High Accuracy Electronics

(at room temperature)

Faraday's toroidal transformer – the first on record (Picture courtesy of IET)

A collection of monographs on inductive voltage dividers, ratio transformers and related circuits plus some nice maths

Version 1.0 (May 2020)

Christopher I. Daykin MA

Acknowledgements

For Di, with sincere thanks, in memory of John David Yewen FRAS (JDY): friend and former colleague. “Hairy” John taught me the art of engineering and deserves most of the credit for the ingenious designs herein.

Thanks, also, to former ASL colleague: Peter Caleb Frederick Wolfendale FIEE [1].

Other contributions came from: -

- Don White [2] for extensive and valuable input on high accuracy resistors and guidance on publishing in general.
- Paul Bramley [3] for feedback on the monograph: The Isotech MicroK “Bridge”.
- Professor Stephen Donnelly [4], for encouraging me to publish “for the benefit of mankind”.

Thank you, also, Isothermal Technology Ltd and WIKA Alexander Wiegand SE & Co. KG for your permission to include certain images.

Thanks guys!

More feedback would be appreciated.

1. Inspirational founder and former Managing Director of Automatic Systems Laboratories Ltd.
2. Measurement Standards Laboratory of New Zealand and co-inventor (with Keith Jones) of the resistance bridge calibrator (RBC) - one of the most significant contributions to this field in recent years (Patent: PCT/NZ95/00022). See part 1, monograph 1, section 4.3 and 4.3.2
3. Metrosol Ltd and Project Manager of the MicroK Bridge design team.
4. University of Huddersfield and a genuine scholar.

Preface

"High Accuracy" in the title refers to measurement uncertainty in the range of ten parts per million (10ppm) to ten parts per billion (10ppb). The subtitle "at Room Temperature" emphasises the fact that this level of accuracy is possible with readily available components, materials and tools. Whereas greater accuracy can be achieved at very low (cryogenic) temperature, such techniques are very expensive. A bridge based on the (superconducting) cryogenic current comparator [1], for example, exploits the Meissner effect and uses a superconducting quantum interference detector (SQUID) [2] to measure the ratio of resistors to better than 10ppb.

This collection is aimed, primarily, at students, teachers and researchers in the physical sciences. Some of the content could, for example, form the basis of a module at first degree (final year) level in electronic engineering. More advanced topics are suited to a Masters level course in metrology or instrumentation. Researchers could apply some of the techniques and practical tips to construct cutting edge instrumentation at minimal cost.

Part 1 introduces the concepts of accurate lumped impedance: resistors and capacitors at low frequency. Three-terminal capacitors, for example, can be constructed for which the AC version of Ohm's law is accurate to within a few parts per billion. The theory is underpinned by Maxwell's field equations, surely one of the greatest scientific achievements of the nineteenth century, if not of all time.

The fundamental mechanism that gives rise to the concept of resistance, on the other hand, is much more complicated. Fortunately we have decades of technological development and empirical evidence, at low frequency, to underpin our faith in Ohm's law, at least to the level of 10ppb, for many types of materials and components.

Part 2 also deals with (mainly) passive components – in circuits to form high accuracy filters. These are not only useful but also instructive - the models developed and "nice" matrix maths is applied in later monographs (for the analysis of two and three-stage inductive voltage dividers and ratio transformers).

Parts 3 and 4 are the core of the collection: inductive voltage dividers (IVDs), ratio transformers (RTs) and related accurate circuits, especially the "inside-out" high accuracy voltage follower (HAVF). A review of basic principles is provided for students but the advanced specialist may also find some useful ideas which have remained unpublished for over 35 years.

Parts 5 and 6 are concerned with other circuits which, despite not being inherently accurate, are indispensible. A well matched low noise pre-amplifier and filtering for potentially large amounts of interference are, for example, essential ingredients of a good null detector. Included in the miscellany are a couple of unusual circuits which may not be found in any text book or application note: A simulated large capacitor and a simulated negative capacitor are further examples of JDY's ingenuity.

Most of the content should be uncontroversial but there are a few notable exceptions. This author "begs to differ" with the generally accepted definition of a four-terminal-pair lumped impedance [3]. An alternative is proposed.

Also, the term "Bridge" as in [4] is debatable; The MicroK "Bridge" is based on a high accuracy analogue to digital converter (ADC) with 10ppb linearity. An attempt is made to explain how this incredible performance is achieved.

Similarly, the HP3458A digital multi-meter (DMM) is based on a proprietary "multi-slope" ADC technique. Much of its workings remain "trade secret". The theory here is based on scraps of published information [5].

Some of the content is speculative – novel but untested designs. The interested reader is encouraged to contact the author (email: hiaccelectronics@gmail.com).

1. Seppä, H. and Satrapinski, A.: "AC Resistance Bridge Based on the Cryogenic Current Comparator". IEEE Trans. on Inst. & Meas., Vol. 46, No 2, April 1997.
2. A SQUID can detect extremely low levels of magnetic flux density (typically $\approx 3aT/\sqrt{Hz}$).
3. Part 1, Monograph 1, section 2.1.
4. Part 4, Monograph 4: The Isotech MicroK "Bridge".
5. Part 4. Monograph 5: The 3458A general purpose digital multi-meter.

CONTENTS

Part 1: High accuracy passive components

Monograph 5: Variable gap capacitive displacement transducers

Monograph 6: Low phase error capacitors and inductors

Part 2: High Accuracy Filters

Monograph 1: Two-stage filters

Part 3: Inductive Voltage Dividers and Ratio Transformers

Monograph 2: Single-stage inductors and transformers

Monograph 3: Two-stage IVDs and RTs

Work in progress: -
Monograph: 9_Coaxial AC bridges
Monograph: 10_A quadrature bridge

Part 4: High accuracy active circuits

Monograph 3: High accuracy amplifiers, integrators and differentiators

Monograph 4: The Isotech microK "bridge"

work in progress: -

Monograph 5: Multislope ADCs

Part 5: Null detector circuits

Monograph 1: Null detectors – the basics

Monograph 2: Low noise BJT pre-amplifiers

Monograph 4: JFET theory

Part 6: Miscellaneous circuits

Monograph 1: A simulated large capacitor circuit

Monograph 2: A simulated negative capacitor circuit

Monograph 3: A circuit for measuring tanδ.

High accuracy resistors

1. Introduction

The world's most accurate resistors are based on the (cryogenic) quantum Hall effect and beyond the scope of this collection. Fortunately it is possible to construct practical (room temperature) resistors and resistive sensors which obey Ohm's law (within limits) with an accuracy approaching 1ppb. For the highest accuracy, however, operating temperature must be controlled (including self-heating) to within ±0.01°C or better.

High accuracy resistors are constructed in three principal ways, depending on the resistor value and the length of the connecting leads. The main issues are resistance and capacitance of the connecting leads, series inductance of the resistor and the importance of phase error for the particular application: -

The three types are: -

1. Single-terminal-pair (1TP, sometimes referred to as "two-terminal-pair" resistors [1]).
2. Two-terminal-pair connection (2TP, sometimes referred to as "four-wire" or "Kelvin" connection).
3. Four-terminal-pair connection (4TP).

The first method is appropriate for high value resistors where the lead resistance is negligible. Parallel capacitance and electrical interference may be a problem in which case the resistor and connections need to be screened. The main application is transfer standards and high accuracy circuits (e.g. amplifiers, integrators and oscillators).

The second method is appropriate for low to medium value resistors where the lead resistance is significant. One pair delivers the current while the other pair is used to measure the voltage. Bridge techniques can distinguish between the in-phase and quadrature (real and imaginary) components of impedance and a small phase error (e.g. due to parallel capacitance of the leads or series inductance) is not usually a problem. Clearly, however, the size of the problem increases with frequency and high accuracy resistors and measurement techniques are designed to operate at low (or very low) frequency (sinewave or alternating DC). Very low frequency measurement is severely limited by thermal emfs and 1/f noise and so low frequency (10Hz – 1kHz) sine wave methods remain popular. The main applications are transfer standards and resistance thermometry.

Fig. 1.1 The "Wilkins" type transfer standard 2TP resistor
(designed for immersion in a temperature controlled bath of mineral oil - courtesy ASL U.S.)

The third type is the ultimate method, for the very highest accuracy, reducing the effect of lead capacitance. The main application is for medium to high value transfer standards.

1. Awan, S., Kibble, B., and Schurr, J: "Coaxial Circuits for Interference-free Measurements" Electrical Measurements Series 13 published by the IET. www.theiet.org. For a different perspective see chapter 5 "General Principles of Accurate Impedance Measurement". For "two-terminal–pair" see section 5.3.6.

2. Single-terminal-pair plus screen

If the resistance value is high, compared to the connecting leads, then a single pair of connections is all that is required. Interference from electric and magnetic fields can be an issue but a screen around the resistor (fig. 2.1a) and coaxial connections provide a simple solution (see, also, fig.2.4). It can easily be arranged that the screen/outer conductor is at local 0V, at least approximately, with low impedance to "earth" (fig. 2.1a). The screen can also provide the route for the return current in order to minimise external magnetic flux - the electric and magnetic fields are contained within the coax cable (fig. 2.1b).

