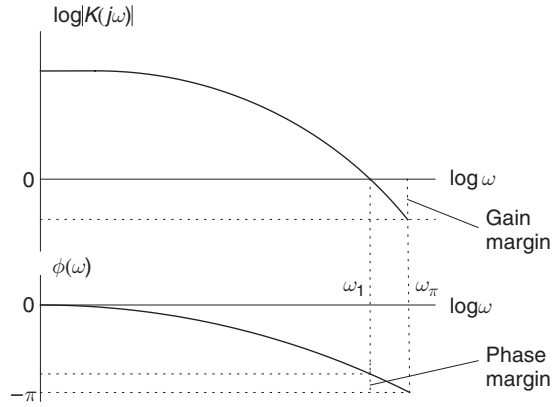


Figure 6.38 Schematic (not to scale) Bode plots illustrating the Nyquist stability criterion and the relative stability factors: gain and phase margins

same plot one has to utilise the logarithmic scales for both the magnitude and frequency axes. It became common to plot logarithmic magnitude axes in dB units defined as $\log L[\text{dB}] = 20 \log |L(j\omega)|$. All frequencies and magnitudes on the log–log scales are equally emphasised. Such a plot is much easier to read when we try to identify the positions of poles and zeroes, and evaluate their significance for the system stability through the weighting function.

The logarithmic representation adds some other convenient features to the analysis and design in the frequency domain. Since polynomials can be factored, the loop gain with which we deal in the frequency domain analysis can be viewed as a product of simple complex terms. When complex terms are multiplied together, the result can be presented as a product of amplitudes of each term with the total phase shift resulting from the algebraic sum of the individual phases. The product of amplitudes is known to be equivalent to addition of their logarithms. Therefore, the overall magnitude–frequency response on the log–log scale can be constructed simply by addition of known individual responses as shown in Figure 6.39.

The full Bode equations are normally used at the requirement capture as well as the final design stage when we need to know the exact magnitude and phase frequency curves. Such a precise analysis would be justified in cases when a system must carry particularly important functions (safety, security, etc.). Otherwise, practising electronic engineers operate with asymptotic approximations that are sufficient for deriving significant time constants, ball park gain and phase characteristics, approximate

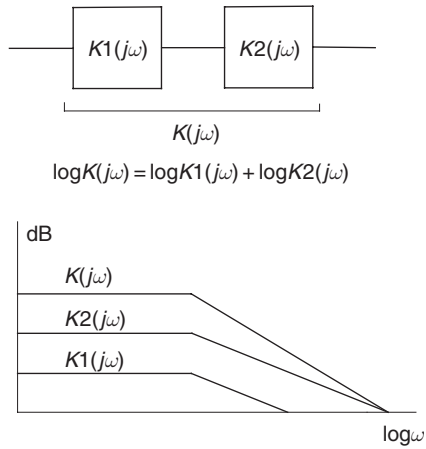


Figure 6.39 Since the product of amplitudes is known to be equivalent to addition of their logarithms the overall log–log magnitude–frequency response can be constructed by simply addition of known individual responses

transfer functions to predict the time response, and so on. For this, we initially concentrate upon the frequency response for small and large values n. According to the Bode theorems, the decibel gain and the phase are proportional to the log of the frequency. Thus, they are reduced to straight lines on the log–log scale, which can be drawn by simple inspection without the equations (6.1.69) and (6.1.70).

Exercise 6.7

In this exercise, we will draw asymptotic approximation of the magnitude–frequency response of a single-pole operational amplifier with the unity feedback gain, $\beta = 1$.

The closed loop gain of an operational amplifier is given by the following general equation

$$T(j\omega) = \frac{K(j\omega)}{1 + \beta(j\omega)K(j\omega)} \tag{6.1.73}$$

This can be re-written in the modified form

$$T(j\omega) = \frac{1}{\beta(j\omega)}\left[\frac{K(j\omega)\beta(j\omega)}{1 + K(j\omega)\beta(j\omega)}\right] \tag{6.1.74}$$

The equation can now be approximated by two simple equations

$$T(j\omega) \approx K(j\omega) \quad \text{for} \quad |K(j\omega)\beta(j\omega)| \ll 1 \tag{6.1.75}$$

and

$$T(j\omega) \approx \frac{1}{\beta(j\omega)} \quad \text{for} \quad |K(j\omega)\beta(j\omega)| \gg 1 \tag{6.1.76}$$

Now, it is straightforward to construct an asymptotic magnitude–frequency response of the operational amplifier with the closed feedback. The plot is shown in Figure 6.40.

Asymptotic plots intersect each other in a well-defined corner frequency, ω_1 corresponding to a 3-dB attenuation of the closed loop gain. Asymptotic plots suggest that the gain is $|K(j\omega_1)|\beta = 1$. Thus, at this frequency, the asymptotic approximation introduces the largest error. In general, the largest approximation error corresponds to $n = 1$. It can be evaluated either from experiment or analytically. In our case, the approximation error can be estimated from the equations

$$\left[\frac{K(j\omega)\beta}{1 + K(j\omega)\beta}\right] \quad \text{for} \quad |K(j\omega\beta)| > 1 \tag{6.1.77}$$

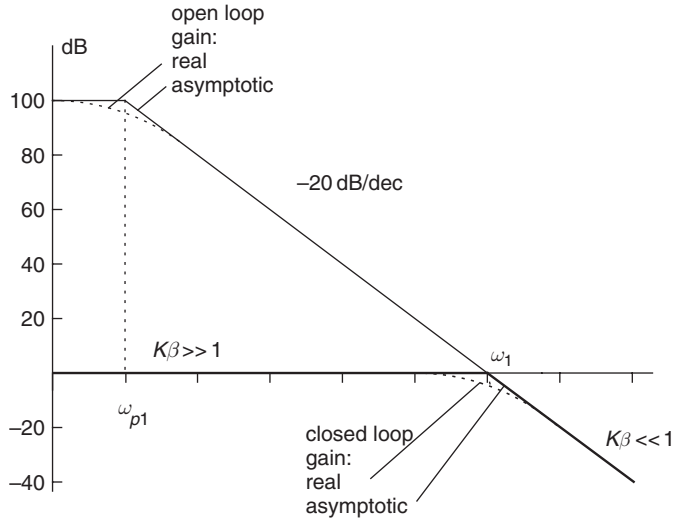


Figure 6.40 An asymptotic and actual magnitude–frequency response of an operational amplifier with the closed feedback on the log–log scale. The characteristic can be approximated by a number of straight lines. The largest error occurs at the point where the approximation lines intersect

and

$$\left[\frac{1}{1+K(j\omega)\beta}\right] \quad \text{for} \quad |K(j\omega)|\beta < 1 \tag{6.1.78}$$

that are derivatives of the equations (6.1.73–6.1.76).

We require a limited amount of basic functions to represent the factorised transfer function of a linear minimum phase network. In fact, implementation of only four of them would be sufficient to approximate the Bode plots of any complexity with the sufficient accuracy. These functions are:

1. frequency-independent constant factor, K_0;
2. a term containing one zero p or a pole $\frac{1}{p}$ positioned at the origin of the p-plane;
3. a term containing a first-order zero $(1 + ap)$ or a first-order pole $(1 + bp)^{-1}$;
4. a quadratic term containing a zero

$$\left(\frac{p}{\omega_p}\right)^2 + \frac{Q}{\omega_p}p + 1$$

or a pole

$$\frac{1}{\left(\frac{p}{\omega_p}\right)^2 + \frac{Q}{\omega_p}p + 1}.$$

In this respect, the synthesis of the Bode plots is quite similar to the design of linear filters. The coincidence is not accidental since the Bode plots are actively utilised in the filter design.

In what follows, we will give the Bode graphical representations of these functions.

Frequency-independent constant factor

The constant factor, K_0, represents the product of all frequency-independent terms in the loop amplification. On the Bode plot, it is a straight horizontal line $20 \log_{10} K_0$ with a zero phase shift as shown in Figure 6.41.

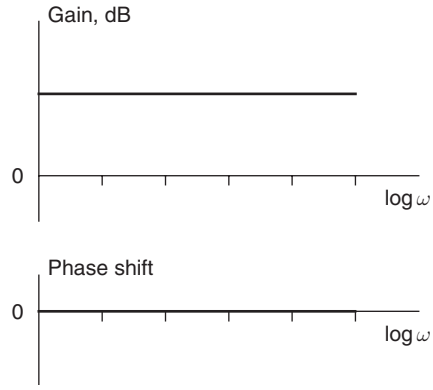


Figure 6.41 The Bode plot of the frequency-independent constant factor

Zero positioned at the origin of the p*-plane*
The logarithmic magnitude is given by

$$L_{\log}(\omega)[\text{dB}] = 20\log_{10}(|j\omega|) = 20\log_{10}(\omega) \tag{6.1.79}$$

whereas the phase is

$$\phi(\omega) = \arg[(j\omega)] = \frac{\pi}{2}\text{rad.} \tag{6.1.80}$$

If an expression of the transfer function contains a number m zeroes, the term $(j\omega)$ is raised to the power m. In the logarithmic form, this increases the slope of the magnitude–frequency response by a factor of m, that is,

$$L(\omega)[\text{dB}] = 20\log_{10}(|j\omega|^m) = 20m\log_{10}\omega \tag{6.1.81}$$

and

$$\phi(\omega) = \arg(j\omega)^m = \frac{\pi}{2}m\text{ rad} \tag{6.1.82}$$

This is illustrated in Figure 6.42.

Pole positioned at the origin of the p*-plane*
For a transfer function including m such poles the magnitude–frequency and the phase–frequency characteristics are given by

$$L(\omega)[\text{dB}] = 20\log_{10}(|j\omega|^{-m}) = -20m\log_{10}\omega \tag{6.1.83}$$

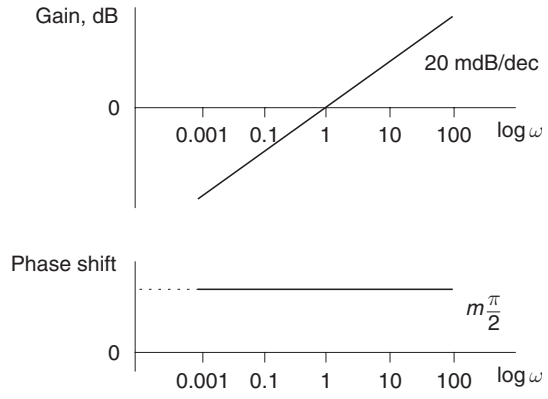


Figure 6.42 The Bode plot of a network with the zero positioned at the origin of the *p*-plane

and

$$\phi(\omega) = \arg(j\omega)^{-m} = -\frac{\pi}{2}m \text{ rad} \tag{6.1.84}$$

respectively. The Bode plots of such poles are depicted in Figure 6.43.

Simple zero (not at the origin of the p*-plane)*
The Fourier transfer function of the simple zero is given by

$$T(j\omega) = 1 + j\omega\tau \tag{6.1.85}$$

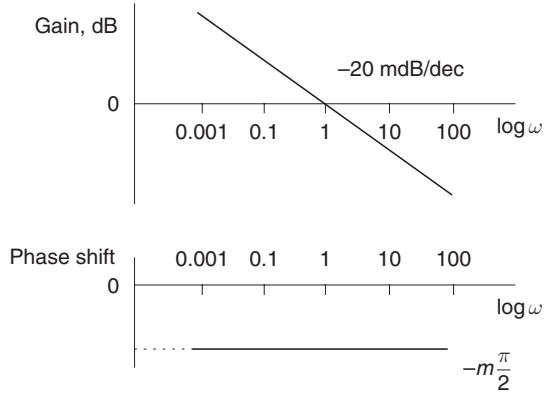


Figure 6.43 The Bode plot of a network with the pole positioned at the origin of the *p*-plane

The Bode plots of this transfer function is no longer a straight line. In order to be able to construct the frequency response by inspection, we have to apply the asymptotic approximation. To do that, we will calculate the logarithmic magnitude–frequency characteristics at $\omega\tau \ll 1$ as

$$L(\omega)[\text{dB}] = 20\log_{10}(|j\omega\tau + 1|) \approx 20m\log_{10} 1 = 0 \text{ dB} \qquad (6.1.86)$$

and at $\omega\tau \gg 1$ when we can neglect the constant term to get

$$L(\omega)[\text{dB}] = 20\log_{10}(|j\omega\tau + 1|) \approx 20m\log_{10}\omega\tau \qquad (6.1.87)$$

The equation (6.1.87) is similar to the expression describing behaviour of the zero at the origin with one exception that the log-magnitude curve intersects the 0-dB line at a frequency $\omega = \frac{1}{\tau}$ rather than at $\omega = 1$. Otherwise, at $\omega\tau \gg 1$ the slope and phase remain constant at 20 dB/decade and $\frac{\pi}{2}$ respectively. The point where two approximation lines intersect is called the corner frequency. The exact and approximated Bode plots are depicted in Figure 6.44.

The maximum error introduced by the asymptotic approximation occurs at the corner frequency. It equals to $20\log_{10}(|1 + j1|) = 20\log\sqrt{2} = 3$ dB. Knowing that enables us to restore the exact curve of a simple zero also by inspection from existing asymptotic lines.

The phase–frequency characteristics of the transfer function described by the equation (6.1.85) is derived from the ratio of the imaginary to the

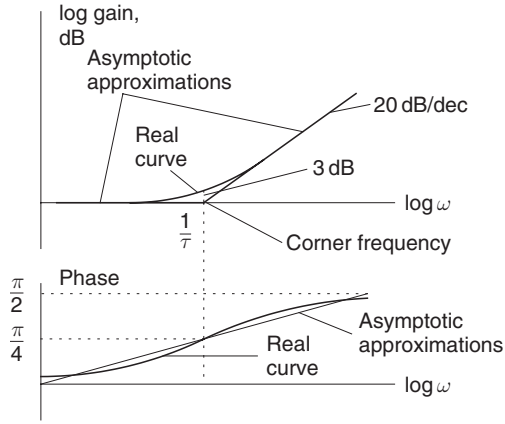


Figure 6.44 The Bode plot of a network with a simple zero (not at the origin of the *p*-plane)

real part of the complex function whose tangent gives us the sought angle. Following that we get

$$\phi(\omega) = \arg(1 + j\omega\tau) = \tan^{-1}\omega\tau \tag{6.1.88}$$

The zero shows a phase shift of $\frac{\pi}{4}$ at the corner frequency approaching asymptotically $\frac{\pi}{2}$ at frequencies $\omega\tau \gg 1$. The first order linear approximation of the phase curve would require minimum three lines. One way to arrange it is to use a straight line with a slope $\frac{\pi}{4}$/decade starting from a point corresponding to a one tenth fraction of the corner frequency, passing through a value of $\frac{\pi}{4}$ at the corner frequency and finishing at ten times the corner frequency with a value of $\frac{\pi}{2}$. At frequencies below $0.1\,\omega\tau$ and above $10\,\omega\tau$ one can use horizontal lines. Such an approximation produces a maximum error of $\frac{\pi}{20}$.

The presence of m simple zeroes in the transfer function gives m-fold increase in the slope of the approximation line at $\omega\tau \gg 1$ and m times larger phase shift compared to one produced by the single simple zero.

Simple pole (not at the origin of the p*-plane)*
The simple pole has the following transfer function

$$T(j\omega) = \frac{1}{1 + j\omega\tau} \tag{6.1.89}$$

The treatment of this problem is very similar to that of simple zero factors. Since the logarithm of a reciprocal quantity gains just an opposite sign we are getting Bode plots that are the mirror images of those shown in Figure 6.44. We will leave it to readers to derive asymptotic equations. For reference, results are summarised in Figure 6.45.

Quadratic poles
A general expression for the frequency response of the quadratic pole is given by

$$T(j\omega) = \frac{1}{1 - (\omega\tau)^2 + 2j\zeta\omega\tau} \tag{6.1.90}$$

where ζ is the damping ratio. Approximation curves follow from the equations:

$$L(\omega)[\text{dB}] \approx 20m\log_{10} 1 = 0\,\text{dB} \quad \text{at} \quad \omega\tau \ll 1 \tag{6.1.91}$$

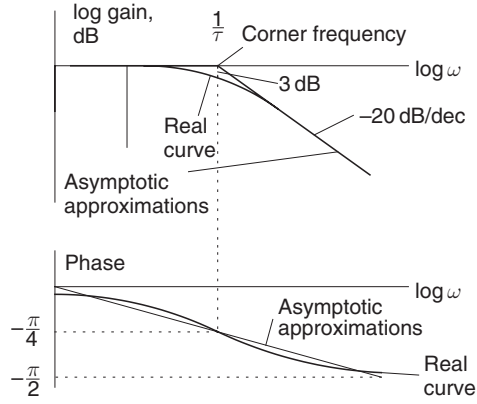


Figure 6.45 The Bode plot of a network with a simple pole (not at the origin of the p-plane)

and

$$L(\omega)[\text{dB}] = 20\log_{10}\left(\left|\frac{1}{(\omega\tau)^2}\right|\right) \approx -40m\log_{10}\omega\tau \text{ at } \omega\tau \gg 1 \quad (6.1.92)$$

where m is the number of quadratic poles in the transfer function. Asymptotic lines again intersect at the corner frequency $\omega\tau = 1$. Unlike previous cases, however, the approximation error depends strongly on the damping factor. These errors can be obtained from exact plots of $|T(j\omega)|$. Figure 6.46 demonstrates the Bode plots of the quadratic pole at $\zeta = 1$ and $\zeta = 0.1$. Note that the former case corresponds to the presence of the double simple pole in the transfer function.