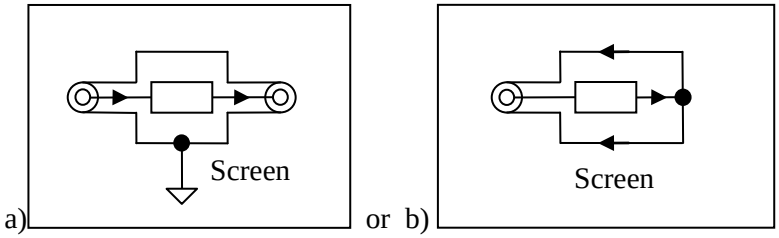


Fig. 2.1 A screened resistor

Commercially available miniature component resistors, while not quite matching the long term stability of the Wilkins type, have excellent temperature stability and AC characteristics. Single resistors, matched pairs and networks with an initial tolerance of ±0.01% and temperature coefficients <1ppm/°C (20-30°C) are readily available.

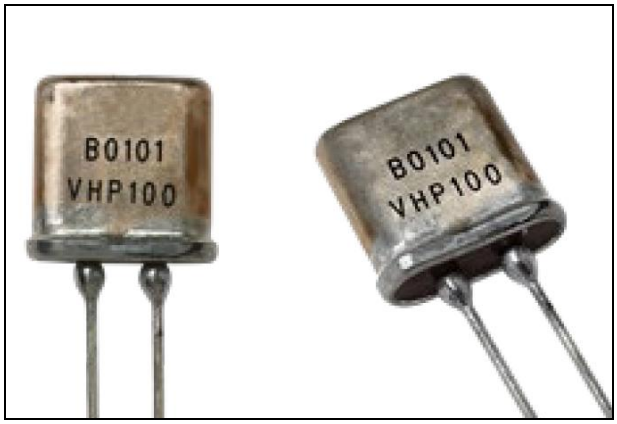


Fig. 2.2 High accuracy component resistors (picture courtesy Vishay Precision Group, Inc)

With pairs, triples and networks the accuracy and stability of ratio can be even better: -

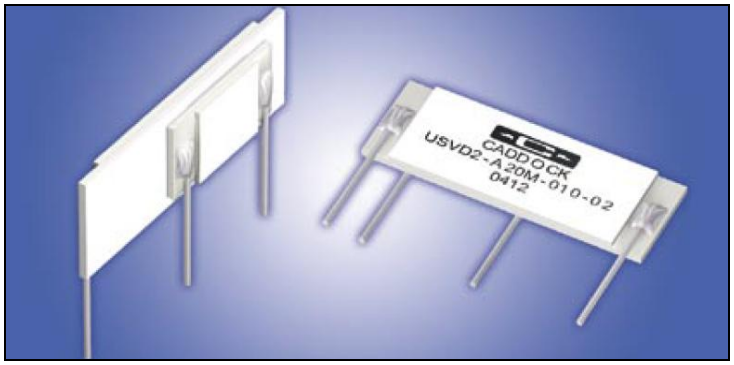


Fig. 2.3 High accuracy 100:1 ratio resistor pair (courtesy Caddock Inc.)

The main applications are transfer standards and accurate amplifiers and integrators [1].

1. Part 4, Monograph 3: "High accuracy amplifiers, integrators and differentiators".

One simple method for eliminating the effect of interference and stray capacitance (across the resistor) is to connect one terminal to a virtual earth with the screen connected to 0V. An overall 0V screen around the high gain block front-end may also be necessary: -

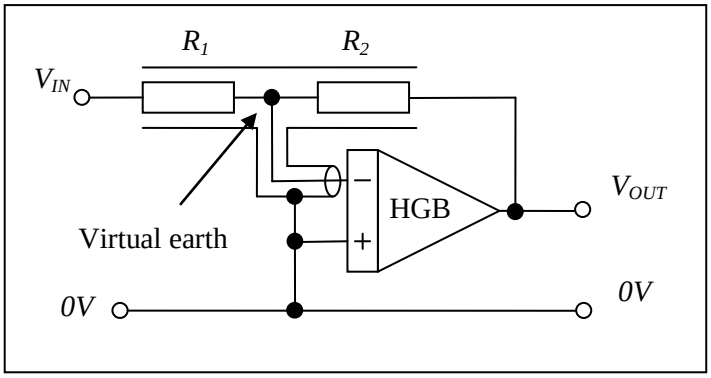


Fig. 2.4 High accuracy inverting amplifier with screened resistors [1]

If the resistor is at the end of a pair of coaxial cables the outer conductors (screens) should not be connected together at the resistor end. If they are connected the current flowing through C_1 (signal input side) can develop a small voltage drop down the screen resistances (R_1 and R_2 in parallel) which then injects an in-phase current* into the virtual earth via the cable capacitance C_2: -

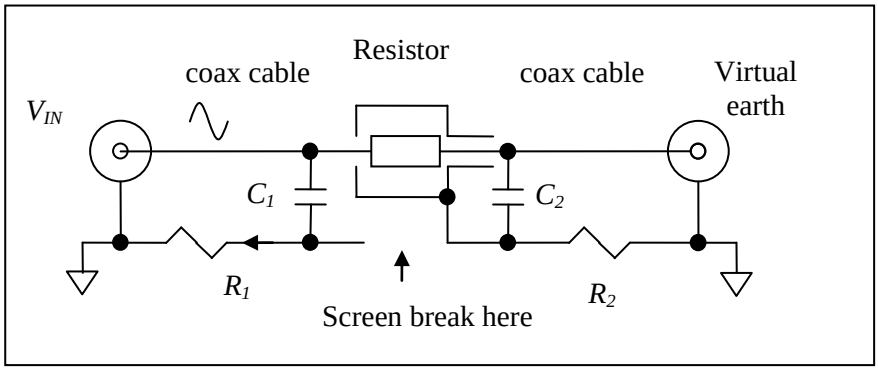


Fig. 2.5 Simplified equivalent circuit of a remote resistor

Magnetic interference can be kept to a minimum by routing a pair of coaxial cables together as a twisted pair. Any stray AC magnetic flux passing through a loop formed by the cables induces a voltage which is added in series with the resistor. Thin and flexible coax cable with a less than perfect screen (c.f. earphone cables) may be better, in practice, than the more usual kind of high quality (less flexible) type. The twisted pair could always be routed through a flexible metal conduit, connected to local 0V, if electric field interference is a particular problem.

*Such a current can also be eliminated with a high accuracy voltage follower/active guard [2]: -

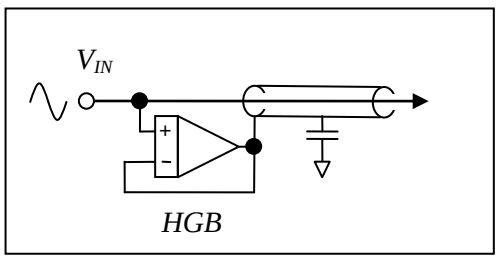


Fig. 2.6 An active guard circuit

(Caution – this method involves an amount of positive feedback, depending on V_{IN} source resistance)

1. Part 4, monograph 3: “High accuracy amplifiers, integrators and differentiators”.
2. Part 4, monograph 2: “High accuracy voltage followers (HAVFs)”. See section 1.

2.1 Converting a single to a four-terminal-pair resistor

For the very highest accuracy some experts advise that the effect of lead resistance can be significantly reduced by adding coaxial T-connectors physically close to the device [1]. This is then "regarded" as a four-terminal-pair resistor. The result is described as a complex network (fig. 2.1.2), conforming to the generally accepted definition, though they do not include the resistance of the screen in the diagram. This must be included, therefore, in the defining parameter *Z* (see section 5).

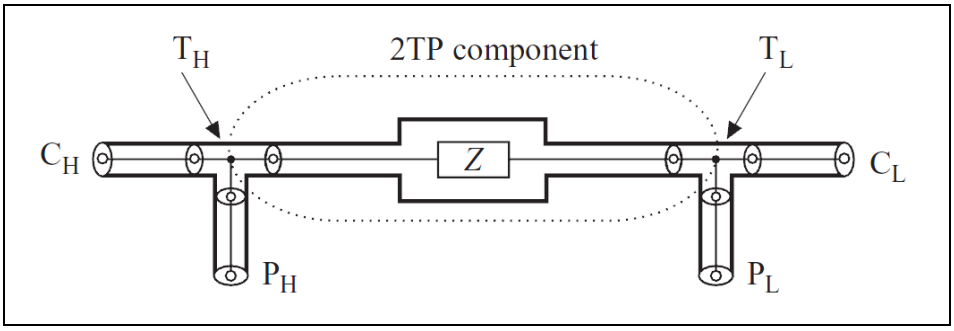


Fig. 2.1.1 Converting a 1TP plus screen into a 4TP (courtesy Awan et al)

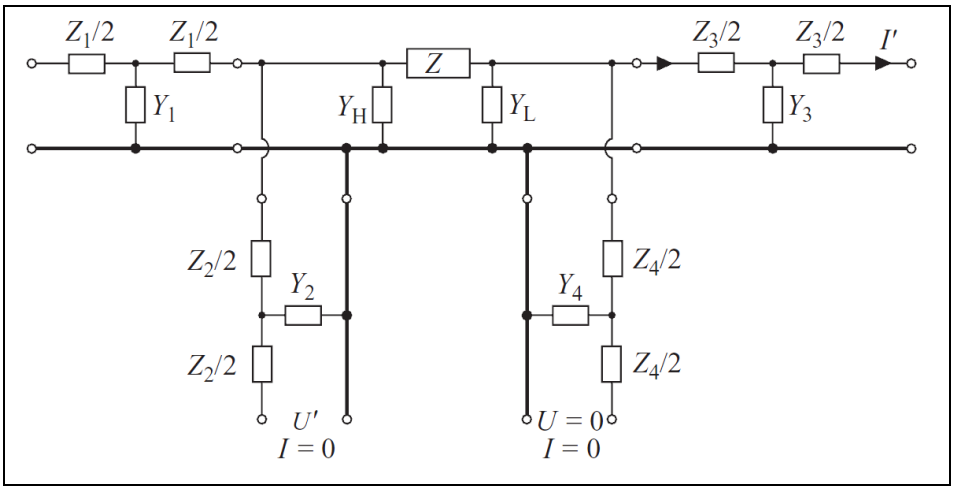


Fig. 2.1.2 Equivalent circuit of fig. 2.1.1 (courtesy Awan et al)

This author begs to differ!
1. The device in fig. 2.1.1 is, in effect, a two-terminal-pair resistor with screen (see the next section).
2. It is far better to break the screen connection thus avoiding the problem cited in the previous section. The return current is best routed via another conductor.
3. If it is unavoidable that the screen must carry the return current then it is important that the screen/connector resistances are as small as possible.
4. If the manufacturers don't know any better the experts should advise them accordingly.

A more satisfactory definition of a four-terminal-pair is offered in section 5.

This topic and the issues of "earth loops" and "chokes" are explored in the monograph "Coax AC bridges" [2]

1. According to Awan et al (see footnote page 1): "Many standard impedances made and sold for measurements of the highest accuracy have two-terminal-pair terminations, and no information is available for the length or impedance and shunt admittance of the internal cables between the impedance itself and these terminations. To improve their electrical definition, T-connectors can be added at the terminations. The impedance can then be regarded as being defined as a four terminal pair having its internal defining points at these T-connectors."
2. Part 3, monograph 9: "Coaxial AC bridges".

3. Two-terminal-pair resistors

Most applications are for low to medium resistance measurement (1 - 1000Ω). Lead resistance is not negligible and so an extra pair of voltage sensing connections is required. The effect of resistance in the current carrying pair is eliminated as long as the current flowing in the voltage sensing leads is negligible. In most applications the effect of resistance in the connecting leads is insignificant and so they are not usually shown: -

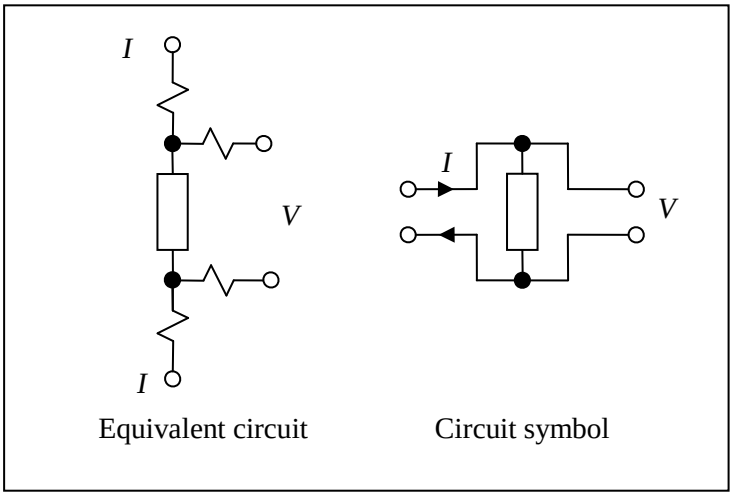


Fig. 3.1 2TP connections

For high accuracy each pair needs to be co-axial or tightly twisted. This ensures that the magnetic flux transmitted from the current carrying pair and the voltage induced in the voltage sensing pair by any stray flux are kept to a minimum.

A typical bridge configuration is to use some kind of multiplier with sufficiently high input impedance (e.g. an actively energised ratio transformer). The null balance ensures that no current flows in the voltage connections from R_T: -

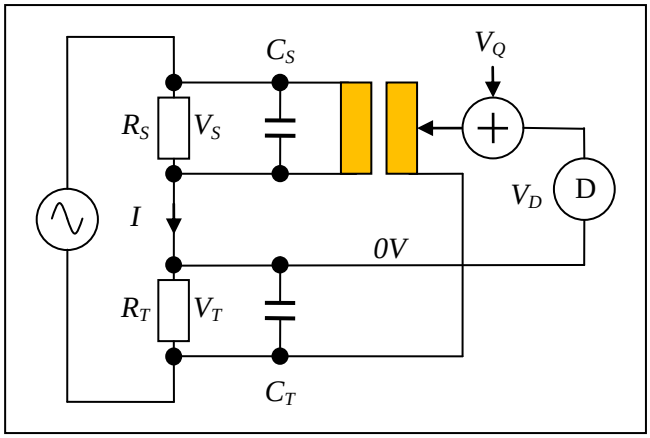


Fig. 3.2 A typical bridge configuration for 2TP resistors

The main issue with two-terminal-pair resistors is the effect of parallel capacitance. One metre of coaxial or twisted pair cable typically represents 100pF of capacitance. The two pairs act in parallel and so the total capacitance is roughly 200pF per metre of connecting cable.

The real part (in-phase) is balanced by the main ratio device – usually a very accurate ratio transformer or similar. The imaginary part (quadrature) is relatively small and balanced by a quadrature servo [1].

1. Part 5, monograph 1: “Null detectors – the basics”.

3.1 Analysis of parallel capacitance with quadrature balance

The quadrature servo adds a signal to the output of the ratio transformer so that a null is achieved to both in-phase and quadrature components: -

$$V_D = aV_S - V_T + V_Q = 0$$

$a = \frac{N_S}{N_P}$ is the transformer ratio setting. The phase error of a ratio transformer is usually much smaller than that due to the parallel capacitance and is simply added to it – equivalent to a small change in the capacitance. The second order term (in-phase error) is negligible. One can assume, therefore, that a is a real number.

It is shown elsewhere [1] that a practical approach is to derive the quadrature servo signal from V_S: -

$$V_Q \approx jbV_S \quad \Rightarrow \quad V_D = (a + jb)V_S - V_T$$

$$V_D = 0 \;\Rightarrow\; \frac{V_T}{V_S} = a + jb$$

Define parameters: $\theta_S = \omega_C R_S C_S$ and $\theta_T = \omega_C R_T C_T$. Both are small (typically <0.1mrad or 100ppm).

With a little algebra: $a + jb = \frac{R_T}{R_S} \times \frac{1 + j\theta_S}{1 + j\theta_T} \;\Rightarrow\; a = \frac{R_T}{R_S}\left(\frac{1 + \theta_S\theta_T}{1 + \theta_T^2}\right)$ and $b = \frac{R_T}{R_S}\left(\frac{\theta_S - \theta_T}{1 + \theta_T^2}\right)$

Values for the quadrature components are typically <100ppm so that the effect on accuracy can be reduced to less than 10ppb.

For the very highest accuracy it is quite simple to add capacitance or, better still, matching cable to one side or the other so that the quadrature imbalance is reduced to almost zero. To a very good approximation, therefore: -

$$\theta_S \approx \theta_T \text{ and } \theta_S << 1 \text{ and } \theta_T << 1 \;\Rightarrow\; a \approx \frac{R_T}{R_S} \text{ and } b \approx \frac{R_T}{R_S}(\theta_S - \theta_T)$$

Matching cable is the better choice because any loss factor ("tan(δ)") of the dielectric, equivalent to a small imaginary component for both θ_S and θ_T, is the same for both numerator and denominator.

$$\theta_S = \theta_T \text{ and } \theta_S \to \theta_S(1 + j\delta) \text{ and } \theta_T \to \theta_T(1 + j\delta) \;\Rightarrow\; a = \frac{R_T}{R_S}$$

3.1.1 Example calculation: -

$R_S = 25\Omega$ with 1 metre of cable $(C_S = 200\,pF)$; $R_T = 100\Omega$ at the end of 5 metres of cable $(C_T = 1nF)$ at an operating frequency of 75Hz ($\omega = 2\pi f = 471$ radians per second).

$$\theta_S = 471 \times 25 \times 2 \times 10^{-10} \approx 2.4 \times 10^{-6} \quad \theta_T = 471 \times 100 \times 10^{-9} \approx 4.7 \times 10^{-5}$$

$$\frac{\delta a}{a} \approx 2.2 \times 10^{-9} \text{ (2.2ppb and negligible)}$$