We will omit the discussion for the case of the quadratic zero as self-explanatory.

6.1.4.4 Compensation networks

What should we do if our designed preamplifier exhibits a substantial overshoot in the step response or a SQUID current amplifier oscillates at high frequencies? A possible answer is: we need to ask a practising electronic engineer to resolve the problem. If we are practising electronic engineers or there is none around, then we need to look for solutions of both the problems using primarily the Nyquist diagrams and the Bode plots. The first problem is likely to be caused by the fact that the group delay of the network is not maximally flat, whereas the second problem

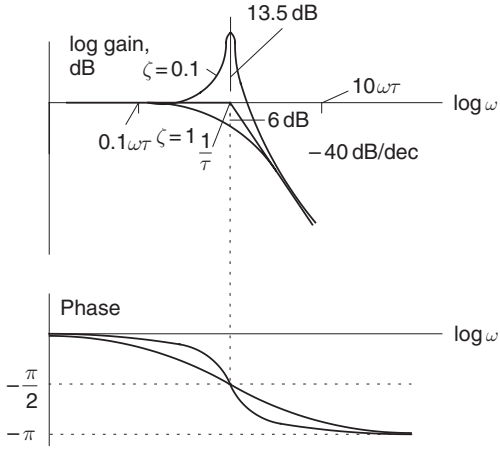


Figure 6.46 The Bode plot of a network with a quadratic pole (not at the origin of the p-plane). The error at the intersection point of two approximation lines depends on the damping factor in the transfer function of the network

originates from the insufficient gain or/and phase margins. Once the problems have been identified, the next step will be to modify the frequency responses of both the devices. In the case of the preamplifier, one has to achieve the proportionality of the phase shift with respect to frequency. The SQUID amplifier must have either a lower gain at the cross-over frequency ω_π or, alternatively, a smaller phase shift at the unity gain frequency, ω_1. Here we assume that amplifiers meet the Nyquist criterion for the absolute stability.

Re-shaping of the frequency response is achieved with specially designed compensation networks or correctors. The presence of a corrector corresponds to adding zeroes or poles to the overall transfer function of the system so that

$$T_{\text{cor}}(j\omega) = [K_0 W(j\omega)]T(j\omega) \tag{6.1.93}$$

where $K_0 W(p)$ is the transmittance function of the compensation network.

Compensating transmittance can assume numerous forms. Yet, all of them can be classified into three base types:

1. lag compensation;
2. lead compensation; and
3. lag–lead combined compensation.

In what follows, we will give examples of correctors of each type.

6.1.4.5 Lag compensation

The transfer function of the lag compensator is given by

$$W(j\omega) = \frac{1 + \frac{j\omega}{\omega_{\text{lag}}}}{1 + \alpha \frac{j\omega}{\omega_{\text{lag}}}} \tag{6.1.94}$$

where $\alpha > 1$.

A circuit diagram of the network that realises the lag transfer function is depicted in Figure 6.47. The transfer function $W(j\omega)$ can be re-written as a product of the simple pole and the simple zero with corner frequencies $\frac{\omega_{\text{lag}}}{\alpha}$ and ω_{lag} respectively. The asymptotic approximation of the Bode plot can be drawn immediately by inspection referring to the equations (6.1.85) and (6.1.89). The Bode plot is presented in Figure 6.48(a). Figure 6.48(b) shows also the Nyquist polar diagram of the transfer function, which completes the picture regarding the frequency response of the network.

In what follows, we will give some remarks on properties and usage of lag compensators:

1. The lag compensator serves to decrease the return ratio by as much as a factor of α at frequencies $\omega > \omega_1$. In other words, the lag networks are essentially low-pass filters. They are normally tuned in such a way that the system preserves a high gain at low frequencies but lowers the gain in the frequency range centred at ω_1 to avoid system instability. Under ω_1 we mean the unity gain frequency of the system. The compensation of magnitude–frequency

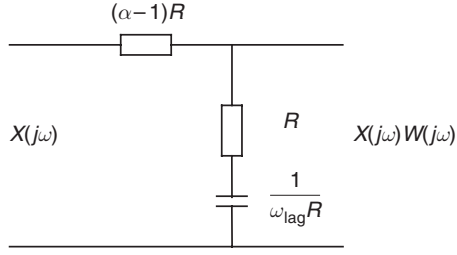


Figure 6.47 A circuit diagram of the network that realises the lag transfer function

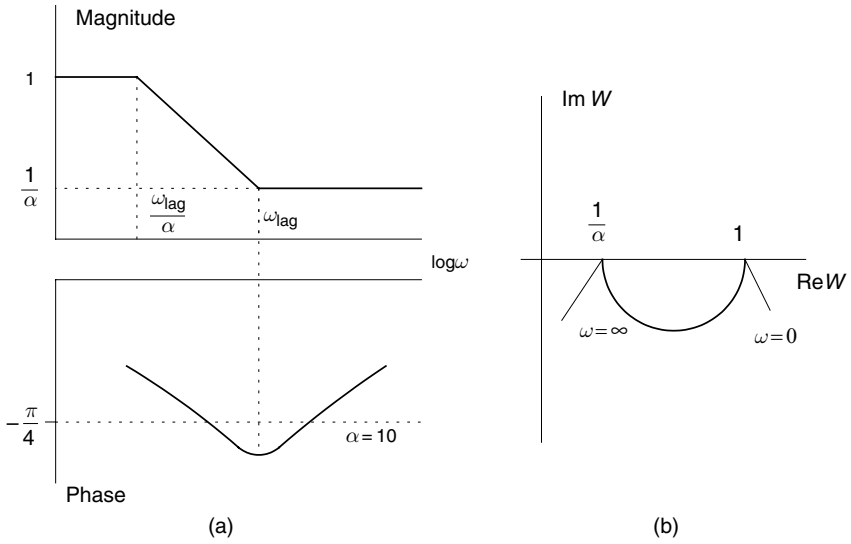


Figure 6.48 The Bode plot (a) and the Nyquist polar diagram (b) of the transfer function of the lag compensator

characteristic results in shifting the unity gain frequency to a lower point where the phase margin is acceptable.

2. According to Figure 6.48(a), the lag network adds to the phase shift. Therefore, it is of no use for the phase compensation purposes.
3. In order to avoid instability caused by the introduction of an extra pole and additional phase shift, the corner frequency of the compensator must be much less compared to the unity gain frequency of the non-compensated system.
4. One has to remember that being essentially the low-pass filters, all lag compensators reduce the bandwidth of the system and slow down the transient response.

6.1.4.6 Lead compensation

An example of the lead compensation network involving passive components only is depicted in Figure 6.49. Its transfer function has the following form:

$$W(j\omega) = \frac{1 + \frac{j\omega}{\omega_{\text{lead}}}}{1 + \frac{j\omega}{\beta\omega_{\text{lead}}}} \tag{6.1.95}$$

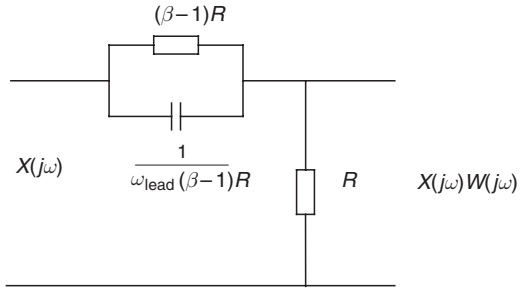


Figure 6.49 A circuit diagram of the network that realises the lead transfer function

where the parameter $\beta > 1$.

The lead network provides compensation by reducing the overall phase shift of a system in the low-to-medium frequency range, whereas introducing negligible attenuation at high frequencies. Figure 6.50 presents the frequency characteristics of the network including Bode plots (asymptotic approximation) and the Nyquist diagram. The phase-to-frequency curve was calculated for $\beta = 10$.

According to the Nyquist diagram depicted in Figure 6.50(b), the maximum lead angle of the network can be found from the following equation:

$$\sin \phi_{\mathrm{m}} = \frac{1 - \frac{1}{\beta}}{1 + \frac{1}{\beta}} \tag{6.1.96}$$

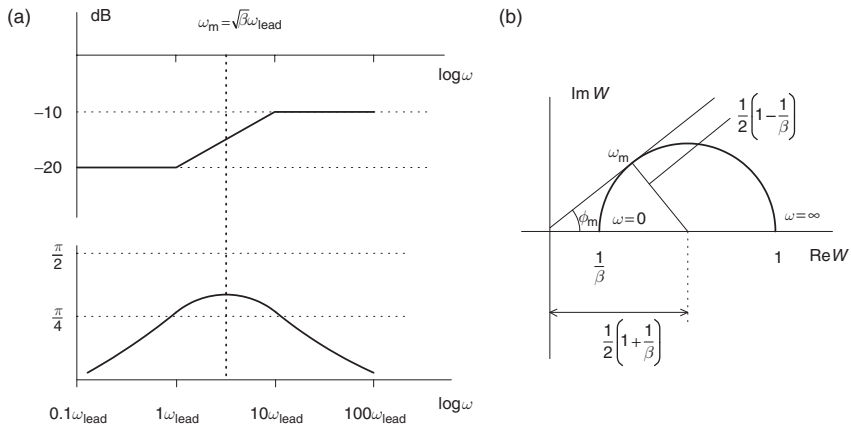


Figure 6.50 The frequency characteristics of the network including Bode plots (asymptotic approximation) (a) and the Nyquist diagram (b). The phase-to-frequency curve was calculated for $\beta = 10$

The lead compensator represents a class of proportional differentiating circuits. In order to maintain its transfer function in the wide frequency range, the real circuit will require presence of active components adding some gain at high frequencies and buffering its input and output. From the point of view of the dynamics, the implementation of the circuit is very beneficial. It increases the phase margin as well as the bandwidth followed by reduced transient response (including both the rise and settling time). Nevertheless, such a compensation technique may not be acceptable in low-noise electronic devices. A larger bandwidth means an increased level of high-frequency noise circulating in the loop. We have already learned in Chapter 2 discussing SQUID parametric amplifiers that the high-frequency noise mixes up with the pumping signal producing undesirable non-linear effects. These non-linear effects can upset the system stability on their own. A high-frequency noise above a certain level is capable to suppress the SQUID response altogether. In order to circumvent this and other problems one recommends utilising a combined lag–lead network.

6.1.4.7 Lag–lead compensation network

The lag–lead compensator achieves phase lag in one range of frequencies and phase lead in another. Its transfer function is given by

$$W(j\omega) = \frac{\left(1 + \frac{j\omega}{\omega_{\text{tag}}}\right)\left(1 + \frac{j\omega}{\omega_{\text{lead}}}\right)}{\left(1 + \alpha\frac{j\omega}{\omega_{\text{tag}}}\right)\left(1 + \frac{j\omega}{\beta\omega_{\text{lead}}}\right)} \tag{6.1.97}$$

where $\alpha > 1$, $\beta > 1$. $W(j\omega)$ is, in fact, a product of individual transfer function of the lag and lead networks

$$W(j\omega) = W_{\text{lag}}(j\omega)W_{\text{lead}}(j\omega) \tag{6.1.98}$$

which we discussed earlier. Therefore, the lag–lead compensation combines the advantages of the both the lag compensation and the lead compensation. This is achieved without such undesirable side effects like an excessive bandwidth or the fact that the time constant of lag network must be much larger than the largest time constant of the system to be corrected, and so on. Since the choice of parameters influencing the frequency response of the compensator is governed by dynamics as much as by noise properties (or other considerations), it is difficult to give general recommendation on a technique for its employment. If, however,

we concentrate for a moment on the system frequency response only, some results of the corrective action can be predicted. To do that, we give an example of the lag–lead passive network, its Nyquist diagram and the Bode plots for parameters $\alpha = \beta = 10$ and $\omega_{\text{lead}} = 10\omega_{\text{lag}}$ in Figures 6.51 and 6.52 respectively. It can be seen that the corrector acts as a lag network in the frequency range $0 < \omega < \omega_0$, while it acts as a lead network at other frequencies. The cross-over frequency at which the phase angle is zero depends on corner frequencies ω_{lag} and ω_{lead} in the following form:

$$\omega_0 = \sqrt{\omega_{\text{lag}}\omega_{\text{lead}}} \tag{6.1.99}$$

This equation follows directly from the Nyquist polar plot. The phase lag portion of the lag–lead network provides attenuation near and

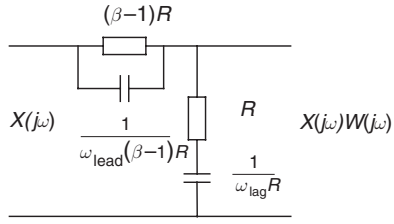


Figure 6.51 A circuit diagram of the network that realises the lag–lead transfer function

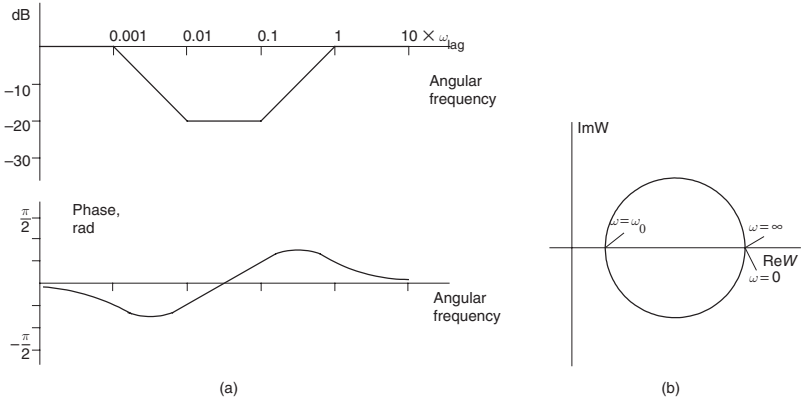


Figure 6.52 The frequency characteristics of the network including Bode plots (asymptotic approximation) (a) and the Nyquist diagram (b). Both the frequency curves were calculated for $\alpha = \beta = 10$

above the unity gain cross-over frequency of a system under correction. This enables an increase of its gain at a low-frequency range to improve the steady state performance, for instance reducing the non-compensation error. The phase lead part of the lag–lead network modifies the frequency response curves via the compensation of the phase shift of the system by an amount of the lead angle. This action increases the phase margin at the unity gain cross-over frequency. The extra margin can again be used to increase the loop amplification at the low-frequency range or, alternatively, to speed up attenuation above the unity gain frequency. The latter is often necessary to reduce circulating in the loop noise and high-order harmonics produced by existing non-linearities or/and switching components.

The presence of non-linear or switching elements in the loop may have a serious effect on the system time dynamics and frequency response. Now, we will proceed to the discussion of these problems.

6.1.5 Describing Function of Non-linear Elements in the Frequency Domain

In Chapter 2, we analysed the SQUID amplifier in two limits: (a) a small-signal limit and (b) a large-signal limit. This is done because the amplifier contains a non-linear element, which is the SQUID itself. In the small-signal limit, we assume the transfer function to be linear and are able to apply all analysis and design methods and techniques developed for linear continuous time circuits. In the large-signal limit, we take into account the effect of the non-linearities on the system dynamics and resort to the describing functions, non-trivial analysis methods such as Lyapunov criteria or the phase planes, and so on.

The approximation to the first order of the working branch of the SQUID flux-to-voltage characteristic yields a pattern known as the saturation non-linearity. The non-linearity is shown in Figure 6.53. Indeed, for small-input signals, the output of such an element would be their product with the transfer function, η, equal to the gradient of the flux-to-voltage curve in its middle, that is, $\eta = (\frac{\partial V}{\partial \Phi})|_{o}$. As the input signal becomes larger, so that the operating point reaches the saturation (horizontal) branches, the output is no longer proportional. Finally, for very large-input signals, the system loses its ability to respond. Its output falls into another saturation set this time by the voltage of a power supply.