1. Part 5, monograph 1: "Null detectors – the basics". See section 6.1.

3.2 Analysis of parallel capacitance without quadrature balance

One of the main reasons for a quadrature balance is the necessity for filtering in the null detector. This usually consists of a band pass filter at the operating frequency combined with notch filters at power supply frequency harmonics. The filtering can introduce phase shift and any quadrature component in the out-of balance signal appears, at the phase sensitive detector, as an in-phase error. The result is an error in the in-phase balance. In less demanding applications, with low levels of both quadrature and interference expected, there may be no need for a quadrature balance, in which case any phase error in the null detector needs to be kept low, as the following analysis shows: -

Define the phase shift in the null detector: θ_D. The effect is equivalent to multiplying the bridge output by the phase factor $\exp(j\theta_D) = \cos(\theta_D) + j\sin(\theta_D)$.

$$V_{OUT} = \{\cos(\theta_D) + j\sin(\theta_D)\}\left(\frac{aR_1}{1+j\theta_1} - \frac{R_2}{1+j\theta_2}\right)$$

The phase errors of the resistors are very small but the null detector phase error may not be negligible. One can make only the following approximation: -

$$|\theta_1| \text{ and } |\theta_2| << 1 \Rightarrow \quad V_{OUT} \approx \{\cos(\theta_D) + j\sin(\theta_D)\}\{aR_1(1 - j\theta_1) - R_2(1 - j\theta_2)\}$$

The bridge ratio, *a*, is adjusted so that the real part is zero.

$$\mathrm{Re}(V_{OUT}) = 0 \qquad \Rightarrow \qquad aR_1\{\cos(\theta_D) + \theta_1\sin(\theta_D)\} - R_2\{\cos(\theta_D) + \theta_2\sin(\theta_D)\} = 0$$

$$a = \frac{R_2}{R_1} \times \left(\frac{\cos(\theta_D) + \theta_2\sin(\theta_D)}{\cos(\theta_D) + \theta_1\sin(\theta_D)}\right) = \frac{R_2}{R_1} \times \left(\frac{1 + \theta_2\tan(\theta_D)}{1 + \theta_1\tan(\theta_D)}\right)$$

The result is exact if the detector phase error is zero and/or if the resistor phase errors are the same.

To a useful degree of accuracy, assuming small phase errors: -

$$|\theta_1|, |\theta_2| \text{ and } |\theta_D| << 1 \quad \Rightarrow \quad a \approx \frac{R_2}{R_1} \times (1 + (\theta_2 - \theta_1)\theta_D)$$

The proportionate error is, therefore: $\dfrac{\delta a}{a} \approx (\theta_2 - \theta_1)\theta_D$

3.2.1 Example calculation

If one uses the same example as above with a null detector phase error of about 0.1 radians: -

$$\theta_1 = -2.4\mu rad \qquad \theta_2 = -47\mu rad \qquad \text{and} \qquad \tan(\theta_D) \approx 0.1$$

$$\frac{\delta a}{a} \approx -4.5 ppm$$

Clearly a bridge without a separate quadrature balance can only cope with a small amount of quadrature imbalance.

3.3 Quadrature due to series inductance

With low value resistors the problem is often the series inductance. Extra care needs to be taken in the construction of low value resistors to keep the self-inductance and mutual inductance (between the current carrying pair and the voltage sensing pair) as low as possible.

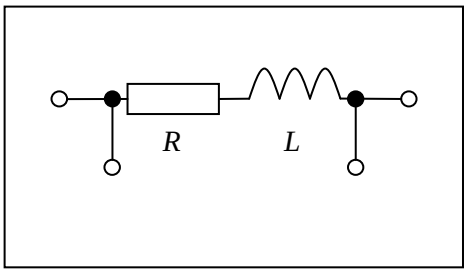


Fig. 3.3.1 Equivalent circuit

The impedance is: $$Z = R + j\omega L$$

The phase error is now positive and inversely proportional to the resistance: -

$$\theta = \arctan\left(\frac{\omega L}{R}\right) \approx \frac{\omega L}{R} \qquad \text{in radians}$$

The analysis is otherwise the same as with quadrature due to parallel capacitance.

3.4 Two-terminal-pair zero-Ohm junctions

The basic idea of a zero-Ohm junction (ZOJ) is that current passing through one pair of connections results in negligible voltage developed across the other pair. In practice a junction resistance as low as a few nΩ is possible. They are useful in the construction of two-terminal-pair resistor networks (see section 4) and four-terminal-pair resistors (see section 5). The most basic type (type 1) is easy to construct (with bits of wire): The current carrying pair and the voltage sensing pair are physically separated by a low value resistor. The pairs are interchangeable $(V \leftrightarrow I)$ but may not be mixed. Type 1 junctions are used in four-terminal-pair resistors.

Type 2 junctions, have one dedicated current and one dedicated voltage sensing terminal. The other two connections can be used for either current carrying or voltage sensing. A high degree of symmetry is required between the dedicated current terminal and the other three. The construction is typically a solid cylinder of copper for the dedicated current terminal with the other three connections emerging radially at 120 degree intervals. Type 2 junctions are used in series connected (Hamon type) networks. See section 4.3.1 for more detail.

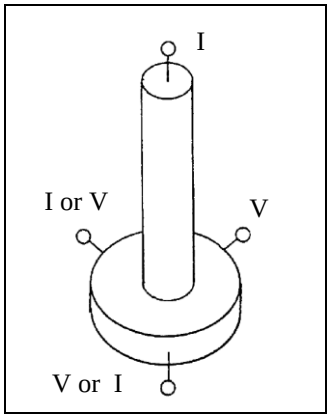


Fig. 3.4.1 A type 2 zero-Ohm junction

Type three is fully symmetrical and requires much higher precision with regards to the (tetrahedral) symmetry of construction. The main advantage is that all four connections are interchangeable (i.e. can be used for current carrying or voltage sensing). Type 3 junctions are used in White/Jones type networks. See section 4.3.2 for more detail.

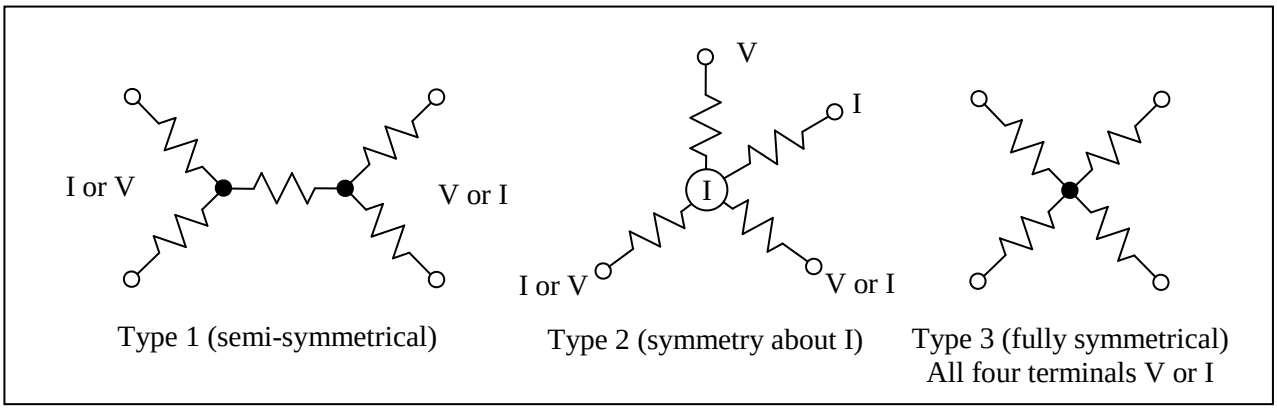


Fig. 3.4.2 Zero-Ohm junctions (equivalent circuits)

Good AC performance is achieved by ensuring, as near as possible, the voltage sensing pair is perpendicular to the current carrying pair and the areas are kept to a minimum thus minimising mutual inductance.

4. Two-terminal-pair resistor networks

4.1 Introduction

Whereas connecting separately constructed two-terminal-pairs in parallel is possible it is not very practicable. This is demonstrated in the next section. Connecting two or more in series is virtually impossible if high accuracy is required. The solution is to construct two or more resistors with permanent series connections using zero-Ohm junctions. The small resistance of the junction connections forms part of the accurately defined resistor which is trimmed accordingly. The other two connections provide the means for measuring each individual resistor as a two-terminal-pair resistor.

The main application is resistance bridge calibration with an accuracy approaching 10ppb (with accurate and stable temperature control) from DC to 400Hz AC. The following, for example, provides the means of producing three values of resistance (R_1, R_2 and R_1+R_2). The latter can be calculated as well as measured.