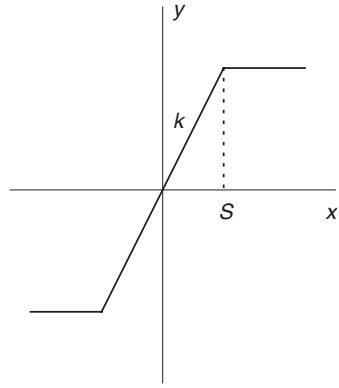


Figure 6.53 The plot of the saturation non-linearity. It adequately describes to the first-order approximation the working branch of the SQUID flux-to-voltage characteristics

In a range of the input signals, where the response of the non-linear element ceases to be proportional, but the overall system still operates to requirement specifications, we need to switch from the transfer function to the describing function. Let us assume that a non-linear element was exposed to an input sinusoidal signal

$$x(t) = X \sin \omega t$$

The output of the element will contain the first, fundamental harmonic as well as a series of the higher order harmonics. The describing function analysis in the frequency domain takes into consideration only the fundamental harmonic component. Neglecting the higher-order harmonics is justified since they are often of a much smaller amplitude compared to the first one from the beginning. In addition to that, inertial networks responsible for stability of the system loop act as low-pass filters. They further attenuate unwanted harmonics together with circulating high-frequency spurious signals and noise.

The Fourier series of the signal at the output of a non-linear unit has the following form:

$$\begin{aligned} y(t) &= A_0 + \sum_{n=1}^{\infty} Y_n \sin(n\omega t + \phi_n) \\ &= A_0 + \sum_{n=1}^{\infty} (A_n \cos n\omega t + B_n \sin n\omega t) \end{aligned} \tag{6.1.100}$$

where

$$A_n = \frac{1}{\pi}\int_0^{2\pi} y(t)\cos(n\omega t)\mathrm{d}(\omega t) \tag{6.1.101}$$

$$B_n = \frac{1}{\pi}\int_0^{2\pi} y(t)\sin(n\omega t)\mathrm{d}(\omega t) \tag{6.1.102}$$

$$Y_n = \sqrt{A_n^2 + B_n^2} \tag{6.1.103}$$

$$\phi_n = \tan^{-1}\left(\frac{A_n}{B_n}\right) \tag{6.1.104}$$

Note that for symmetric non-linearity a factor $A_0 = 0$.

Following that, the fundamental harmonic component at the output is

$$\begin{aligned} y_1(t) &= A_1\cos\omega t + B_1\sin\omega t \\ &= Y_1\sin(\omega t + \phi_1) \end{aligned} \tag{6.1.105}$$

Now, we can write down the definition of the describing function as

$$\begin{aligned} N &= \frac{Y_1}{X}\mathrm{e}^{j\phi_1} \\ &= \frac{\sqrt{A_1^2 + B_1^2}}{X}\mathrm{e}^{j\tan^{-1}\left(\frac{A_1}{B_1}\right)} \end{aligned} \tag{6.1.106}$$

where Y_1 and ϕ_1 are the amplitude and the phase of the fundamental harmonic of the output signal respectively.

When $\phi_1 \neq 0$, the describing function becomes a complex quantity. N is always a function of the input amplitude X. Yet, if a non-linear element contains any energy storage elements it will depend also on the frequency of the input signal. For instance, there is a small parasitic capacitance between the input coil and the SQUID washer. It can be safely ignored at relatively low frequencies where we assume the input coil and the SQUID washer are coupled through the mutual inductance only. But as an input-frequency approaches closer to the transmission line resonance of the input coil, the capacitance has an increasingly stronger by-passing effect and, as a result, modifies the describing function. This modification must be taken into account in the design of HF-SQUID amplifiers intended to operate near 1 GHz.

Following the last statement, we will re-write the definition of the describing function in a more general form (Figure 6.54)

$$N(X,\omega) = \frac{Y_1(X,\omega)}{X} e^{j\phi_1(X,\omega)} \tag{6.1.107}$$

In order to perform analysis and design in the frequency domain, system non-linearities existing in the loop should be replaced by describing functions. After that, we can apply all analytical as well as the graphic analysis and design techniques developed for linear continuous systems, including those presented earlier in this chapter.

A describing function of a non-linear element with saturation, shown in Figure 6.53, is given by

$$N = \frac{2k}{\pi}\left[\sin^{-1}\left(\frac{S}{X}\right) + \frac{S}{X}\sqrt{1-\left(\frac{S}{X}\right)^2}\right] \tag{6.1.108}$$

where k is the gradient of the $Y(X)$ in the linear range, S is the X-co-ordinate of a point, at which the slope and saturation lines intersect.

More close approximation of the SQUID-like flux-to-voltage curve can be achieved with a shape presented in Figure 6.55. In this case, the describing function takes the following form

$$N = k_2 + \frac{2k_1 - k_2}{\pi}\left(\sin\frac{S}{X} + \frac{S}{X}\sqrt{1+\frac{S^2}{X^2}}\right) \tag{6.1.109}$$

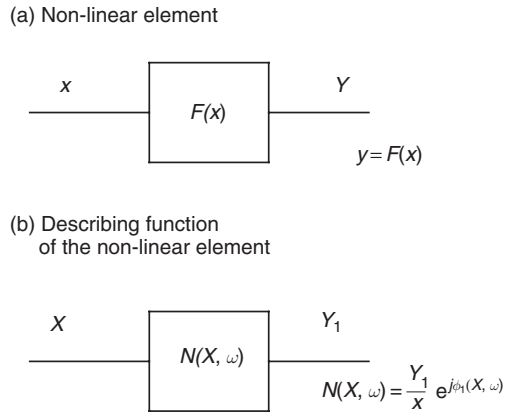


Figure 6.54 A block diagram of a non-linear element with the transfer function $F(x)$ (a), and its equivalent block diagram with the describing function $N(X, \omega)$

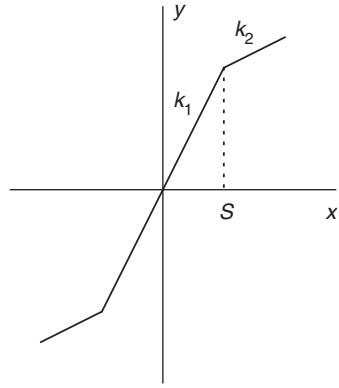


Figure 6.55 More close approximation of the SQUID-like flux-to-voltage curve

The describing function analysis works particularly well for systems containing a single non-linear element. The SQUID amplifier illustrated in Figure 6.56 is a good example of such a system. Its closed loop describing function is given by

$$T(j\omega) = \frac{N(X,\omega)G(j\omega)}{1 + N(X,\omega)\beta} \tag{6.1.110}$$

where X represents in this case a magnitude of the sinusoidal reduced magnetic flux. As usual, the frequency analysis starts from looking at pole positions of the characteristic equation $1 + N(X, \omega)G(j\omega)\beta = 0$. We need to check whether there exist any critical values of X and ω that satisfy the characteristic equation or, in other words, would cause the

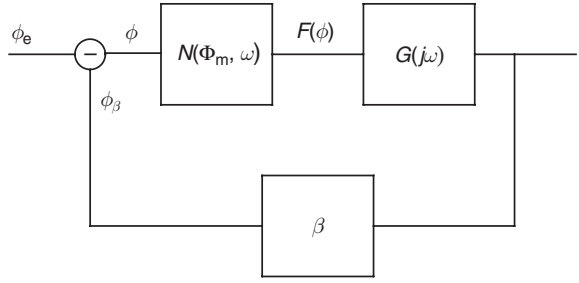


Figure 6.56 An equivalent block diagram of the SQUID amplifier incorporating a non-linear SQUID block with the describing function $N(\Phi_m, \omega)$

oscillations in the loop. It helps if we re-write the equation in a modified form as follows

$$G(j\omega) = -\frac{1}{N(X,\omega)\beta} \tag{6.1.111}$$

Now, we can draw separate plots for G(jω) and $-\frac{1}{N(X,\omega)\beta}$ on the same Nyquist diagram. Absence of any intersection of these two plots would indicate that the critical values do not exist.

Popov formulated a general stability criterion for non-linear feedback systems containing a single non-linear element. It states that a system is stable if

1. the linear inertial element $G(j\omega)$ is stable;
2. $\mathrm{Re}G(j\omega) > -\frac{1}{K}$; and
3. $F(0) = 0$ and $0 < \frac{\beta F(\phi)}{\beta\phi} < K$.

In the criterion, $F(\phi)$ is the non-linear function of the non-linear element. It should not be confused with the describing function (Chapter 2).

6.2 DISCRETE-TIME SYSTEMS AND SYSTEMS WITH PERIODICALLY CHANGING PARAMETERS

As we mentioned before, the electronic world becomes increasingly digital. Practising electronic designers tend to implement the digitisation of signals at the earliest possible stage in order to eliminate problems arising from analogue components. These include the variations of nominal values with time and temperature, extra noise and lack of flexibility. Under the latter, we mean that some transfer functions either cannot be realised in the analogue form or would make the realisation exceedingly complex. Some adjustments to the transfer function of an analogue unit often mean the re-design, while, with a digital system, these require mere changing of some factors in the processor algorithm. Yet, the introduction of the digitisation stage does not change the essence of the system analysis and design. A digital system can equally be intrinsically unstable or produce the overshoots in its output strings, and so on. In this respect, the analogue and digital design skills blend in the digital environment. For instance, it is common to read job advertisements offering positions of an analogue electronic engineer who is

expected to develop digital instrumentation. There is no contradiction in the announcements. Employers are likely to be involved in the development of some variety of the digital control systems. Digital designers are often not familiar with the analysis and design techniques we have discussed so far. Therefore, by tradition, the dynamics of systems in the time and frequency domains is attributed to analogue designers.

Despite the fact that the analysis and design of analogue and discrete-time systems are conceptually very similar, the presence of a sampler or a number of samplers in the loop introduces some specific features. We will outline them in the following paragraphs.

6.2.1 *z*-Transform Method

The discrete-time systems differ from the continuous-time ones in that the signals propagate in the sampled data form through, at least, a part of the loop. A good example to this definition would be a system containing an analogue-to-digital converter (ADC) which performs two consecutive functions of the sampling (digitisation) and then the quantisation.

In terms of sampling operations, the systems based on (or including) superconducting devices have to deal mostly with a periodic sampling when the sampling instants are equally spaced in time, that is, $t_k = kT(k = 0,1,2,\dots)$. This case is well handled by a method called *z*-transform.

The *z*-transform in the discrete-time systems is a counterpart of the Laplace transform in the continuous time systems. In fact, its definition follows from the Laplace transform. Let us consider an ideal signal multiplier, shown in Figure 6.57. During the sampling process an input signal $x(t)$ is sampled, for instance, by closure of an analogue switch.

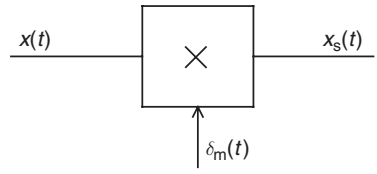


Figure 6.57 A schematic representation of an ideal signal multiplier. During the sampling process an input signal x(t) is sampled, for instance, by the closure of an analogue switch. $\delta_m(t)$ represents a signal controlling the switch. In the time domain, it is a train of very short unit pulses occurring at time moments $t = kT$

$\delta_m(t)$ represents a signal controlling the switch. In the time domain, it is a train of very short unit pulses occurring at time moments $t = kT$

$$\delta_{\mathrm{m}}(t) = \sum_{k=0}^{\infty} \delta(t - kT) \tag{6.2.1}$$

The sampler performs the function of a modulator. The output signal of the sampler, $x_s(t)$, is a product of the two time functions

$$x_{\mathrm{s}}(t) = x(t)\delta_{\mathrm{m}}(t)$$

Thus, the output signal is also a train of pulses with the widths controlled by the pulse widths of $\delta_{\mathrm{m}}(t)$ and the amplitudes equal to instantaneous values of $x(t)$ at the time moments kT. The area of each individual pulse is in fact a mean value of $x(t)$ during the sampling interval. Assuming $\delta_{\mathrm{m}}(t)$ being a train of δ-functions (the unit area and infinitely small duration), we derive the output signal in the following mathematical form:

$$x_{\mathrm{s}}(t) = \sum_{k=0}^{\infty} x(kT)\delta(t - kT) \tag{6.2.2}$$

Taking the Laplace transform of the output signal yields

$$X_{\mathrm{s}}(p) = L[x_{\mathrm{s}}(t)] = \sum_{k=0}^{\infty} x(kT)\mathrm{e}^{-kTp} \tag{6.2.3}$$

Now, introducing a new variable $z = \mathrm{e}^{Tp}$, we obtain the z-transform of the function $x_{\mathrm{s}}(t)$ as follows:

$$X(z) = X_{\mathrm{s}}(p) = x_{\mathrm{s}}\left(\frac{1}{T}\ln z\right) = \sum_{k=0}^{\infty} x(kT)z^{-k} \tag{6.2.4}$$

Its notation may also look like $Z\{x_{\mathrm{s}}(t)\}$.

The z-transform deals with signals at sampling instants only defined by a sampler. In this respect, there is no difference between two notations $Z\{x(t)\}$ and $Z\{x_{\mathrm{s}}(\mathrm{t})\}$.

Some functions and their z-transforms are listed in a short Table 6.4. Their applicability can be expanded by using Table 6.5 that summarises the properties of the z-transforms. More complete lists of z-transform

Table 6.4 Table of z-transform

Function, $f(t)$	z-Transform, $F(z)$
$\delta(t)$	1
$\delta(t-kT)$	z^{-k}
$1(t)$	$\dfrac{z}{z-1}$
t	$\dfrac{Tz}{(z-1)^2}$
$e^{-\alpha t}$	$\dfrac{z}{z-e^{-\alpha T}}$
$\sin \omega t$	$\dfrac{z \sin \omega T}{z^2 - 2z \cos \omega T + 1}$
$\cos \omega t$	$\dfrac{z(z - \cos \omega T)}{z^2 - 2z \cos \omega T + 1}$
$e^{-\alpha t} \sin \omega t$	$\dfrac{ze^{\alpha T} \sin \omega T}{z^2 - 2ze^{-\alpha T} \cos \omega T + e^{-2\alpha T}}$
$e^{-\alpha t} \cos \omega t$	$\dfrac{z(z - e^{\alpha T} \cos \omega T)}{z^2 - 2ze^{-\alpha T} \cos \omega T + e^{-2\alpha T}}$
$\dfrac{1}{(z-a)(z-b)}$	$\dfrac{1}{a-b}(a^{k-1} - b^{k-1})$ for $k > 0$ 0 for $k = 0$
$\dfrac{z}{(z-a)(z-b)}$	$\dfrac{1}{a-b}(a^k - b^k)$
$\dfrac{z(1-a)}{(z-a)(z-b)}$	$1 - a^k$

pairs can be found in the specialised literature on the analysis and design of the discrete systems and the digital control.

6.2.1.1 Inverse z-transform

The usage of the z-transform method is similar to the Laplace counterpart. First, we find a z-transform of the output response of a system or a network

$$Y(z) = T(z)X(z) \tag{6.2.5}$$

The next step is taking an inverse z-transform of $Y(z)$ to derive the output response in the time domain, that is,

$$y(t)\big|_{t=kT} = Z^{-1}[Y(z)] \tag{6.2.6}$$

This is illustrated in Figure 6.58.

Table 6.5 Properties of z-transform

$f(t)$ or $f(k)$	$Z[x(t)]$ or $Z[x(k)]$
$af(t)$	$aF(z)$
$f_1(t) + f_2(t)$	$F_1(z) + F_2(z)$
$f(t+T)$ or $f(k+1)$	$zF(z) - zF(0)$
$f(t+kT)$	$z^kF(z) - z^kf(0) - z^{k-1}f(T) - \cdots - zf(kT-T)$
$f(k+m)$	$z^mF(z) - z^mf(0) - z^{m-1}f(1) - \cdots - zf(m-1)$
$tx(t)$	$-Tz\dfrac{\mathrm{d}}{\mathrm{d}z}[F(z)]$
$kf(k)$	$-z\dfrac{\mathrm{d}}{\mathrm{d}z}[F(z)]$
$\mathrm{e}^{-at}f(t)$	$F(z\mathrm{e}^{aT})$
$\mathrm{e}^{-ak}f(k)$	$F(z\mathrm{e}^{a})$
$a^k\ f(k)$	$F\left(\dfrac{z}{a}\right)$
$ka^k\ f(k)$	$-z\dfrac{\mathrm{d}}{\mathrm{d}z}\left[x\left(\dfrac{z}{a}\right)\right]$
$\sum_{k=0}^{\infty} f(k)$	$F(1)$
$\sum_{k=0}^{\infty} f(kT)y(nT-kT)$	$F(z)Y(z)$
$f(0)$	$\lim F(z)$ at $z \to \infty$ if exists
$f(\infty)$	$\lim[(z-1)F(z)]$ at $z \to \infty$ if $\dfrac{z-1}{z}F(z)$ is analytical on and outside the unit circle

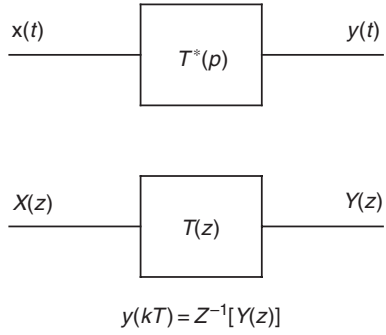


Figure 6.58 A discrete time system with a transfer function $T(p)$ and its equivalent block diagram in the z-domain

Since $Y(z)$ is a z-transform of $y(t)$ defined at $t = kT$ ($k = 0, 1, 2, \ldots$), the inverse z-transform of $Y(z)$ restores $y(t)$ only at these sampling instants. The definition of the inverse z-transform is given by

$$f_{\mathrm{k}} = \frac{1}{2\pi j} \oint_C F(z) z^{k-1} \mathrm{d}z \tag{6.2.7}$$

where C is a closed path that encloses all the singularities of the integral. The circle C is centred at the origin of the z-plane. Its radius z_{r} must be greater than z-values at which the function $F(z)$ converges.