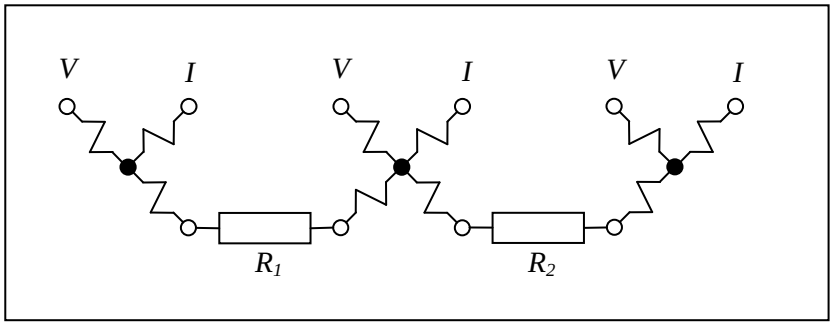


Fig. 4.1.1 Two-terminal pair resistors in series

The basic idea of networks is that series and parallel combinations of two-terminal pair resistors can be calculated and measured with great accuracy. They provide a convenient means for comparing transfer standards and checking/calibrating DC and AC ratio measuring instruments.

4.2 Balancing (potential sharing) resistors

Connecting two separately constructed two-terminal pair resistors in parallel results in the following: -

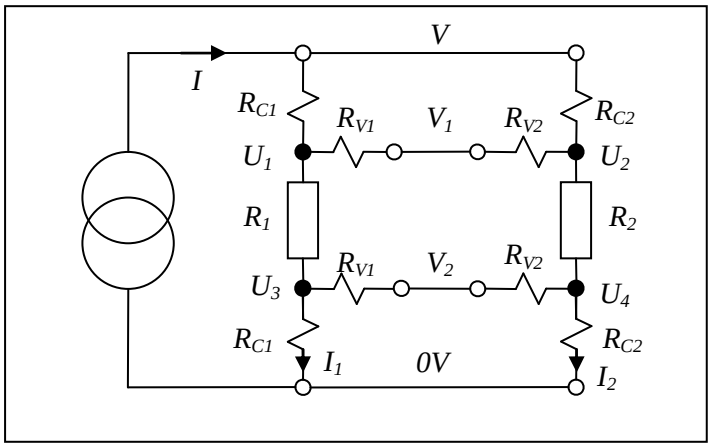


Fig. 4.2.1 A pair of 2TP resistors in parallel

The aim is to be able to calculate the parallel combination accurately, in terms of the individual resistances: -

Ideally, for a parallel combination: $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2}$$

From the symmetry it is clear that there can be no advantage in having different resistances in the voltage sensing or current carrying connections (flip vertically and apply reductio ad absurdum). One can assume, therefore, that both voltage sensing connections for R_1 have resistance R_{V1} and both current carrying connections have resistances R_{C1} and similarly for R_2.

The trick is to adjust the resistance in the current carrying connections so that no current flows in the voltage sensing resistances (i.e. $V_1 = U_1 = U_2$ and $V_2 = U_3 = U_4$).

$$U_3 = U_4 \qquad \Rightarrow V\frac{R_{C1}}{2R_{C1}+R_1} = V\frac{R_{C2}}{2R_{C2}+R_2}$$

Divide by V and take the reciprocal: $\Rightarrow 2+\frac{R_1}{R_{C1}} = 2+\frac{R_2}{R_{C2}}$

From which one can derive the simple condition and define the parameter: $\alpha = \frac{R_{C1}}{R_1} = \frac{R_{C2}}{R_2}$

From the symmetry the upper voltages must also be the same as the following confirms: -

$$U_1 = V\frac{R_{C1}+R_1}{2R_{C1}+R_1} \qquad \text{and} \qquad U_2 = V\frac{R_{C2}+R_2}{2R_{C2}+R_2}$$

In the first divide top and bottom by R_1 and in the second by R_2: -

$$U_1 = V\frac{\alpha+1}{2\alpha+1} \qquad \text{and} \qquad U_2 = V\frac{\alpha+1}{2\alpha+1} \qquad \text{QED}$$

With no current flowing through the voltage sensing connections one can confirm that the aim is achieved. According to Ohm's law: -

$$I_1 = \frac{U_1-U_3}{R_1} = \frac{V_1-V_2}{R_1} \qquad I_2 = \frac{U_2-U_4}{R_2} = \frac{V_1-V_2}{R_2} \qquad \text{and} \qquad I_1+I_2 = I = \frac{V_1-V_2}{R_T}$$

The net effective resistance is, therefore, exactly that for two basic resistors in parallel: -

$$\frac{1}{R_T} = \frac{I_1+I_2}{V_1-V_2} = \frac{1}{R_1}+\frac{1}{R_2}$$

One could, of course, add a third resistor, or as many as required, as long as the resistance in the current carrying pairs conforms to the requirement: -

$$\frac{R_{C1}}{R_1} = \frac{R_{C2}}{R_2} = \frac{R_{C3}}{R_3} = \ldots = \frac{R_{CN}}{R_N}$$

From the symmetry it would also make sense if the resistances in the voltage sensing pairs were also matched: -

$$\frac{R_{V1}}{R_1} = \frac{R_{V2}}{R_2} = \frac{R_{V3}}{R_3} = \ldots = \frac{R_{VN}}{R_N}$$

This is made clear by more detailed analysis, not included here as the algebra is a little tedious. For two resistors (see section 4.3.2 for a reference to the user manual): -

$$R_T = \frac{R_1 R_2}{R_1 + R_2}\left\{1 + \frac{R_1 R_2}{R_1 + R_2}\left(\frac{R_{V1}}{R_1} - \frac{R_{V2}}{R_2}\right)\left(\frac{R_{C1}}{R_1} - \frac{R_{C2}}{R_2}\right)\left(\frac{1}{R_{V1} + R_{V2}}\right)\right\}$$

It would seem a good idea, therefore, to match as closely as practicable both current carrying and voltage sensing connections. The most practical solution, however, given that resistance matching is only feasible to about 1mΩ, is to have very low resistance in the current connections (and switches) and larger, more closely matched resistances in the voltage sensing connections. Typical values are 10mΩ ±1mΩ and 1 - 10Ω ±1mΩ respectively. This is the basis of the White/Jones market leading resistance bridge calibrators.

4.3 Resistance bridge calibrators

There are two main types of resistor networks for bridge calibration: -
a). The Hamon type [1] was originally used to provide 10Ω and 100Ω transfer standards, calibrated against a 1Ω primary standard. It also provided limited means for checking the accuracy of ratio measuring instruments. The technique is now obsolete due to the development of the second type but is included for its ingenuity and for historical interest.
b). The White/Jones type [2] and [3] provides a much more practical means for checking the accuracy of ratio measuring instruments. Four accurate base resistors connected by a zero-Ohm junction provide a convenient method for generating an additional 31 calculable resistance values. When compared with a stable transfer standard resistor (e.g. a Wilkins) a full range of ratios can be measured and compared with calculated values. Unlike the Hamon type commercially available resistance bridge calibrators (RBC) also retain high accuracy at higher frequency (400Hz). Ratio accuracy can approach 10ppb especially with temperature control (±0.01°C).

The US National institute for Science and Technology (NIST) used the White/Jones type to compare five of the world's best selling resistance bridges with some very interesting results [4]: -

Fig 4.3.1 A report from NIST

1. Hamon, B. V.: "A 1-100Ω build-up resistor for the calibration of standard resistors". J. Sci. Instr. 31, 450-453 (Dec 1954).
2. US Patent 5,867,018 (Inventors: Rod White and Keith Jones): PCT/NZ95/00022
3. White, D.R. Jones, K. Williams, J.M. and Ramsey, I.E.: "A simple resistance network for calibrating resistance bridges". IEEE trans. on inst. and meas. Vol.46(5): pp.1068-1074. (1997)
4. Chojnacky, M. Kosior, J. Chaves-Santacruz, L. and Strouse, G.: "Performance assessment of thermometer resistance bridges". NIST Thermodynamic Metrology Group, Sensor Science Division

4.3.1 The Hamon "build-up" network

The original Hamon network [1] consisted of eleven accurate 10Ω resistors permanently connected in series using type two zero-Ohm junctions. The two connections to each junction provide for two-terminal-pair measurement and comparison with a reference and with each other. Groups of resistors can then be connected in parallel or in series-parallel combinations, with suitable external matching resistors, to produce a wide range of resistor values (nominally 1Ω - 100Ω) and accurate ratios in the range 1:10 to 10:1. Hamon's analysis, later confirmed in more detail by Page [2] and Riley [3], showed that an accuracy of ±10ppb, or better, was "feasible". Although obsolete no monograph on high accuracy resistors would be complete without a mention of this ingenious design.

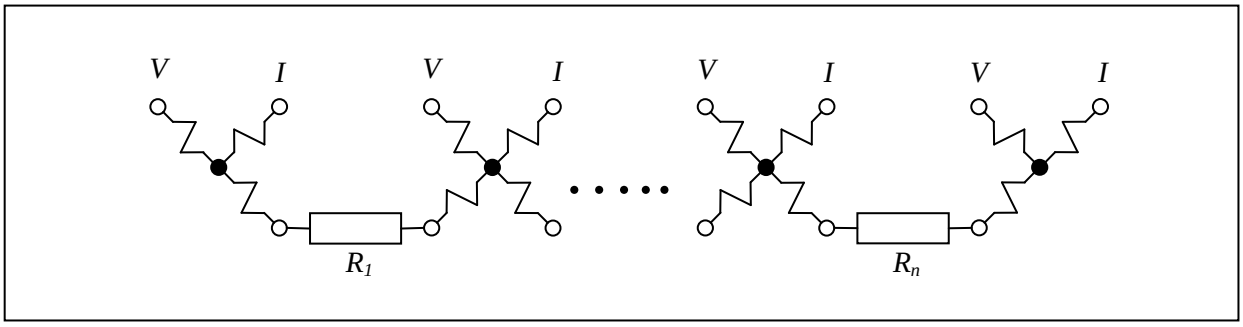


Fig. 4.3.1.1 A Hamon type network

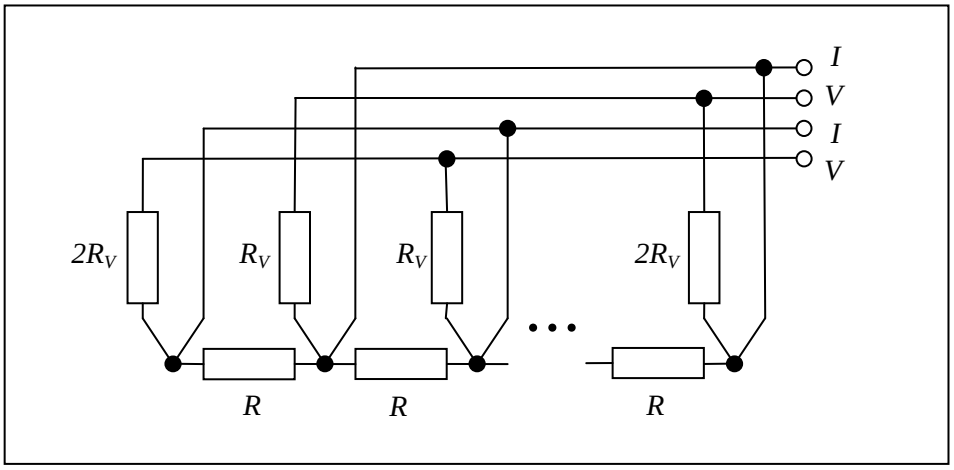


Fig. 4.3.1.2 The Hamon network with external matching resistors

N.B. The original design employed pots of mercury to provide the very low resistance connections!

1. Hamon, B.V.: "A 1-100Ω build-up resistor for the calibration of standard resistors". J. Sci. Instr. 31, 450-453 (Dec 1954).
2. Page, C. H.: "Errors in the Series-Parallel Build up of Four-Terminal Resistors" Journal of research of the National Bureau of Standards-C. Engineering and Instrumentation Vol. 69C, No.3, July-September 1965.
3. Riley, J. C.: "The accuracy of series and parallel connections of four terminal resistors." IEEE Trans. on Instr. & Meas. vol. IM-16. pp 258-268.

4.3.2 The White/Jones resistance bridge calibrator

This ingenious design [1] supercedes the Hamon type and, when automated [2] and kept in a temperature controlled environment, provides a convenient method of checking resistance ratio measurement accuracy at the highest level. The following omits the small resistor symbols for clarity.

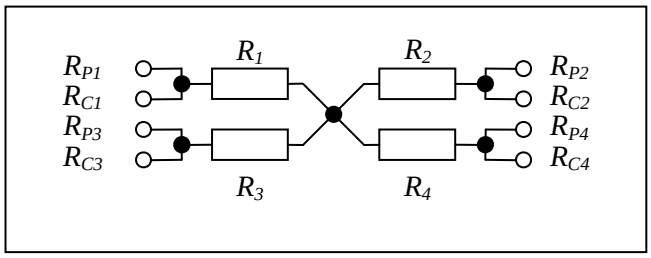


Fig 4.3.2.1 The White/Jones resistance bridge calibrator

Remote controlled RBC [2]

Manual RBC

Fig. 4.3.2.2 Resistance bridge calibrators (courtesy Isotech Ltd [1])

(The USB controlled version is designed to be submerged in a temperature controlled oil bath)

Three models are available suitable for calibrating most commercially available resistance bridges, each with four very accurate resistors, four moderately accurate potential sharing resistors and a fully symmetric (type 3) zero-Ohm junction. The internal current carrying and switch contact resistances are kept very low. Each model is designed to provide a suitable range of ratios with transfer standard resistors of typically 10Ω, 25Ω and 100 Ω respectively, depending on the measuring range of the instrument being calibrated. The potential sharing resistors are critical and are included below (resistance values in Ohms): -

Model	R_1	R_{P1}	R_2	R_{P2}	R_3	R_{P3}	R_4	R_{P4}
RBC13	10.63652	2.635	6.26322	1.551	4.74705	1.176	4.06149	1.006
RBC100	81.81935	2.645	48.17864	1.558	36.51578	1.180	31.24221	1.010
RBC400	251.9874	3.177	147.9912	1.866	116.3731	1.467	40.4480	0.510

The values of the base resistors are chosen not only to provide an optimum range of ratios, when compared to a suitable transfer standard resistor, but also to exercise all nine digits of each decade, at least once, for detecting bridge malfunctions (e.g. a faulty relay).

1. The RBC User manual is available from Isothermal Technology Limited. www.isotech.co.uk
2. White, D.R., Edgar, H., McLennan, B.E., Saunders, P.: "Automation of the resistance bridge calibrator". AIP Conference Proceedings. Vol.1552, no. 8, pp 392-397 (2013).

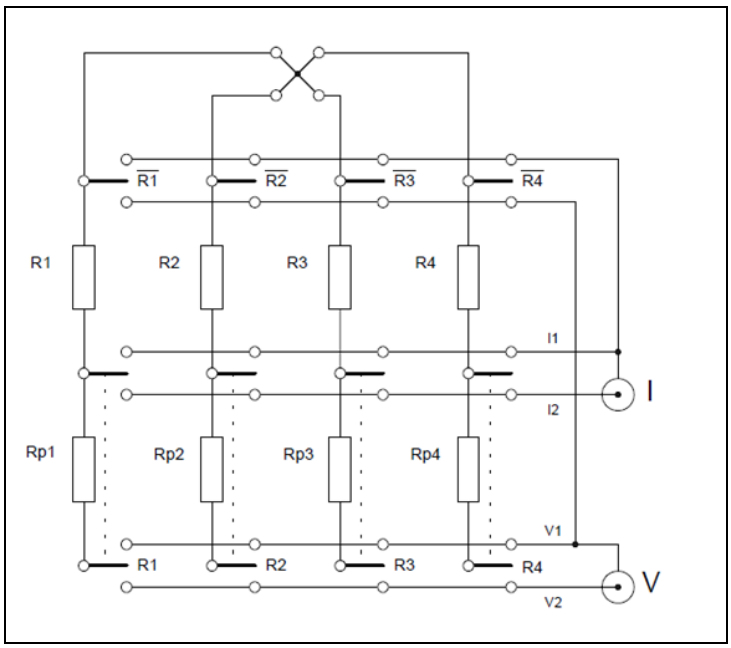


Fig. 4.3.2.3 The internal connection scheme (the manual version has eight switches in total)

The four base resistors can be combined in various series, parallel and series-parallel combinations to provide an additional 31 resistor values which can be calculated (35 values in total). The range of values provides a relatively simple check on linearity of the instrument being tested to better than 100ppb. This is usually sufficient for most applications (e.g. resistance thermometry) as absolute values of resistance are not required.

Even more information (e.g. absolute ratio accuracy) can be gleaned if two or more (i.e. external to the bridge) transfer standard resistor values are available, providing more data points and permitting more complement and three resistor checks, subject to the range of the instrument: -

Complement check (swap resistors): $$\frac{R_2}{R_1}\times\frac{R_1}{R_2}=1$$

Three resistor check: $$\frac{R_2}{R_1}\times\frac{R_1}{R_3}\times\frac{R_3}{R_2}=1$$

5. Four-terminal-pair resistors

In much of the established literature the generally accepted definition of a four-terminal-pair resistor is a two-terminal-pair resistor with overall screen. This dates back many years. In one of the earliest papers Cutkosky, for example, regarded the screen connections as valid terminals of a four-terminal-pair device [1]. According to Awan et al "The result… is a four-terminal-pair definition, which is precise enough for all present practical applications and all values of impedances." [2].