In all practical systems, $F(z)$ is a rational function. Therefore, the contour integration is rarely used by system designers. Instead, the sought values of f_k can be evaluated using the theory of residues. Once we have identified poles of the function $F(z)$, its residue at a selected pole p_i can be derived from the Laurent expansion of $F(z)$ in the vicinity of p_i. Let us write down a Laurent series of $F(z)$ as

$$F(z) = a_{-n}(z - p_i)^{-n} + \cdots + a_{-1}(z - p_i)^{-1} + a_0 + a_1(z - p_i) + \cdots \tag{6.2.8}$$

The residue of $F(z)$ at a pole p_i, is a coefficient a_{-1}.

If a function $F(z)$ contains only simple poles (not repeated poles) the sampled values of $f(t)$ can be found from the following equation:

$$f_{\mathrm{k}} = \sum_{i=1}^{n} r_i \tag{6.2.9}$$

where

$$r_i = \lim(z - p_i) F(z) z^{k-1} \quad \text{at} \quad z \to p_i \tag{6.2.10}$$

If $F(z)$ contains repeated poles of the order m at $z = p_0$ the residue takes the modified form which is

$$r_i = \frac{1}{(m-1)!} \lim \frac{\mathrm{d}^{m-1}}{\mathrm{d}z^{m-1}} \left[(z - p_0)^m F(z) z^{k-1} \right] \tag{6.2.11}$$

There are a number of other approaches to derive the inverse z-transform. They are based on the modification of $F(z)$ to a form which can be handled with a table of existing transform pairs. Here, we will briefly discuss one modification method by expansion of $F(z)$ into partial

fractions. Provided that $F(z)$ is rational function we can represent it in the following form:

$$F(z) = \frac{\sum_{i=0}^{n} b_i z^i}{\prod_{i=1}^{m} (z + p_i)^{n_i}} = b_n + \sum_{i=1}^{m} \sum_{k=1}^{n_i} \frac{c_{ik}}{(z + p_i)^k} \tag{6.2.12}$$

where the second term constitutes the partial fraction of $F(z)$ with the factors c_{ik} given by

$$c_{ik} = \frac{1}{(n_i - k)!} \frac{\mathrm{d}^{n_i - k}}{\mathrm{d}z^{n_i - k}} \left[(z + p_i)^{n_i} F(z) \right] \Big|_{z = -p_i} \tag{6.2.13}$$

Thus, the inverse z-transform is obtained as the sum of the inverse z-transforms of the partial fractions. The partial fractions are likely to be ratios of simple first-order zeroes to simple poles, or just simple poles only. Therefore, all terms can be represented in the time domain by the use of existing tables.

Example 6.1

In this example, we will find an inverse z-transform of a function given by

$$F(z) = \frac{az}{(z - b)(z - c)}$$

Expanding it into partial fractions, we obtain

$$F(z) = \frac{a}{(c - b)} \left(-\frac{z}{z - b} + \frac{z}{z - c} \right)$$

From Table 6.4,

$$Z^{-1}\left[\frac{z}{z - b}\right] = b^k \quad \text{and} \quad Z^{-1}\left[\frac{z}{z - c}\right] = c^k$$

According to Table 6.5, the samples of the time function are

$$f_k = \frac{a}{c-b}(-b^k + c^k)$$

6.2.1.2 *z*-Plane

In general, z is the complex variable. Therefore, poles and zeroes of a rational function $F(z)$ will be positioned on the 2D complex z-plane. If we represent z in the complex form as $z = a + jb$, then a and jb form the axes of the z-plane as shown in Figure 6.59.

Due to the fact that p and z variables are mutually related, locations of poles on the z-plane equally must characterise the stability of a discrete-time network or system. To illustrate the statement, we show four examples of the network-unforced time responses in Figure 6.60. First three examples correspond to poles located within/on the unity circle while the fourth one lies on the real axis at a point $a > 1$. According to Figure 6.60(a–c), poles located within the unit circle give rise to sequences that decay with k. The pole outside the unit circle, however, (Figure 6.60[d]) characterises response terms that grow in magnitude with k. This behaviour pattern is in line with the stability criterion of the discrete-time systems: a discrete-time system is said to be asymptotically stable if and only if its poles are all inside the unit circle on the z-plane. A system is asymptotically stable if its unforced output response $y(t) \to 0$ at $t \to \infty$ with any initial conditions. Under the unforced response, we understand a case when no input signal applied.

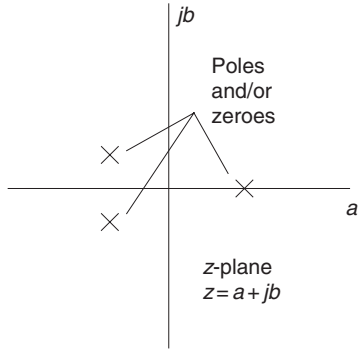


Figure 6.59 A graphical representation of z as a complex plane $z = a + jb$ with a and jb forming its real and imaginary axes

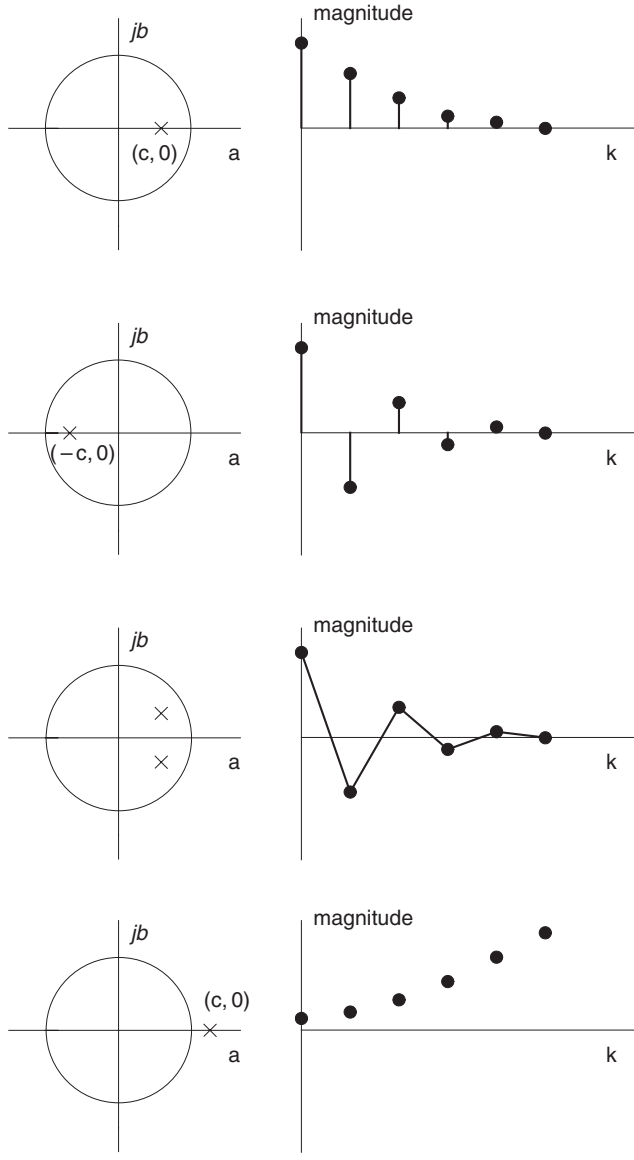


Figure 6.60 Four examples of the network-unforced time responses. First three examples correspond to poles located within/on the unity circle while the fourth one lies on the real axis at a point where $a > 1$. Poles located within the unit circle give rise to sequences that decay with k. The pole outside the unit circle characterises response terms that grow in magnitude with k

6.2.1.3 Transfer function of discrete-time systems

Discrete-time systems normally include discrete- and continuous-time elements and signals. When dealing with z-transform we need to remember that $x(t)$ and $x_s(t)$ are not interchangeable even at very small sampling interval T. Impulses $x_s(t)$ are of infinite height and infinitely small width. They cannot be handled by any practical circuit unless we hold a sampled value for some time, which would be sufficient for the components following the hold circuit to react.

The hold circuitry certainly affects the overall frequency response of the system. Its transfer function must be included into the direct channel or the loop at the design stage. Figure 6.61(a) depicts an example of a system containing a digitiser and a continuous-time network $T(p)$. Both elements are connected in series. A digitiser is represented by a sampler (switch S) and a hold circuit. In general, the order of holding devices is determined by the degree of a polynomial with which the signal is approximated between samples. Three types of hold circuits are most widely used in the digital systems. These are:

1. the zero-order hold;
2. the extrapolative first-order hold; and
3. the polygonal hold.

The circuits and their corresponding approximation principles are illustrated in Figure 6.61(b). The zero-order hold maintains an i-sampled value unchanged until the following $i+1$ sample is taken. The extrapolative first-order hold circuit anticipates the $(i+1)$-sampled value that lies on the line drawn through the previous two samples. Finally, the polygonal hold contains a delay unit with a delay time T to draw a line from one delayed sample value to the next one.

The Laplace transform of the zero-order sample and hold circuit is given by

$$H(p) = \frac{1 - \mathrm{e}^{-Tp}}{p} \tag{6.2.14}$$

In the frequency domain, it takes the following form

$$H(j\omega) = \frac{1 - \mathrm{e}^{-j\omega T}}{j\omega} = T \frac{\sin\left(\frac{\omega T}{2}\right)}{\frac{\omega T}{2}} \mathrm{e}^{-\frac{j\omega T}{2}} \tag{6.2.15}$$

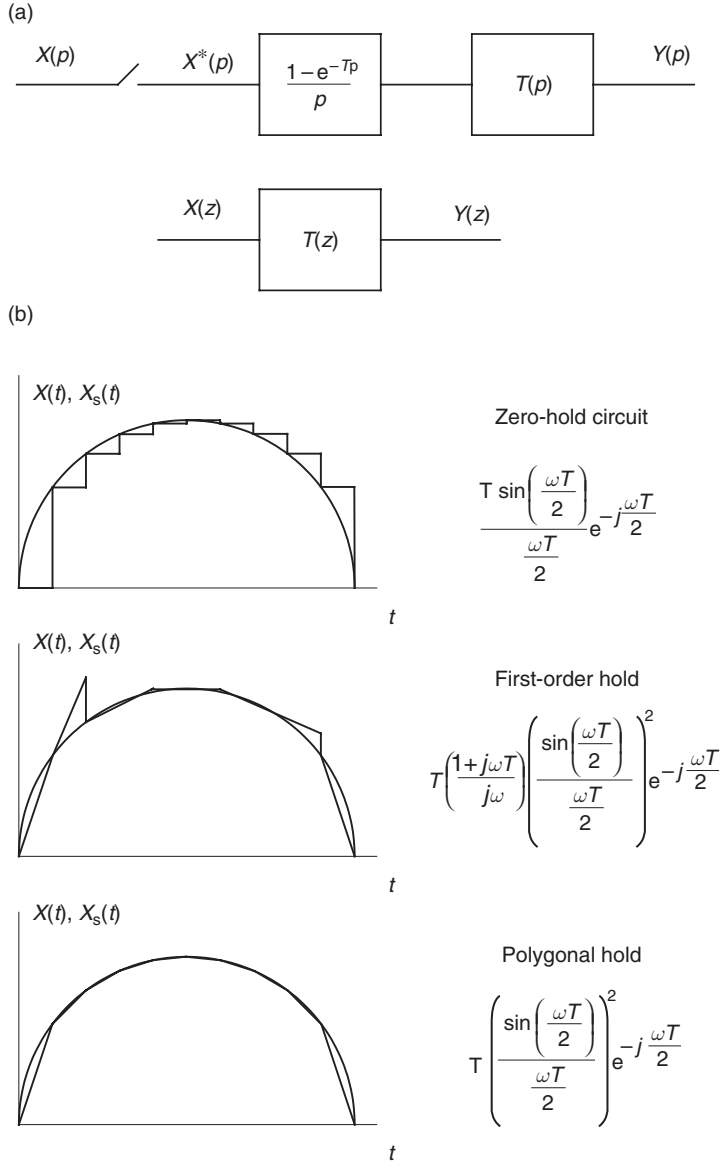


Figure 6.61 An example of a system containing a digitiser and a continuous-time network $T(p)$. (a) Both elements are connected in series. A digitiser is represented by a sampler and a hold circuit. (b) explains principles and gives the Fourier transfer functions of three types of the hold circuits most widely used in the digital filtration. These are: the zero-order hold; the extrapolative first-order hold; the polygonal hold (from top to bottom)

Introducing the sampling frequency $\omega_s = \frac{2\pi}{T}$, we obtain

$$H(j\omega) = \frac{2\pi}{\omega_s} \frac{\sin\left(\pi \dfrac{\omega}{\omega_s}\right)}{\pi \dfrac{\omega}{\omega_s}} \tag{6.2.16}$$

The overall z-transform transfer function of the channel, shown in Figure 6.61(a), can be written down as [9]

$$T(z) = \frac{Y(z)}{X(z)} = (1 - z^{-1}) Z\left\{ L^{-1}\left[\frac{T(p)}{p}\right]\Big|_{t=kT} \right\} \tag{6.2.17}$$

According to the equation (6.2.17), the presence of the sample-and-hold circuit modifies the transfer function $T(p)$. The phenomenon is often referred to as the pre-distortion of the input signal. The pre-distortion has to be addressed in the digital filtering routine following the digitisation and quantisation stage (the ADC) in the spectrometer. Usually this is done by the selection of weighting coefficients that offset the effect of the sample-and-hold circuit.

Digital filtering can be realised as a program routine in a microprocessor, hardwired logic or using the analogue delay lines. While the hardwired logic demonstrates the fastest response time, the microprocessor offers the flexibility, much better accuracy and a larger choice of transfer functions. Both the methods thus can deal with several data streams simultaneously using the time-sharing technique.

The superconducting detector systems are likely to employ microprocessors to perform the pulse shaping routine. There is a number of reasons for that. Microprocessors work well with any of sample-and-hold circuits. In fact, circuits can be interchangeable depending on the requirement specifications for the accuracy of the signal approximation. Besides, the nanoanalysis does not require a high throughput due to a relatively low event rate generated by super-fine current beams. Finally, the pulse shapes of superconducting detectors carry potentially a large-ballistic deficit that needs to be compensated in the filtering routine. The hardware cost of a logic filter that offsets the pre-distortion and the ballistic deficit can be unacceptably high. The latter, however, is a preferred option in high-throughput semiconductor detector systems employing relatively simple pulse shaping transfer functions.