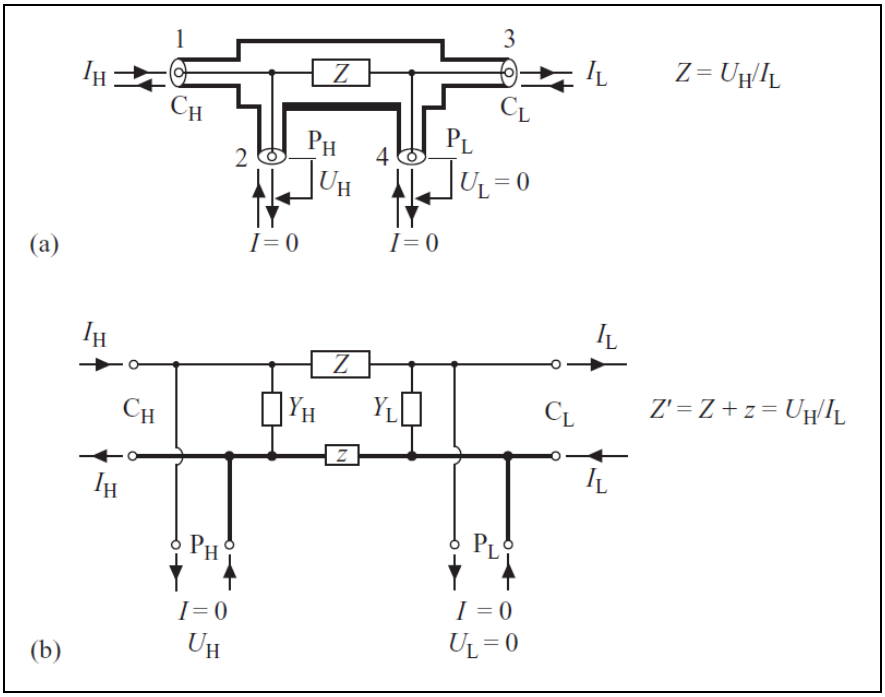


Fig. 5.1 A widely accepted definition of four-terminal-pair impedance [2]

The result, over the years, has been complex bridge designs with earth loops and associated problems [3]. Fortunately there is a better way. The basic idea is to combine a two-terminal-pair resistor with a zero-Ohm junction (ZOJ type 1) plus an overall screen. The defining conditions remain the same: -

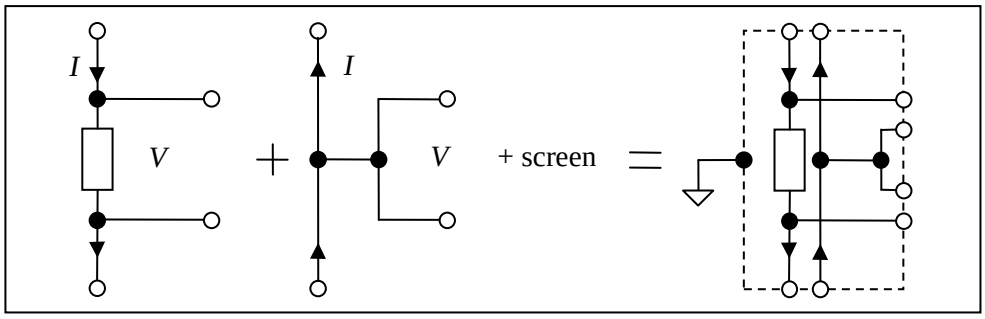


Fig. 5.2 The author's concept of a four-terminal-pair resistor

The measuring current is returned via the ZOJ in such a way that minimises the transmission of magnetic flux. Similarly, the voltage sensing connections are routed to minimise induced voltages due to stray magnetic flux. The screen is provided with a separate connection to local 0V and is insulated from the measuring terminals. Suitably insulated connectors (e.g. BNC co-axial) and insulating washers are readily available.

1. Cutkosky, R. D.: "Four-terminal-pair networks as precision admittance and impedance standards". Trans. IEEE Commun. Electron. 1964; **83**: pp19–22.
2. Awan et al. See footnote page 1.
3. Part 3, monograph 9: "Coaxial AC bridges".

A typical measurement configuration is shown in fig. 5.3. The devices being measured are labelled as general impedances but can be resistive with small phase errors. The configuration ensures accurate equal and opposite currents and the minimal transmission of stray flux at the operating frequency without the need for high permeability toroidal chokes [1]: -

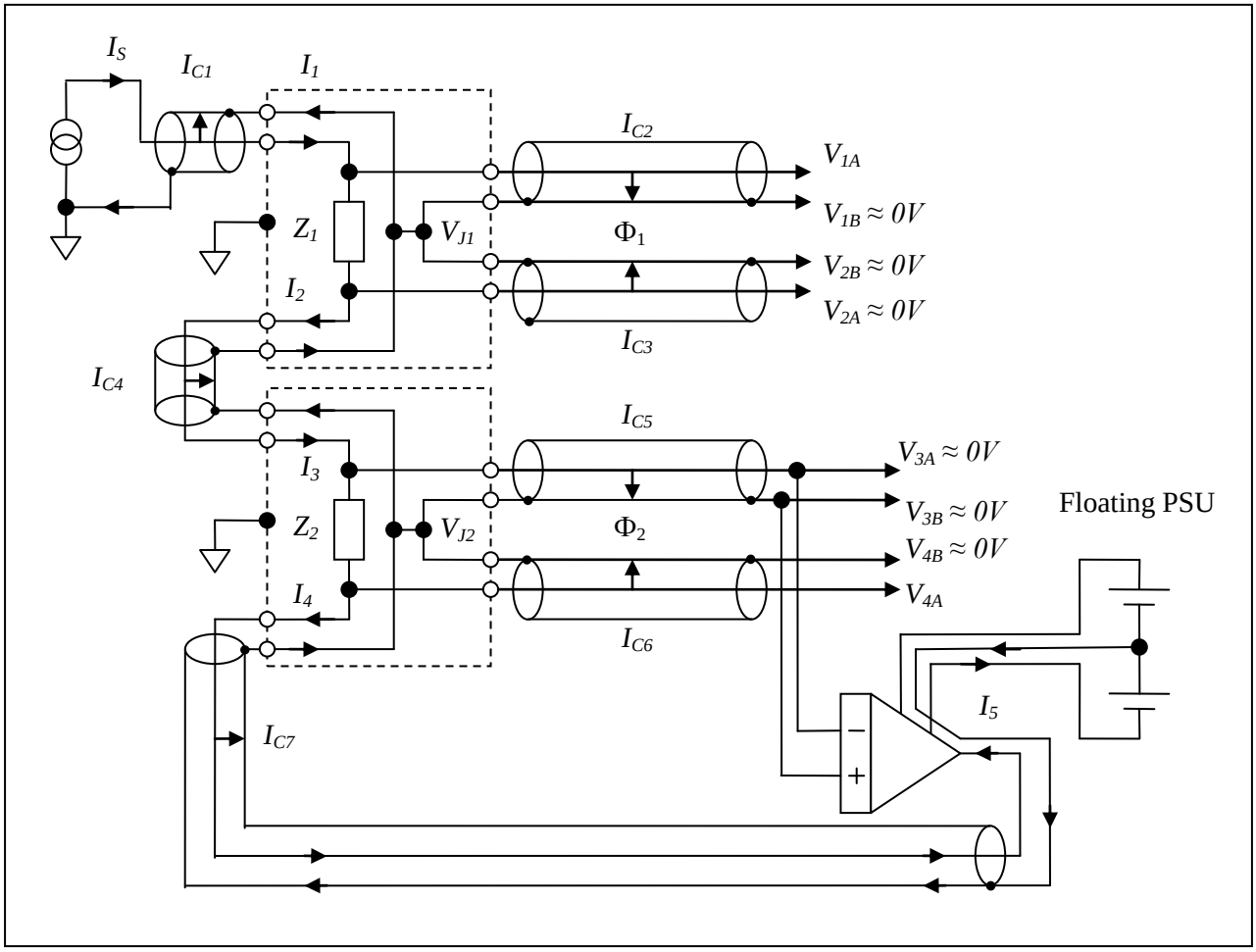


Fig. 5.3 Scheme for measuring the ratio of two four-terminal-pair resistors

Notes: -

1. The current source is shown with the usual symbol for simplicity. See fig. 5.5 for details.

2. The action of feedback of the high gain block is to ensure that the voltage across and, therefore, the currents through cable capacitances C_3, C_4, and C_5 are very small but may not be negligible. According to the definitions above the currents through Z_1 and Z_2 are, therefore: -

$$I_{Z1} = I_2 + I_{C3} \qquad I_{Z2} = I_2 - I_{C4} - I_{C5} \qquad \text{with} \qquad |I_{C3}|, |I_{C4}| \text{ and } |I_{C5}| << I_S \approx I_1 \approx I_2 \approx I_3 \approx I_4$$

3. Assume $V_{1B} \neq V_{2B}$ and $V_{3B} \neq V_{4B}$ because the zero-Ohm junctions may not be perfect and stray flux Φ_1 and Φ_2 between the cables may not be negligible. The relevant voltages are, therefore: -

$$V_{Z1} = (V_{1A} - V_{1B}) - (V_{2A} - V_{2B}) - V_{J1} \qquad \text{and} \qquad V_{Z2} = (V_{3A} - V_{3B}) - (V_{4A} - V_{4B}) - V_{J2}$$

Where V_{J1} and V_{J2} are the voltage generated at the (imperfect) zero-Ohm junctions.