6.2.1.4 Block diagrams and signal flow graphs

Block diagrams and signal flow graphs introduced for continuous systems also apply to the discrete-time systems. Figure 6.62 illustrates examples. The closed loop gain of the depicted system is given by

$$T(z) = \frac{Y(z)}{X(z)} = \frac{G(z)}{1 + G(z)H(z)} \tag{6.2.18}$$

6.2.1.5 Frequency response of discrete-time systems: w-transform

The calculation of the frequency response is a very important part of the system analysis. The magnitude-to-frequency curve and the

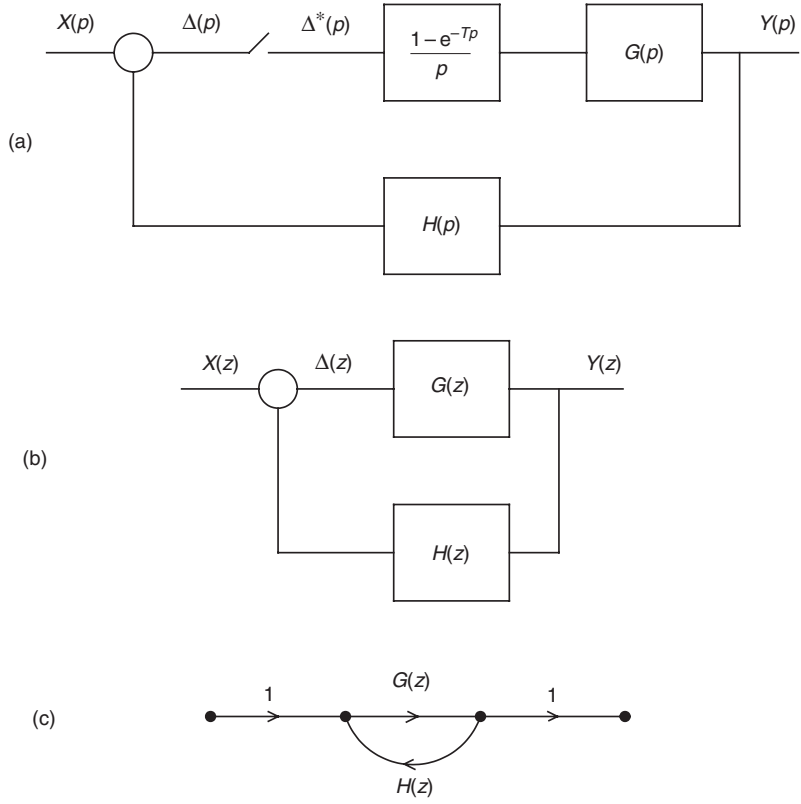


Figure 6.62 An example of a feedback system in the Laplace domain, its equivalent block diagram in the z-domain and the corresponding signal flow graph in the z-domain. All rules and properties of block diagrams and signal flow graphs introduced for continuous systems also apply to the discrete-time systems

frequency-dependent phase shift are those characteristics that can be verified in the experiment. In order to obtain the transfer function of a continuous time network in the frequency domain, we resort to the Fourier analysis by implying $p = j\omega$. A similar study into the discrete-time systems requires operation with the restricted z-transform defined as $G(z)|_{z=e^{j\omega T}}$. Figure 6.63 illustrates specific features of the p-plane and the z-plane in the frequency analysis. While restoring the frequency characteristic of a continuous time network or a system, the frequency of a test signal moves along the positive axis $j\omega$. For the comparison purpose, Figure 6.63(a) shows three frequency points in units defined as a fraction of the sampling frequency of the discrete-time system, ω_s, that is, 0, $\frac{\omega_s}{4}$ and $\frac{\omega_s}{2}$. Since $z = e^{j\omega T}$, the same frequency points on the z-plane lie along the unit circle. We pictured polar plots only up to an angular frequency $\frac{\omega_s}{2}$. By that, we wanted to accentuate that the sampling theorem does not allow the input test signal to exceed this limit.

The presence of the exponent e in the relation $z = e^{j\omega T}$ leads to rather complex mathematical expressions for the transfer functions of the discrete-time systems in the frequency domain. The behaviours of the Nyquist diagrams and the Bode plots derived from these expressions also have their own specifics which are different from what we learned for the continuous systems. Yet, the stability and the stability margins of linear discrete-time systems can be determined using p-plane methods if

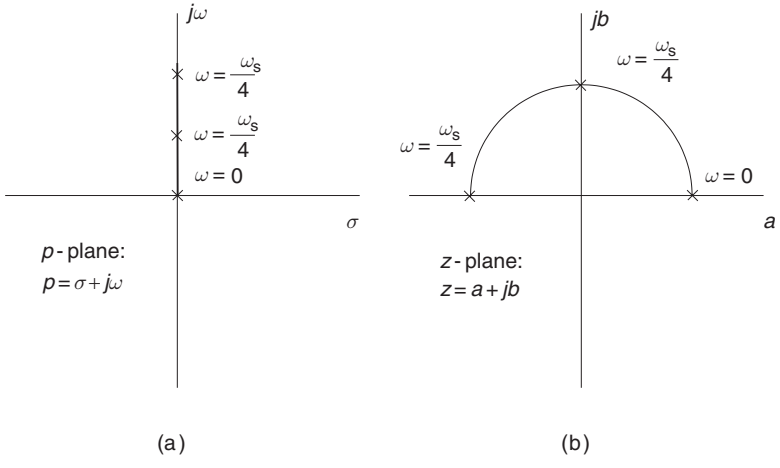


Figure 6.63 The frequency analysis on the p-plane and the z-plane. On the p-plane, the frequency of a test signal moves along the positive axis $j\omega$. Since $z = e^{j\omega T}$, the corresponding frequency points on the z-plane lie along the unit circle

we apply the following bi-linear transformation of z into a new complex variable w given by

$$w = \frac{z-1}{z+1} \tag{6.2.19}$$

The complex variable has a conventional representation via the real and imaginary parts, that is, $w = \sigma_w + j\omega_w$.

The transformation of a transfer function defined in the z-plane into the w-plane is achieved by the substitution

$$z = \frac{1+w}{1-w} \tag{6.2.20}$$

The restrictive condition for the frequency analysis in the w-transform is $\sigma_w = 0$. Following the equation (6.2.20), the equivalent locus in the z-plane takes the form

$$z = \frac{1+j\omega_w}{1-j\omega_w} \tag{6.2.21}$$

Now, since the angular frequency ω_w moves along the imaginary axis $j\omega_w$ from $-\infty$ to $+\infty$, the variable z takes the values

$$z = \mathrm{e}^{j[\tan^{-1}\omega_w - \tan^{-1}(\omega_w)]} = \mathrm{e}^{j2\tan^{-1}\omega_w} \tag{6.2.22}$$

According to the equation (6.2.22), the imaginary axis in the w-plane maps onto the unit circle in the z-plane. Moreover, the interior of the unit circle in the z-plane is entirely positioned on the left hand of the w-plane. This suggests that the frequency domain analysis of the discrete-time systems can be done in the w-plane treating the variable w like the Laplace variable p. In other words, this enables us to draw frequency response plots similar to those for continuous systems and treat them using all continuous time methods in the conventional way. After the analysis has been completed all we need is to substitute the angular frequency ω_w by the true angular frequency. Their relation is

$$\omega_w = \tan\left(\frac{\omega T}{2}\right) \tag{6.2.23}$$

Example 6.2

In this example, we will derive the frequency response of a network, which is described by a z-transform transfer function

$$G(z) = \frac{1}{z - a} \tag{6.2.24}$$

using the w-transform method.

Making the substitution $z = \frac{1+w}{1-w}$ into the equation (6.2.24), we obtain w-transform of the transfer function

$$G(w) = \frac{(1-w)^2}{(1+w)^2 + a(1+w)} \tag{6.2.25}$$

Restricting the variable w to $w = j\omega_w$, we proceed to the frequency analysis on the w-plane, that is,

$$G(j\omega_w) = A(\omega_w) + jB(\omega_w) \tag{6.2.26}$$

where

$$A(\omega_w) = \frac{1 - a + \omega_w^2}{(1-a)^2 + \omega_w^2}, \qquad B(\omega_w) = \frac{(2-a)\omega_w}{(1-a)^2 + \omega_w^2}$$

The inputs to the Nyquist polar diagrams and Bode plots can be found from

$$\left|G(j\omega_w)\right| = \sqrt{A(\omega_w)^2 + B(\omega_w)^2}$$

for the magnitude and

$$\phi(\omega_w) = \tan^{-1}\frac{B(\omega_w)}{A(\omega_w)}$$

for the phase shift. Here $\omega_w = \tan(\frac{\omega T}{2})$.

6.2.2 Digital Filtration

The principle of the digital filtration is explained in Figure 6.64. The ADC yields binary numbers x_k representing the digitised input analogue

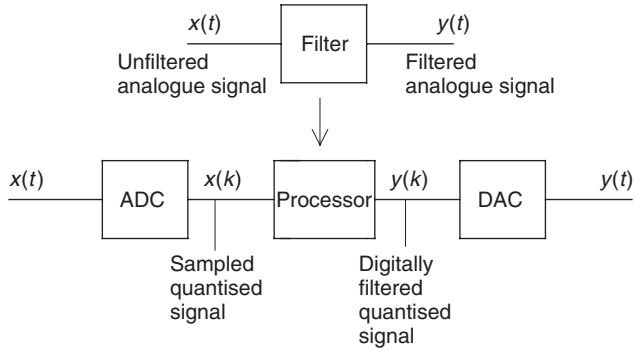


Figure 6.64 A schematic representation of the digital filter

signal $x(t)$. The processor performs a numerical computation on these numbers. The computation involves multiplying x_k by weighting factors defined by a desired transfer function of the filter. The digital filter output signal y_k is essentially a sum of these products. The presence of the DAC is optional. It is needed only if the output signal is to be restored in the analogue form.

Digital filters have a number of advantages over their analogue counterparts. The most important one is their adaptability. Transfer function can be changed or tuned without affecting the hardware (unless, of course, the digital filter is based on a specifically designed hardware logic). The property is particularly valuable for developing general-purpose spectrometers intended to operate with superconducting detectors. The pulse shape of a superconducting detector is quite sensitive to the operating conditions, properties of materials involved and the detector geometry. Therefore, the pulse shaping function of the digital filter may need to be tuned for each individual detector.

Digital filters will never substitute the analogue filters completely. Their application area is restricted by the sampling theorem. Besides, digital filters suffer from finite precision effects caused by the quantisation of the weighting factors as well as product values. Yet, they work extremely well with a suitable class of input signals. By "extremely well", we mean that digital filters have no drift, their transfer function does not vary with time and temperature, a designer does not need to deal with tolerances of components, PCB layouts to ensure a proper stability margin, and so on.

The design of digital filters is based on the z-transform. The classical z-transform that was introduced earlier in the chapter applies to systems with the zero-order sample-and-hold circuit. They are most widely

utilised in the digital spectrometry. High-precision systems may incorporate the first-order sample-and-hold circuit. Such systems will require using the bi-linear z-transform with the relationship between the p-plane and the z-plane given by

$$p = \frac{2(z-1)}{T(z+1)}; \qquad z = \frac{\frac{2}{T}+p}{\frac{2}{T}-p} \tag{6.2.27}$$

We will not discuss details of the bi-linear z-transform. Readers interested in the analysis of discrete systems with the first-order sample-and-hold circuit can find details in the specialised literature (see, for instance, [5] and references therein).

The order of a digital filter is defined by a number of previous sampled values that must be stored in the processor memory to calculate the output y_k. For instance,
The zero-order filter

$$y_k = a_0 x_k \tag{6.2.28}$$

The first-order filter

$$y_k = a_0 x_k + a_1 x_{k-1} \tag{6.2.29}$$

The third-order filter

$$y_k = a_0 x_k + a_1 x_{k-1} + a_2 x_{k-2} + a_3 x_{k-3} \tag{6.2.30}$$

The M-order filter

$$y_k = \sum_{i=0}^{M} a_i x_{k-i} \tag{6.2.31}$$

where a_i are referred to as the weighting factors of the filter.

There are two types of digital filters. The equation (6.2.31) represents the non-recursive filter that calculates the output solely from the current and previous sampled input signal values. Such filters are also known as the FIR filters. The second type embraces filters with the recursive algorithm or infinite impulse response (IIR) filters. The equation

$$y_k = a_0 x_k + a_1 x_{k-1} - b_1 y_{k-1} - b_2 y_{k-2} \tag{6.2.32}$$

gives an example of the second order recursive filter. Following that, the algorithm of the recursive filter can use the current and previous sampled input values x_{k-i} as well as previous computed outputs y_{k-i}. The order of the filter equals to the largest number of previous input values x_i or previous output values required to compute y_k.

A general description of the recursive filter in the discrete time domain is given by

$$y_k = \sum_{i=0}^{M} a_i x_{k-i} - \sum_{i=1}^{N} b_i y_{k-i} \tag{6.2.33}$$

where b_i are the weighting factors of the filter with respect to the previous y_{k-i} values.

To address practical design issues let us introduce a delay operator z^{-1}. When applied to a sampled value x_k this operator gives its previous value x_{k-1}, that is,

$$z^{-1}x_k = x_{k-1} \tag{6.2.34}$$

Thus, z^{-1} introduces a delay of one sampling interval. Applying the delay operator twice produces a delay of two sampling intervals:

$$z^{-1}(z^{-1}x_k) = z^{-2}x_k = x_{k-2} \tag{6.2.35}$$

The notation can be extended to the delay of arbitrary n sampling intervals z^{-n}.

By using a variable z^{-n} for the delay operator, we indicate that it relates directly to the z-transform. Indeed, taking the z-transform of the equation (6.2.33), we obtain

$$Y(z) = X(z)\sum_{k=0}^{M} a_k z^{-k} - Y(z)\sum_{k=1}^{N} b_k z^{-k} \tag{6.2.36}$$

Re-writing it in the symmetrical form

$$Y(z)\left[1 + \sum_{k=1}^{N} b_k z^{-k}\right] = X(z)\sum_{k=0}^{M} a_k z^{-k}$$

we can derive the z-transform of the transfer function of the recursive filter in terms of the delay operator

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{M} a_k z^{-k}}{\left[1 + \sum_{k=1}^{N} b_k z^{-k}\right]} \tag{6.2.37}$$

The z-transform of the transfer function of the non-recursive filter is a special case of $H(z)$ given by the equation (6.2.37), when the weighting factors at previous output values $b_k = 0$.

According to the definition of the z-transform,

$$H(z) = Z[h_k] \tag{6.2.38}$$

where h_k is the unit sample response of the filter. The output sequence can be defined through the unit sample response using the convolution theorem

$$y_k = h_k * x_k = \sum_{i=0}^{k} h_k x_{k-i} \tag{6.2.39}$$

Now, we have all necessary mathematics in place for the filter design from either a desired frequency characteristic or an impulse function. Conceptually, synthesis of the digital filters is very similar to one of the analogue filters, which we discussed in the previous section. The total requirement specifications include the location of passbands, the minimum stopband attenuation, the maximum passband ripple, the order of the analogue filter model, and sometimes the shape of the response in some of the specified bands, for instance, the shaping function in spectrometers. From these requirement specifications, the desired filter response is constructed in the Laplace or frequency domain. In a way, the design starts with a prototype-analogue filter. Then a p-domain to z-domain transformation is used to derive the z-transform transfer function $H(z)$ in the exact form or as a ratio of two polynomials approximating the actual $H(z)$. The transfer function yields the filter order M, N and weighting factors a_i, b_i. At this stage, the design of the digital filter with the floating-point response is completed.

Example 6.3

In this example, we will construct a third-order recursive filter with the floating-point response defined by weighting factors a_0, $-b_1$, $-b_2$, $-b_3$.

In the discrete-time domain, the filter is described by a following equation

$$y_k = a_0 x_k - b_1 y_{k-1} - b_2 y_{k-2} - b_3 y_{k-3} \tag{6.2.40}$$

Using the time delay operator, we obtain

$$y_k = a_0 x_k - b_1 z^{-1} y_k - b_2 z^{-2} y_k - b_3 z^{-3} y_k \tag{6.2.41}$$

The filter described by the equation (6.2.41) can be realised in two direct forms as shown in Figures 6.65 and 6.66.

Converting floating-point weighting factors into fixed point ones requires their quantisation. In this case, the factors can be reproduced only to a certain finite accuracy. For instance, in the 8-bit word, a factor b_i will be quantised to an 8-bit value $\frac{L}{128}$, where L is an integer with a value providing the closest fit for b_i. The quantisation of the weighting factors is illustrated in Figure 6.67.

The finite accuracy of the digital filter is not defined entirely by the quantisation of weighting factors. The output of the multipliers must be quantised as well. Following that, a more complete equivalent circuit diagram takes a configuration shown in Figure 6.68 where we added the extracted quantisation error as a source of the quantisation noise. The resolution of the filter is limited to the least significant bit.

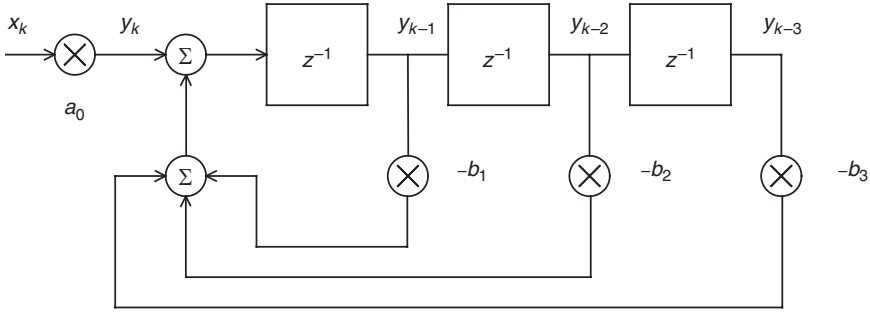


Figure 6.65 A direct form 1 of the third-order digital filter described by the equation (6.2.41)

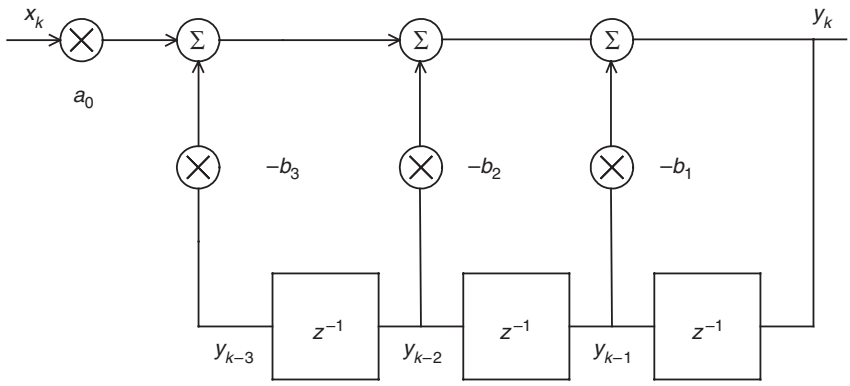


Figure 6.66 A direct form 2 of the third-order digital filter described by the equation (6.2.41)

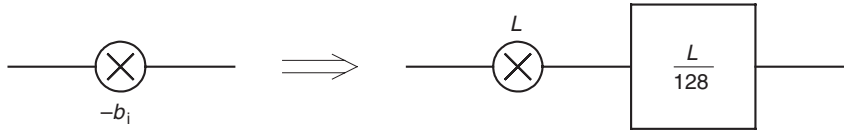


Figure 6.67 An equivalent block diagram of the multiplier unit illustrating the quantisation of the weighting factors

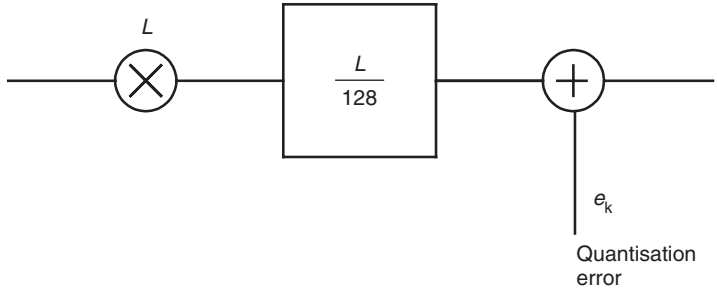


Figure 6.68 A fixed-point equivalent block diagram of the multiplier unit illustrating the quantisation of the output values. The extracted quantisation error is added as a source of the quantisation noise

The quantisation error requires from a designer a particular attention. Added or multiplied by integer values (L in the model presented in Figure 6.68) numbers anywhere in the filter should not exceed a range established by the wordlength of the processor. In the 8-bit processor the range extends from -127 to 128. The overflow may introduce a severe non-linearity into the transfer function of the filter. If this happens, such

a filter can become unstable or/and exhibits a distorted response/no-response in some signal/frequency ranges. The phenomenon is similar to overloading an active analogue filter when the output signal exceeds the power supply voltage. The remedy for that is to reduce the amplitude of the input signal.

Small deviations in weighting factors due to quantisation (even those relating to zeroes of $H(z)$) can translate into a substantial departure of the actual frequency characteristic from its initially specified form. Therefore, the frequency response of the filter model must be repeatedly verified until it meets the required specifications. If it does not, either the number of weighting factor bits can be increased the filter response can be redefined, or the filter arithmetic can be redesigned, and so on. Usually, a designer would apply the combination of these and other measures. When filter hardware is at a premium, sophisticated simulated annealing techniques can be used to produce the best set of filter coefficients, given a fixed filter order and the wordlength.

6.2.2.1 The stability criterion of digital filters

The non-recursive digital filters without the overflow are stable by definition. The design is mostly concerned with right choices of the window, the sampling rate and an optimal wordlength. We will not discuss the procedure here as it is described in general signal processing texts and is standard in most commercially available digital filter design software packages.

When dealing with the recursive digital filters, we must resort to the bounded-input and bounded output stability requirement that follows from the convolution theorem (equation (6.2.39)):

$$|y_k| \leq \sum_{i=0}^{k} |h_k||x_{k-i}| \tag{6.2.42}$$

Provided that the absolute values of the input signal samples do not exceed a certain maximum value x_m, that is,

$$|x_{k-i}| < x_m$$

the equation (6.2.42) can be re-written in the following form

$$|y_k| \leq x_m \sum_{k=0}^{K} |h_k| \tag{6.2.43}$$

The sum

$$g = \sum_{k=0}^{\infty} |h_k| \tag{6.2.44}$$

is referred to as the *g*-norm factor of the unit sample response.

Now, we can formulate the stability criterion as follows: *the linear digital filter is stable if and only if the* g*-norm factor has a finite value.*

The *g*-norm factor is a measure of the maximum gain in the filter between the input signal to internal nodes and that from internal nodes to the output. With a known maximum value of the quantisation error, it helps to determine the minimum wordlength of the processor to satisfy requirement specifications of the discrete system with a feedback and the recursive filter as a special case of such systems.

6.2.3 Systems with Synchronous Multipliers

In this section, we will discuss the dynamics of systems with periodically varying parameters. We will mostly concentrate on a case when the direct path of the loop contains two switches operating in a synchronous mode, as shown in Figure 6.69. The idea behind this concept is to eliminate low-frequency noise and fluctuations contributed by the amplification network $K(p)$. To do so, the first switch performs a function of the modulator. It yields a product of an input signal $x(t)$ with a periodic rectangular M/DM signal. In the frequency domain, the modulation of a signal shifts its spectrum into a frequency range extending from to $[\omega_m - \omega_x]$ to $[\omega_m + \omega_x]$. Following that, the modulation

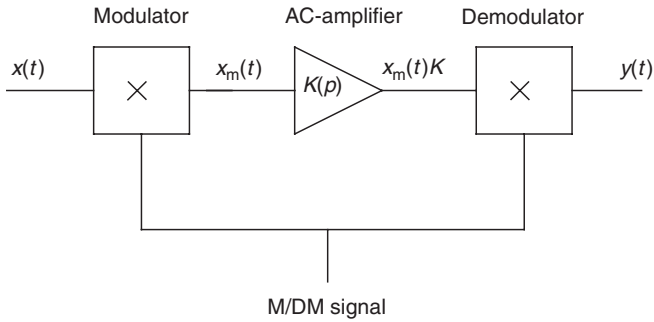


Figure 6.69 A block diagram of a synchronous modulation–demodulation channel

frequency must be at least twice as much as the maximum frequency of the input signal $x(t)$ (the sampling theorem) and well above the range where the amplifier K has substantial 1/f noise. After an AC amplification, $Kx_m(t)$ signal arrives at the input of a second switch working synchronously with the first one. The switch operates as a demodulator that restores an original shape of $x(t)$ with a larger magnitude. Here we assumed that the channel is linear.

This technique is widely used in many commercially available SQUID amplifiers. Due to the periodicity of the flux-to-voltage characteristic, the SQUID can operate as a modulator by mere adding the input and the modulation signal. Figure 6.70 illustrates the principle. For simplicity, we approximated the characteristic by a symmetric triangular shape just for this illustration. The modulated output of the SQUID is then coupled to an AC amplifier via an impedance-matching transformer. Such a configuration enables one using SQUIDs and amplifiers with less

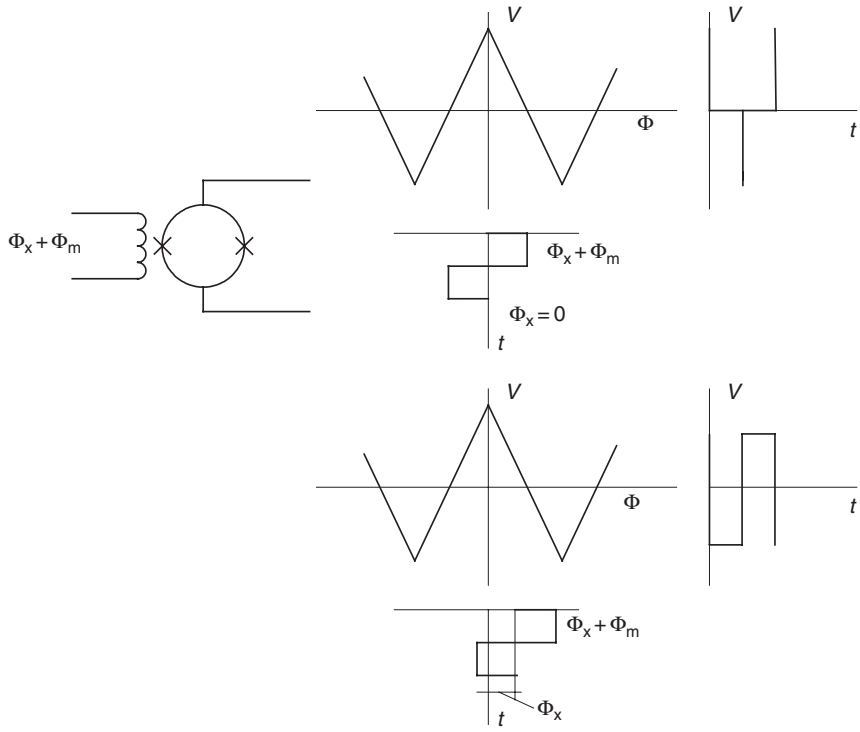


Figure 6.70 Due to the periodicity of the flux-to-voltage characteristic, the SQUID can operate as a modulator by merely adding the input and the modulation signal. For simplicity, we approximated the characteristic by a symmetric triangular shape

stringent specification requirements compared to those applying to SQUIDs employing a direct readout.

A rough analysis of a feedback system comprising a synchronous modulation–demodulation (M–DM) channel can be done by taking into account only the circulation of the fundamental harmonic of the input signal $x(t)$. The approach is viable provided that the following restrictive conditions are met:

- the modulation frequency is much larger than the frequency of the fundamental harmonic;
- AC amplifier is practically non-inertial; and
- an inertial network following the M–DM channel is efficient in filtering out higher-order harmonics.

These restrictions do not cause any problems when systems deal with low- to medium-frequency input signals. In fast systems, in which the unity gain of the loop extends close to ω_m, the harmonics of the higher order may have substantial magnitudes so that their circulation no longer can be neglected. Once the harmonics arrive back to the input of the M–DM channel, they start interacting with the modulation signal and the noise/interference signal. The product of this mixture generates parasitic frequency components within the device bandwidth modifying the overall frequency response of the loop. Since the noise/interference signal and phase shifts of harmonics vary with time, the frequency response becomes non-stationary with adverse consequences for the stability. Even if a system remains stable, the stability margin will certainly shrink.

The method of the equivalent transfer functions proposed in [10] addresses the circulation of the higher-order harmonics in loops containing an M–DM channel and linear inertial networks. It does not restrict the maximum frequency of the input signal with respect to the frequency of the modulation signal. Thus, we are able to predict system performance with respect to a regular input signal $x(t)$. Or alternatively, we can substitute $x(t)$ by the noise/interference signal and observe its influence on the frequency characteristics of the loop.

Let us consider a modulator with an input signal $x(t)$ and a modulation function $m(t)$ shown in Figure 6.57. We assume that the modulation function is a square wave function with the unit magnitude.

The output signal of the modulator in the time domain is given by

$$x_m(t) = x(t)m(t) \tag{6.2.45}$$

Using the definition for the Laplace transform

$$X_{\mathrm{m}}(p) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} X(s)M(p-s)\mathrm{d}s \tag{6.2.46}$$

For the specified modulation function, we obtain

$$X_{\mathrm{m}}(p) = \frac{1}{2j}\left[\frac{1}{p-j\Omega}\tanh\frac{\pi(p-j\Omega)}{2\omega_{\mathrm{m}}} - \frac{1}{p+j\Omega}\tanh\frac{\pi(p+j\Omega)}{2}\right] \tag{6.2.47}$$

if $L\{x(t)\} = L\{\sin\Omega t\} = \frac{\Omega}{p^2+\Omega^2}$ and

$$X_{\mathrm{m}}(p) = \frac{1}{2j}\left[\frac{1}{p-j\Omega}\tanh\frac{\pi(p-j\Omega)}{2\omega_{\mathrm{m}}} + \frac{1}{p+j\Omega}\tanh\frac{\pi(p+j\Omega)}{2}\right] \tag{6.2.48}$$

if $L\{x(t)\} = L\{\cos\Omega t\} = \frac{p}{p^2+\Omega^2}$.

Using the equations (6.2.47) and (6.2.48), we can calculate the Laplace transform of the output response for $x(t)$ represented by the harmonic signals $\sin(h\omega_{\mathrm{m}})$ and $\cos(h\omega_{\mathrm{m}})$. Since each harmonics multiplied by $m(t)$ produces an infinite series of harmonics at the output, the transfer function takes a form of the equivalent transfer matrix. Following [10], it is given in Table 6.6

In the table, the factor σ_{hk} is defined as follows:

$\sigma_{hk} = 1$ at a combination of $h = 0, 2, 4, \ldots$ and $k = 1, 3, 5, \ldots$
or at $h = 1, 3, 5, \ldots$, and $k = 0, 2, 4, \ldots$

$\sigma_{hk} = 0$ at a combination of $h = 0, 2, 4, \ldots$, and $k = 0, 2, 4, \ldots$
or at $h = 1, 3, 5, \ldots$ and $k = 1, 3, 5, \ldots$

Note, that $R_{h0} = \frac{2}{\pi}\frac{1}{h}\sigma_{h0}$ at $k = 0$.

As we mentioned earlier, an M–DM channel contains the amplification network with a limited bandwidth $K(p)$. It is essentially a low-pass filter with one or several poles. This network obviously cannot be described with a transfer function of a zero-order sample-and-hold

Table 6.6 The equivalent transfer matrix of the modulator with the square wave modulation function

Output harmonics / Input harmonics	$\sin(k\omega_{\mathrm{m}})$	$\cos(k\omega_{\mathrm{m}})$
$\sin(h\omega_{\mathrm{m}})$	0	$R_{hk} = \frac{4}{\pi}\frac{h}{h^2-k^2}\sigma_{hk}$ $(k \neq h)$
$\cos(h\omega_{\mathrm{m}})$	$T_{hk} = -\frac{4}{\pi}\frac{h}{h^2-k^2}\sigma_{hk}$	0

circuit. Therefore, the analysis of sampler-inertial network must be performed using an interpretation of the modified z-transform. A version of the modified z-transform called m-transform was adapted for solving problems similar to one discussed here. Its definition is given by the following equation

$$M[F(p)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)\frac{1}{p-s}\tanh\frac{\pi(p-s)}{2\omega_{\mathrm{m}}}\mathrm{d}s \tag{6.2.49}$$

Interested readers can find further details on the m-transform in specialised literature [11]. Here, we will only present several useful transform pairs listed in Table 6.7.

The output signal of a channel comprising a synchronous switch and a linear network $K(p)$ can be found as a product of the m-transform of the input signal and the Laplace transfer function $K(p)$, that is,

$$U(p) = M[X(p)]K(p) \tag{6.2.50}$$

Table 6.7 m-Transform pairs

$F(p)$	$M[F(p)]$
$\frac{a}{p}$	$\frac{a}{p}\tanh\frac{\pi p}{2\omega_{\mathrm{m}}}$
$\frac{a}{a+p}$	$\frac{a}{a+p}\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}}$
$\frac{a_1p+a_0}{p(p+b)}$	$\frac{1}{b}\left[\frac{a_0}{p}\tanh\frac{\pi p}{2\omega_{\mathrm{m}}} - \frac{a_0-a_1b}{p+b}\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}}\right]$
$\frac{1}{p^2}$	$\frac{1}{p^2}\tanh\frac{\pi p}{2\omega_{\mathrm{m}}} - \frac{\pi \mathrm{e}^{-\frac{\pi p}{\omega_{\mathrm{m}}}}}{\omega_m p(1+\mathrm{e}^{-\frac{\pi p}{\omega_{\mathrm{m}}}})}(1+\tanh\frac{\pi p}{2\omega_{\mathrm{m}}})$
$\frac{a_1p+a_0}{(p+b)^2+\Omega^2}$	$\frac{1}{2j\Omega}\left[\begin{array}{l}\frac{a_0-a_1b+j\Omega a_1}{p+b-j\Omega}\tanh\frac{\pi(p+b-j\Omega)}{2\omega_{\mathrm{m}}} \\ -\frac{a_0-a_1b-j\Omega a_1}{p+b+j\Omega}\tanh\frac{\pi(p+b+j\Omega)}{2\omega_{\mathrm{m}}}\end{array}\right]$
$\mathrm{e}^{-k\pi\frac{p+b}{\omega_{\mathrm{m}}}}F(p)$	$(-1)^k\mathrm{e}^{-k\pi\frac{p+b}{\omega_{\mathrm{m}}}}M[F(p)]$
$\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}}F(p)$	$c\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}}M[F(p)]$
$c\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}}F(p)$	$\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}}M[F(p)]$
$\frac{1}{(p+a)(p+b)}$	$\frac{1}{a-b}\left[\frac{1}{p+b}\tanh\frac{\pi(p+b)}{2\omega_{\mathrm{m}}} - \frac{1}{p+a}\tanh\frac{\pi(p+a)}{2\omega_{\mathrm{m}}}\right]$

The output signal of the second synchronous switch results from the m-transform of $U(p)$

$$Y(p) = M[U(p)] \tag{6.2.51}$$

All parameters of the equations (6.2.50) and (6.2.51) are defined in an equivalent block diagram of the synchronous modulation–demodulation channel shown in Figure 6.71.

Let us assume that the network $K(p)$ has one simple pole and is represented by the following Laplace transform

$$K(p) = \frac{b_v}{p + b_v} \tag{6.2.52}$$

Then, following [10], the equivalent transfer matrix of the whole M–DM channel with $K(p)$ has coefficients defined in Table 6.8.

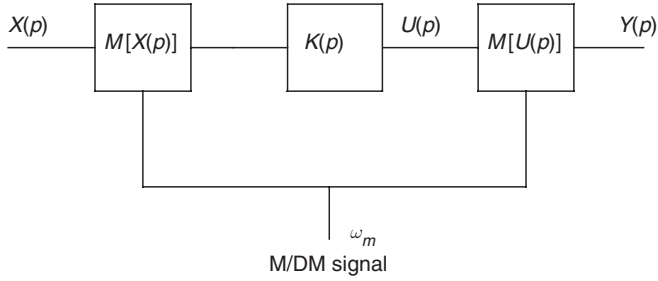


Figure 6.71 A Laplace block diagram of the synchronous modulation–demodulation channel

Table 6.8 The equivalent transfer matrix of the inertial circuit $k(p) = \frac{b_v}{p+b_v}$

Output harmonics / Input harmonics	$\sin(\mathbf{k}\omega_{\mathbf{m}})$	$\cos(\mathbf{k}\omega_{\mathbf{m}})$
$\sin(h\omega_{\mathrm{m}})$	$K_{\mathrm{dhk}}(jw) = \dfrac{A_v}{(A_v + jw)^2 + h^2} \times \left[(A_v + jw)\delta_{\mathrm{hk}} + \dfrac{4hk}{\pi} \times \dfrac{\coth\frac{\pi}{2}(A_v + jw)}{(A_v + jw)^2 + k^2}\varepsilon_{\mathrm{hk}} \right]$	$K_{\mathrm{qhk}}(jw) = \dfrac{hA_v}{(A_v + jw)^2 + h^2} \times \left[\delta_{\mathrm{hk}} - \dfrac{4(A_v + jw)}{\pi} \times \dfrac{\coth\frac{\pi}{2}(A_v + jw)}{(A_v + jw)^2 + k^2}\varepsilon_{\mathrm{hk}} \right]$
$\cos(h\omega_{\mathrm{m}})$	$K_{\mathrm{dhk}} = -\dfrac{k}{h}K_{\mathrm{qhk}}(jw)$	$K_{\mathrm{qhk}} = -\dfrac{A_v + jw}{h}K_{\mathrm{qhk}}(jw)$

The coefficient of the equivalent transfer matrix corresponding to other linear networks $K(p)$ can be found in [11] and references therein or using the equations (6.2.31–6.2.33).

Following the stability criteria, a feedback system with an M–DM channel should include a major inertial network, $T(p)$, that essentially shapes the frequency characteristic of the loop. Besides, the system unavoidably contains a number of parasitic capacitances and inductances outside the M–DM channel. Many of them introduce additional losses or phase shifts into the overall loop. The influence of parasitic components grows as the bandwidth increases until a point when they become a dominant factor causing the system instability. Parasitic components impose the principle limitations on the ultimate system performance. In the simplified model describing the propagation of the Ω-carrier only, dealing with parasitic networks is relatively straightforward. In many cases, we can evaluate simply their joint transfer function as a product (or another combination) of individual transfer functions. At the experimental stage, the function is derived from the difference between expected and actual frequency characteristics. With the more precise model that examines higher-order harmonics in addition to the Ω-carrier, the procedure becomes much more complex. We need to write down the individual equivalent transfer matrices for all non-inertial networks in the loop that are separated by M of DM units. Then their joint transfer function is derived as a product of these individual matrices. Equally, the compensation networks should be incorporated using the equivalent transfer matrices since we need to inspect the compensation effect on the Ω-carrier as well as the higher-order harmonics. The latter is particularly important when we attempt to address problems caused by circulating harmonics.

A signal at the output of the M–DM channel can be represented in the following general form:

$$\begin{aligned} y(t) &= y_1(t)\sin k\omega_m t + y_2(t)\cos k\omega_m t \\ &= y_1(t)\frac{e^{jk\omega_m t} - e^{-jk\omega_m t}}{2j} + y_2(t)\frac{e^{jk\omega_m t} + e^{-jk\omega_m t}}{2} \end{aligned} \tag{6.2.53}$$

Using a property of the Laplace transform on the frequency shift in the complex p-plane, we obtain

$$Y(p) = \frac{1}{2j}\left[Y_1(p - jk\omega_m) - Y_1(p + jk\omega_m)\right] + \frac{1}{2}\left[Y_2(p - jk\omega_m) - Y_2(p + jk\omega_m)\right]$$

The Laplace transform of the signal at the output of a network with a transfer function $Q(p)$ is

$$Z(p) = Y(p)Q(p) \tag{6.2.54}$$

The inverse Laplace transform of the equation (6.2.54) yields the output signal $z(t)$ in the time domain with coefficients at $\sin k\omega_m t$ and $\cos k\omega_m t$ summarised in Table 6.9 [12].

Now, let us consider a closed-loop system that includes an M–DM channel. The overall equivalent transfer matrix of the loop is given by

$$\vec{L}(j\omega) = \vec{K}(j\omega) \times \vec{Q}(j\omega)\beta \tag{6.2.55}$$

where $\vec{K}(j\omega)$, $\vec{Q}(j\omega)$ are the equivalent transfer matrices of the M–DM channel and the inertial network $Q(p)$ respectively, β is the non-inertial feedback circuit.

Following the equation (6.2.55), the equivalent transfer matrix of the closed-loop gain takes the form

$$\vec{W}(j\omega) = \frac{\vec{K}(j\omega)\vec{Q}(j\omega)}{I + \vec{L}(j\omega)} \tag{6.2.56}$$

In this equation, I is unity matrix of the same dimension as the loop amplification matrix.

An example of $[I + \vec{L}]$ matrix is shown in Table 6.10. This matrix takes into account the circulation of the first and second harmonics only. In the majority of cases, the transfer functions involving first three harmonics would adequately characterise the system dynamics.

Table 6.9 The equivalent transfer matrix of the inertial circuit outside the MDM unit

Output harmonics / Input harmonics	$\sin(k\omega_m)$	$\cos(k\omega_m)$
$\sin(k\omega_m)$	$\frac{Q[j\omega_m(K+k)] + Q[j\omega_m(K-k)]}{2}$	$\frac{Q[j\omega_m(K+k)] - Q[j\omega_m(K-k)]}{2j}$
$\cos(k\omega_m)$	$\frac{Q[j\omega_m(K+k)] - Q[j\omega_m(K-k)]}{2j}$	$\frac{Q[j\omega_m(K+k)] + Q[j\omega_m(K-k)]}{2}$

Table 6.10 The equivalent transfer matrix of the loop amplification $[I + \vec{L}]$

Input signal & harmonics \ Output signal & harmonics	Ω	$\sin(\omega_m t)$	$\cos(\omega_m t)$	$\sin(2\omega_m t)$	$\cos(2\omega_m t)$
Ω	$(1 + a_{11})$	a_{12}	a_{13}	a_{14}	a_{15}
$\sin(\omega_m t)$	a_{21}	$1 + a_{22}$	a_{23}	a_{24}	a_{25}
$\cos(\omega_m t)$	a_{31}	a_{32}	$1 + a_{33}$	a_{34}	a_{35}
$\sin(2\omega_m t)$	a_{41}	a_{42}	a_{43}	$1 + a_{44}$	a_{45}
$\cos(2\omega_m t)$	a_{51}	a_{52}	a_{53}	a_{54}	$1 + a_{55}$

Using usual matrix manipulation rules, we can present the equivalent transfer function of the loop in more readable form as follows

$$L(j\omega) = a_{11} + a_{12}\frac{\Delta_{12}}{\Delta_{11}} + a_{13}\frac{\Delta_{13}}{\Delta_{11}} + a_{14}\frac{\Delta_{14}}{\Delta_{11}} + a_{15}\frac{\Delta_{15}}{\Delta_{11}} \tag{6.2.57}$$

where Δ_{ij} are minors of the $\vec{L}$ matrix. Other forms of the equation (6.2.57) are possible.
It is obvious that the expression

$$L_0(j\omega) = a_{11}(j\omega) \tag{6.2.58}$$

is the transfer function of the system which would be derived with a simplified model where only the fundamental Ω-carrier of the input signal is taken into account. Or in other words,

$$a_{11}(j\omega) = \beta K Q(j\omega) \tag{6.2.59}$$

A further analysis of the function derived from the equation (6.2.59) can be done using all methods and techniques developed for the continuous linear systems. These certainly include the Nyquist polar diagram and the Bode polar plots.

The dynamic error caused by the circulation of higher-order harmonic components is defined as

$$\delta(\omega) = \frac{|L(j\omega)|}{|a_{11}(j\omega)|} - 1 \tag{6.2.60}$$

Exercise 6.8

In this exercise, we propose to calculate the frequency characteristics of a SQUID amplifier that employs the modulation–demodulation principle. The block diagram of the amplifier is shown in Figure 6.72.

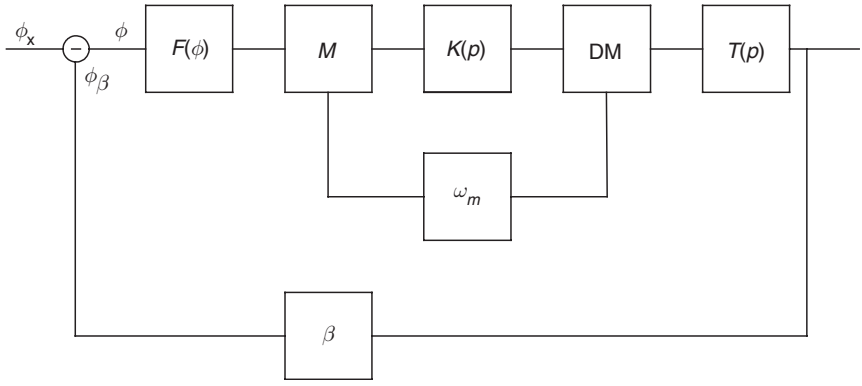


Figure 6.72 SQUID amplifier employing the modulation–demodulation read out principle

The full analysis has been done elsewhere [13]. Here, we will give the final result: the Bode plots of the system with a unity gain frequency 20 kHz, a modulation frequency 60 kHz and 3-dB bandwidth of $K(\omega)\Delta F = 600$ kHz. The Bode plots are shown in Figure 6.73. The effect of higher-order harmonics on the frequency characteristics is significant even with a choice of $\Delta F = 10 f_m$ ($f_m = \frac{\omega_m}{2\pi}$). The system

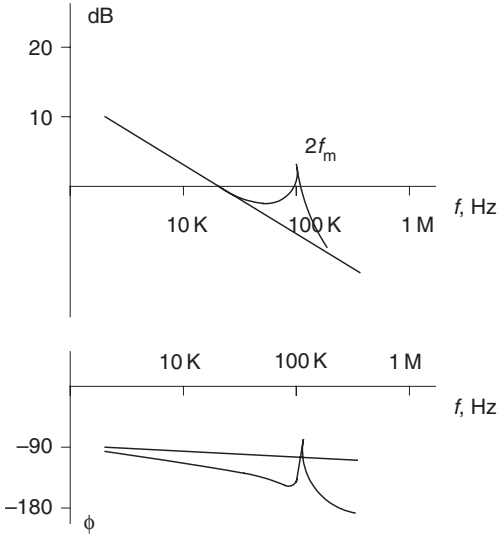


Figure 6.73 The Bode plots of the M/DM SQUID amplifier with and without taking into account the circulation of higher order harmonics produced by the modulation signal

opens a window for high-frequency noise and pick-up signals which may affect the SQUID functionality. The window becomes larger as f_m approaches ΔF.

6.3 INDUCTANCE CALCULATIONS OF THE SUPERCONDUCTING STRUCTURES

The specific inductance plays a crucial role in the superconductor electronics. We saw in Chapter 2 that it defines the frequency of parametric oscillations and the coupled energy resolution of analogue SQUIDs. A similar statement is also true with regard to the dynamics of the digital Josephson gates, the pulse propagation via connecting leads, and so on.

The Meissner effect changes the distribution of the electromagnetic field around a superconductor structure compared to the distribution that would otherwise form around the normal metal structure of the same geometry. It also restricts the supercurrent to flow on the surface of the material. Since the specific inductance (inductance per unit length) follows from localised values of the field-to-current ratios, the inductance computation models should incorporate inputs that reflect duly the properties of the superconducting materials.

Most of the inductance computation models employ the same fundamental concept presented, for instance, in [14]. According to the concept, the field-to-current distribution can be derived from the minimised free energy of the system. The free energy of a superconductor includes two major components [15]. These are:

1. the static magnetic field energy; and
2. the kinetic energy of the superelectrons.

Let us consider a system of N superconducting elements. A generic example of one element of the system is shown in Figure 6.74. Following [16], the total stored magnetostatic energy in the free space is given by

$$E_m = \frac{1}{2}\int \vec{B}\cdot\vec{H}\,d\nu = \frac{\mu_0}{2}\int \vec{H}^2\,d\nu \tag{6.3.1}$$

and the total superconducting electron energy is

$$E_k = \frac{\mu_0}{2}\int_{se} \lambda^2 \vec{J}^2\,d\nu \tag{6.3.2}$$

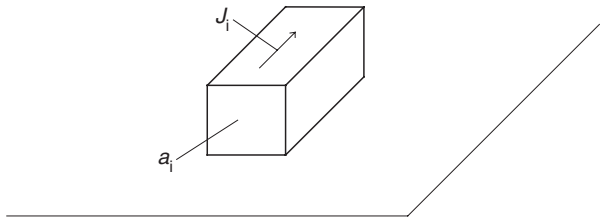


Figure 6.74 A generic example of one element of the inductance system

where λ is the penetration depth of the superconductor, μ_0 is the permeability of the free space, $\vec{B}$ and $\vec{H}$ represent the magnetic field and the magnetic field strength respectively and $\vec{J}$ is the current density distribution in the superconducting system.

The equation (6.3.1) does integration over the entire free space. Restricted to the volume of the superconducting system it takes a modified form as follows

$$E_{\mathrm{m}} = \frac{1}{2} \oint_{\mathrm{sc}} \vec{A} \cdot \vec{J} \mathrm{d}\nu \tag{6.3.3}$$

where

$$\vec{A}(\vec{r}) = \oint_{\mathrm{se}} \vec{J}(\vec{r}\,') G(\vec{r} | \vec{r}\,') d\vec{r}\,' \tag{6.3.4}$$

is the vector potential at a co-ordinate $\vec{r}$ generated by a unit current flowing at $\vec{r}'$.

$$G(\vec{r} | \vec{r}\,') = \frac{\mu_0}{4\pi} \frac{1}{|\vec{r} - \vec{r}\,'|} \tag{6.3.5}$$

is the Green function. The current is derived from the Maxwell equation

$$\vec{J} = \nabla \times \vec{H} \tag{6.3.6}$$

W. Chang [17] proposed a variational numerical technique to derive the specific inductance in superconductor structures. The finite element analysis assumes that the conductor is divided into a number of volumes. If the volumes are small enough to regard the current density within each

volume element quasiuniform, the total energy and element inductances have the following functional dependence:

$$W_{\min} = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} L_{ij} I_i I_j. \tag{6.3.7}$$

where $W_{\min}$ is the total free energy. It is derived from the minimised expression given by [17]

$$W = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_i\int_j \vec{J}_i(\vec{r}_i)\cdot\vec{J}_j(\vec{r}_j')G(\vec{r}_i|\vec{r}_j')\mathrm{d}\vec{r}_i\mathrm{d}\vec{r}_j' + \frac{\mu_0}{2}\sum_{i=1}^{N}\int_i \lambda_i^2\vec{J}_i^2(\vec{r}_i)\mathrm{d}\vec{r}_i \tag{6.3.8}$$

In what follows, we will compare two approaches in the evaluation basic parameters of planar transmission lines: a first one is conventional way usually employed with respect to normal metal structures and a second one that takes into account specific properties of superconductor materials.

6.3.0.1 Superconducting strip transmission line

Stripline is one of the most commonly used transmission lines in electronics. The transmission line is a structure that supports the transverse electromagnetic mode of wave propagation. An example of a planar transmission line is shown in Figure 6.75. The line has the width w, thickness t and separated from the ground plane by a distance of h. For $w > h$, the characteristic impedance

$$Z_0 \approx \sqrt{\frac{L}{C}} \tag{6.3.9}$$

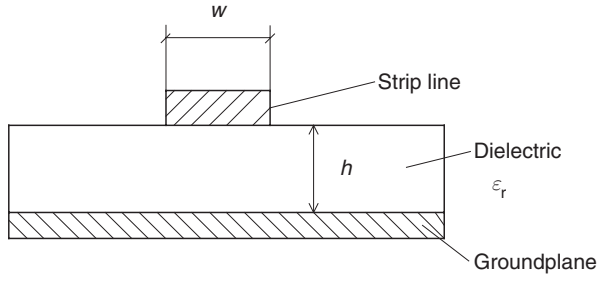


Figure 6.75 An example of a planar transmission line

can be expressed via the geometric factors by electrostatic analysis. Using the conformal mapping the characteristic impedance is given by [18]

$$Z_0 = \frac{1}{c}\sqrt{CC_0} \tag{6.3.10}$$

with $c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$, ε_0 is the permittivity of the free space.

The capacitance per unit length can be found from the following expression

$$C = \frac{\varepsilon_r \varepsilon_0 w}{h} \tag{6.3.11}$$

where ε_r is the dielectric constant of the insulation separating the strip from the ground plane.

The configuration, shown in Figure 6.75, is simple enough to derive an analytical expression for the specific inductance of the superconducting line lying above a superconducting ground plane. In real terms, the total energy can be represented in the following form:

$$W = \frac{1}{2}L_m I^2 + \frac{1}{2}L_k I^2 \tag{6.3.12}$$

In this equation, L_m is the geometrical inductance, whereas L_k is the kinetic inductance that takes into account the inertial mass of the charge carriers. For a uniform current in a homogeneous conductor the kinetic inductance is

$$L_k = \frac{m_q}{nq^2}\frac{l}{a} \tag{6.3.13}$$

where n is the number of carriers, m_q, q represent their mass and charge respectively. The equation (6.3.13) shows that the kinetic inductance depends on the geometry of the conductor through the length, l, and the cross-sectional area, a. The geometric inductance L_m normally dominates over L_k. Yet, with a ground plane in place, L_m can be reduced to a level when L_m becomes much smaller than L_k. The inductance of a superconducting stripline is given by [19]

$$L = \frac{\mu_0 l}{w}\left\{h + \lambda\left[\coth\left(\frac{t}{\lambda}\right)\right]\right\} \tag{6.3.14}$$

In the equation (6.3.14), the thickness of the groundplane is assumed to be much larger than the penetration depth. The kinetic inductance contributes via the temperature-dependent penetration depth

$$\lambda(T) = \frac{\lambda(0)}{\sqrt{1 + \left(\frac{T}{T_c}\right)^4}} \tag{6.3.15}$$

where $\lambda(0)$, T_c refer to the penetration depth at absolute zero and the critical transition temperature of the superconducting material respectively.

Using the minimised total free energy, W. Chang enhanced the equation (6.3.14) [20]:

$$L = \frac{\mu_0 l}{wK(w,h,t)}\left\{h + \lambda_1\left[\coth\left(\frac{t}{\lambda_1}\right) + \frac{2p^{\frac{1}{2}}}{r_b}\ \text{cosech}\ \left(\frac{t}{\lambda_1}\right)\right]\right\} \tag{6.3.16}$$

The equation (6.3.16) yields an accurate inductance values for the aspect ratio $\frac{w}{h} > 1$. In this case, the fringe field becomes significant. It is taken into account by the fringe field factor defined as

$$K(w,h,t) = \frac{h}{w}\frac{2}{\pi}\ln\frac{2r_b}{r_a} \tag{6.3.17}$$

where

$$\ln r_a = -1 - \frac{\pi W}{2h} - \frac{p+1}{p^{\frac{1}{2}}}\tanh^{-1}\left(p^{-\frac{1}{2}}\right) - \ln\left(\frac{p-1}{4p}\right),$$

$$r_b = r_{b0} \text{ for } \frac{w}{h} \geq 5,$$

$$r_b = r_{b0} - [(r_{b0}-1)(r_{b0}-p)]^{\frac{1}{2}} + (p+1)\tanh^{-1}\left(\frac{r_{b0}-p}{r_{b0}-1}\right)^{\frac{1}{2}}$$

$$- 2p^{\frac{1}{2}}\tanh^{-1}\left(\frac{r_{b0}-p}{p(r_{b0}-1)}\right)^{\frac{1}{2}} + \frac{\pi w}{2h}p^{\frac{1}{2}} \quad \text{for} \quad 1 \leq \frac{w}{h} \leq 5$$

$$r_{b0} = \eta + \frac{p+1}{2}\ln\Delta, \quad \Delta = \text{larger value of } \eta \text{ or p},$$

$$\eta = p^{\frac{1}{2}}\left\{\frac{\pi w}{2h} + \frac{p+1}{2p^{\frac{1}{2}}}\left[1 + \ln\left(\frac{4}{p-1}\right)\right] - 2\tanh^{-1}p^{\frac{1}{2}}\right\},$$

$$p = 2\beta^2 - 1 + \left[(2\beta^2 - 1)^2 - 1\right]^{\frac{1}{2}},$$
$$\beta = 1 + \frac{t}{h}$$

The capacitance is also affected by the fringe field effect, that is,

$$C = \frac{\varepsilon_r \varepsilon_0 wl}{h} K(w, h, t) \tag{6.3.18}$$

The maximum operation frequency of a microstrip line, f_T, is the same for normal metals and superconductors. It is defined at a point at which significant coupling occurs between the transverse electromagnetic mode and the lowest-order surface-wave spurious mode. According to [21] ,

$$f_T = \frac{150}{\pi h} \sqrt{\frac{2}{\varepsilon_r - 1}} \tan^{-1} \varepsilon_r \tag{6.3.19}$$

where f_T is in GHz and h is in millimetres.

The propagation speed is a function of the geometry, the dielectric constant and the superconducting material. It was estimated to be about 100 μm/ps for Nb-strip line at w above 0.5 μm, $h = 150$ nm and a dielectric constant $\varepsilon = 4$ characteristic of SiO_2 [22] .

6.3.0.2 Inductance of a single line

The inductance of a single thin film line is given by

$$L_{sl} = \frac{2}{\pi} \mu_0 l \left[\ln\left(\frac{l}{w}\right) + 0.5\right] \tag{6.3.20}$$

It also can be utilised to evaluate the inductance of the shunt resistors deposited directly on the substrate (no ground plane), as shown in Figure 6.76(a). Such an arrangement has a clear advantage in terms of the power dissipation from the shunt directly to the substrate in addition to the cooling fins. The maximum length of a 2 μm ungrounded shunt cannot exceed approximately 10 μm for a SQUID operating at a plasma frequency of 100 GHz. The necessary resistance can be trimmed by thinning the shunt. This, however, may create a problem for the step coverage from the SQUID washer to the substrate. Longer shunts need to be groundplaned as depicted in Figure 6.76(b). Since the superconductor

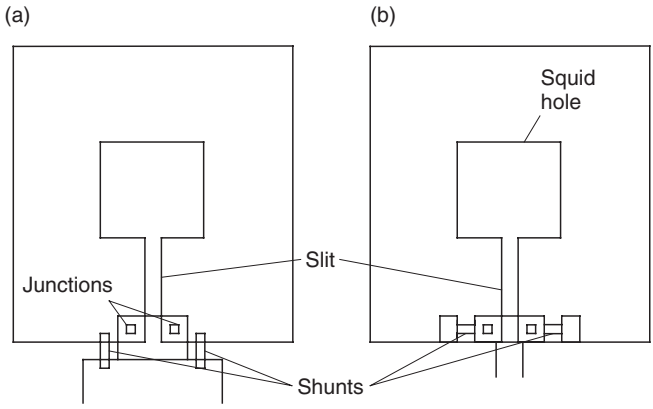


Figure 6.76 A Schematic representation of SQUID configurations with ungrounded shunts (a) and groundplaned shunts (b)

has very poor thermoconductivity, the attachment of designed cooling fins is the "must" option for this configuration to avoid extra thermal noise generated by shunts.

Table 6.11 summarises the calculation of SQUID inductances and an associated input coil using the MathCad software package. It also illustrates how these inductances influence the main parameters of the SQUID.

Table 6.11 The MathCad calculations of main parameters of a SQUID with an input coil

The SQUID inductance required

$$L = 25 \times 10^{-12}$$
$$\pi = 3.14$$
$$\mu_0 = 4 \times \pi \times 10^{-7} \quad \mu_0 = 12.56 \times 10^{-7}$$

The outer dimension of the SQUID hole

$$d = \frac{L}{1.25 \times \mu_0} \quad d = 1.592 \times 10^{-5}$$

The length of the unground slit of the washer

$$w = 5 \times 10^{-6}$$

The inductance of the unground slit

$$L_{s1} = 0.3 \times 10^{-6} \times w \quad L_{s1} = 1.5 \times 10^{-12}$$

The interconnection inductance from washer to Josephson junctions

$$L_j = 5 \times 10^{-12}$$

The total SQUID inductance

$$L_s = L + L_{s1} + L_j \quad L_s = 3.15 \times 10^{-11}$$

The flux quantum

$$F_0 = 2.07 \times 10^{-15}$$

The main parameter of the SQUID

$$\beta_l = 1.0$$

The critical current of the Josephson Junction

$$I_0 = \beta_l \frac{F_0}{2 \times L_s} \quad I_0 = 3.286 \times 10^{-5}$$

TRANSFER FUNCTION AND ENERGY SENSITIVITY

The Boltzmann constant

$$k_b = 1.38 \times 10^{-23}$$

The temperature

$$T = 4.2$$

The dimension of the Josephson junction

$$l_j = 3 \times 10^{-6}$$

The capacitance of the Josephson junction for $j_c = 600\,\text{A/cm}^2$

$$C = 0.05 \times l_j^2 \quad C = 4.5 \times 10^{-13}$$

The main parameter of the Josephson junction

$$\beta_c = \frac{2 \times \pi \times l_0 \times RRC}{F_0}$$

$$\beta_c = 0.7$$

The resistance of the shunt

$$R = \sqrt{\beta_c \times \frac{F_0}{2 \times \pi \times I_0 \times C}} \quad R = 3.95$$

The damping resistance

$$R_d = 2R \quad R_d = 7.901$$

The noise parameter

$$\gamma = 2\pi \times \mathrm{kb} \times \frac{\mathrm{T}}{\mathrm{I}_0 \times \mathrm{F}_0} \quad \gamma = 5.352 \times 10^{-3}$$

Energy resolution $\left(\frac{\mathrm{J}}{\mathrm{Hz}}\right)$ for $\beta_l = \beta_c = 1$, otherwise use $e = 9\,\mathrm{kb}\,\mathrm{TL/R}$

$$e = 16 \times kb \times T \times \frac{\sqrt{L_s C}}{1.0 \times 10^{-34}} \quad l = 34.915$$

$$e_1 = 9 \times kb \times T \times \frac{L_s}{R \times 1.0 \times 10^{-34}} \quad e_1 = 41.596$$

Experimentally measured flux noise, $\frac{F_0}{\sqrt{\mathrm{Hz}}}$

$$ff = 0.8 \times 10^{-6}$$

Experimental energy resolution $\left(\frac{\mathrm{J}}{\mathrm{Hz}}\right)$

$$e_2 = \frac{ff^2 \times F_0^2}{2 \times L_s \times 10^{-34}} \quad e_2 = 435.291$$

The reduction of the swing of the flux-to-voltage transfer function

$$\alpha = 1 - \sqrt{\beta_l \times \gamma} \quad \alpha = 0.927$$

The effective value of the main parameter of the SQUID due to the damping resistor $R_d = R$

$$\beta_{\mathrm{eff}} = \frac{\beta_l}{\sqrt{1 + \beta_l^2}} \quad \beta_{\mathrm{eff}} = 0.707$$

The transfer function with damping resistor

$$V_f = 4 \times \frac{\alpha}{(1 + \beta_{\mathrm{eff}})} \times 10 \times \frac{R}{F_0} \quad V_f = 1.362 \times 10^{11}$$

$$V_f = 2 \times \alpha \times \frac{\beta_l}{1 + \beta_{\mathrm{eff}}} \times \frac{R}{L} \quad V_f = 1.71 \times 10^{11}$$

The peak-to-peak voltage swing

$$d_V = \frac{\alpha}{\pi} \times \frac{4-\alpha}{1+\beta_{eff}} \times I_0 \times R \quad d_V = 6.897 \times 10^{-5} \quad d_V \times 2 \times 10^6 = 137.941$$

The area of the DC squid

$$A = d^2 + 5 \times 10^{-6} \times w$$

The field resolution

$$S_b = \frac{\sqrt{2 \times L \times e_1 \times 1.0 \times 10^{-34}}}{A} \quad S_b = 1.637 \times 10^{-12}\,T$$

The intrinsic magnetic flux resolution

$$S_f = \frac{\sqrt{e_1 \times 1.0 \times 10^{-34} \times 2 \times L_s}}{2.07 \times 10^{-15}} \quad S_f = 2.473 \times 10^{-7}$$

Parasitic capacitance

$$C_p = 1.0 \times 10^{-12}$$

The dynamic resistance

$$R_{dyn} = V_f \times \sqrt{2} \times L_s \quad R_{dyn} = 7.644$$

The magnetic flux resolution

$$S_{f1} = \frac{\left[2 \times k_b \times T \times \left(\frac{L^2}{R} + 2 \times \frac{R_{dyn}}{V_f^2}\right) \times \left(1 + 2 \times \frac{C_p}{C}\right)^{\frac{1}{2}}\right]^{\frac{1}{2}}}{F_0} \quad S_{f1} = 2.068 \times 10^{-7}$$

Number of turn in input coil

$$n = 6$$

Inductance of the input coil

$$L_i = n^2 \left(L + \frac{L_{s1}}{3}\right) \times (1 + 0.05)$$
$$L_i = 9.639 \times 10^{-10}$$

The mutual inductance between input coil and washer

$$M = n \times (L + \frac{L_{s1}}{2}) \quad M = 1.545 \times 10^{-10}$$

The coupling factor between coil and washer

$$k = \frac{M}{\sqrt{L_s L_i}} \quad k = 0.887$$

The coupled energy resolution of SQUID (h)

$$e_c = \left(\frac{e_1}{k^2}\right) \quad e_c = 52.91$$

REFERENCES

[1] W.H. Chang. *IEEE Trans. Magn.* **17**, 764 (1981).
[2] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. Academic press, London, 1994.
[3] R. Sallen and E. Key. *IRE Trans. Circuit Theory*, **CT-2**, 78 (1955).
[4] A. Williams and F. Taylor. *Electronic Filter Design Handbook*. McGraw-Hill, 1988, pp. 3–16.
[5] *Wiley Encyclopedia of Electrical and Electronic Engineering*. J.G. Webster, ed., Wiley, 1999, vol. 7, p. 450.
[6] M. Van Valkenburg. *Analog Filter Design*. Holt Rinehart and Wilson, London, 1987.
[7] G. Tomlinson. *Electrical Networks and Filters: Theory and Design*. Prentice-Hall, NY, 1991, p. 122.
[8] A. Zverev. *Handbook of Filter Synthesis*. NY, Wiley, 1967.
[9] K. Ogata. *Modern Control Engineering*. Prentice-Hall, 1970, p. 628, Schaum, p. 134.
[10] V. Gostev. *Design of Control Systems with Synchronous Multipliers*. Kiev, Technika, 1973.
[11] V. Gostev. *Automatika I telemehanika* (Rus.), Nos 9, 11 (1967).
[12] V. Gostev. *Analysis of Closed Control Systems with Synchronous M–DM Channels Using the Method of Balanced Harmonic Components*(Rus.), In "*Technicheskaja Kibernetika*", vol. 10, Ukranian Academy of Science, 1970.
[13] V. Polushkin. *Izmeritel'naja Technika* (Rus.), No. 6, p. 47 (1989).
[14] A. Kalantarov. *Calculation of inductances* (Rus). Nauka, Moscow, 1970.
[15] N.H. Meyer, *Proc. IRE*, 49, 1640 (1961).
[16] A.R. Sass and W.C. Stewart. *J. Appl. Phys*., 39, 1956 (1968).
[17] W.H. Chang. *IEEE Trans. Magn*., 17, 764 (1981).
[18] K.C. Gupta and A. Singh, eds, *Microwave Integrated Circuits*, Wiley Eastern, New Dehli, India, 1974.
[19] N.M. Meyers, *Proc. IRE*, 49, 1640 (1961).
[20] W.H. Chang. *J. Appl. Phys*. 50, 8129 (1879).
[21] I.J. Bahl and D.K. Triverdi, *Microwaves*, 16, 174 (1977).
[22] M.B. Ketchen. *IEEE Trans. Magn*., 27, 2916 (1991).

Index

Nuclear Electronics: Superconducting Detectors and Processing Techniques V. Polushkin

ISBN: 0-470-85759-5