1. Homan, D.N.: 'Applications of coaxial chokes to AC bridge circuits'.
J. Res. NBS – C. 1968; **72C**: 161–65
(A "Choke" is when the coax cable is passed through a toroidal core the required number of times)

The circuit is basically an inverting amplifier [1] with a high gain block [2] with negligible input current providing an accurate virtual earth: -

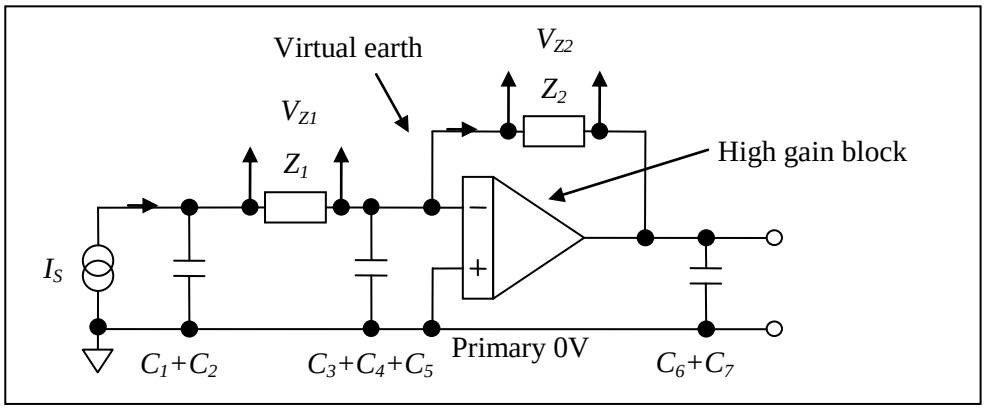


Fig. 5.4 The basic configuration

Notes: -
1. There is only one connection to earth – no earth loops. This is usually at the current source for practical reasons.
2. The non-inverting input of the HGB can be regarded as the primary 0V reference point for the bridge. The virtual earth is maintained at this potential plus the DC offset and random noise of the HGB plus a very small carrier signal depending on the open loop gain.

The current source and guard amp output stages are best constructed on the same PCB module with twisted triples for their (separate) power supplies and twisted pairs for the carrier signals. The principle can be best understood by considering the positive half cycle of the sine wave and following the route by which the current flows. Within the module the PCB tracks are routed to keep the area depicted to a minimum: -

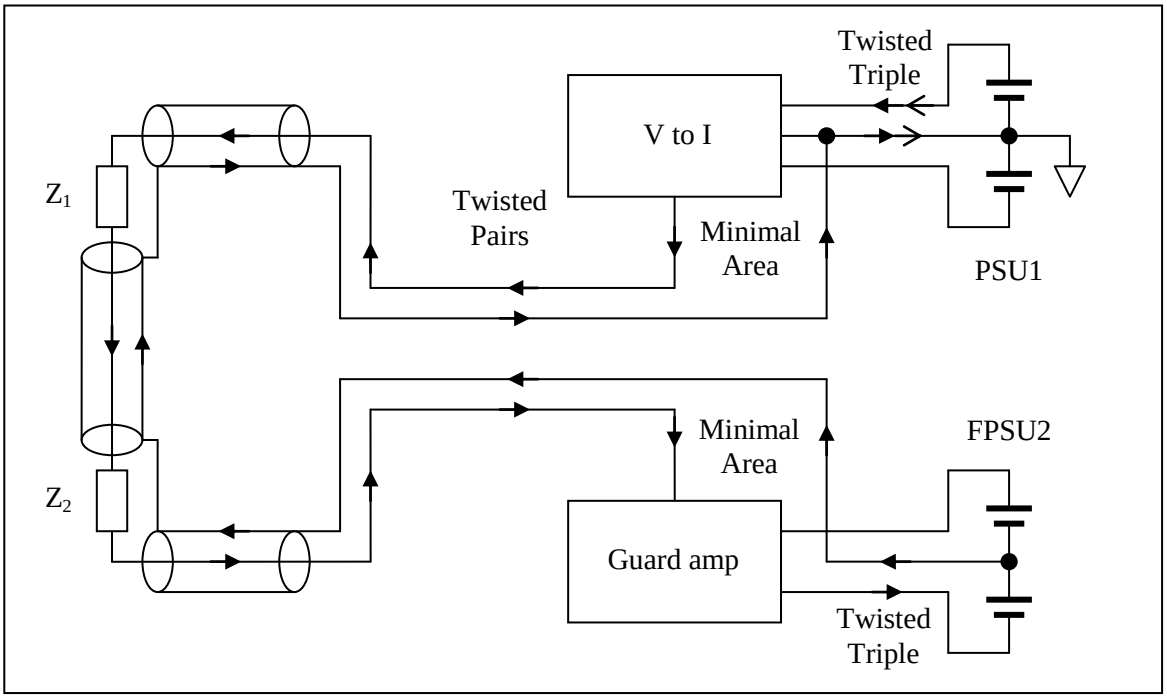


Fig. 5.5 Physical layout and separate power supplies

The current source is usually a "Howland pump" type voltage to current converter which takes and returns a small but not negligible proportion of the total current back to the power supply hence the two pairs of arrows on PSU1.

1. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators".
2. Part 4, monograph 1: "High gain blocks".

Analysis: -

According to Kirchhoff's law (conservation of charge) the currents flowing through the lumped impedances are, in the complex representation (phasors): -

$$I_{Z1} = I_2 + I_{C3} \qquad I_{Z2} = I_2 - I_{C4} - I_{C5} \quad \text{with} \quad |I_{C3}|, |I_{C4}| \text{ and } |I_{C5}| << |I_2| \approx I_S$$

One can take the current I_S as the primary phase reference (i.e. a real number).

One can usually assume that any stray flux is sufficiently uniform so that the voltage induced between the conductors of a pair (coax or twisted) is negligible. According to Kirchhoff's laws, therefore: -

$$(V_{1A} - V_{2A}) = V_{Z1} + V_I \quad \text{where} \quad V_I = \text{the induced voltage due to } \Phi_1$$

$$(V_{1B} - V_{2B}) = V_{J1} + V_I \quad \text{where} \quad V_J = \text{the voltage error of the zero-Ohm junction.}$$

$$\Rightarrow \; V_{Z1} = V_{1A} - V_{2A} - V_I = (V_{1A} - V_{1B}) - (V_{2A} - V_{2B}) - V_{J1}$$

Assume an ideal zero-Ohm junction though it would be wise to measure the voltage to be sure: -

$$V_{J1} = 0V \quad \Rightarrow \quad V_{Z1} = (V_{1A} - V_{1B}) - (V_{2A} - V_{2B})$$

Similarly for cables 5 and 6: $V_{Z2} = (V_{3A} - V_{3B}) - (V_{4A} - V_{4B})$

The ratio of impedances is, therefore: -

$$\frac{Z_2}{Z_1} = \frac{V_{Z2}}{I_{Z2}} \times \frac{I_{Z1}}{V_{Z1}} = \frac{(V_{3A} - V_{3B}) - (V_{4A} - V_{4B})}{(V_{1A} - V_{1B}) - (V_{2A} - V_{2B})} \times \frac{I_2 + I_{C3}}{I_2 - (I_{C4} + I_{C5})}$$

Define voltage ratio: $$\alpha = \frac{(V_{4B} - V_{4A})}{(V_{1A} - V_{1B})} \approx \frac{Z_2}{Z_1}$$

This is very nearly the required ratio and needs to be measured very accurately.

Define voltage ratio: $$\beta = \frac{(V_{3A} - V_{3B})}{(V_{1A} - V_{1B})} \approx \frac{(V_{3A} - V_{3B})}{V_{Z1}} << 1$$

The voltage difference $(V_{3A} - V_{3B})$ is the error at the input of the high gain block and is very small.

Define voltage ratio: $$\gamma = \frac{(V_{2A} - V_{2B})}{(V_{1A} - V_{1B})} \approx \frac{(V_{2A} - V_{2B})}{V_{Z1}} << 1$$

If one assumes an ideal HGB $(V_{3A} - V_{3B} = 0V)$ the voltage difference $(V_{2A} - V_{2B})$ is entirely due to the resistances in cable 4 and is also quite small.

Also, if the capacitive currents through cables 3, 4 and 5 are very small compared to the main current, I_2, then to a very good approximation: -

$$\left|I_{C3}\right|,\left|I_{C4}\right| \text{ and } \left|I_{C5}\right| << I_2 \qquad \Rightarrow \qquad \frac{I_2 + I_{C3}}{I_2 - \left(I_{C4} + I_{C5}\right)} \approx 1 + \frac{I_{C3} + I_{C4} + I_{C5}}{I_2}$$

The ratio thus simplifies. To a very good approximation: -

$$\frac{Z_2}{Z_1} \approx \left(\frac{\alpha + \beta}{1 - \gamma}\right) \times \left(1 + \frac{I_{C3} + I_{C4} + I_{C5}}{I_2}\right)$$

A further approximation is possible: -

$$\left|\beta\right| \text{ and } \left|\gamma\right| << 1 \quad \Rightarrow \quad \frac{Z_2}{Z_1} \approx \left(\alpha\left(1 + \gamma\right) + \beta\right) \times \left(1 + \frac{I_{C3} + I_{C4} + I_{C5}}{I_2}\right)$$

The voltage ratio α must be measured very accurately as it is, by far, the major contributor. The much smaller voltage ratios β and γ can be measured with reduced accuracy.

In many cases, with fairly short lengths of cable, a two-stage high gain block and a low measuring frequency the correction factors are negligible and one can assume, with sufficient accuracy, the simple result: -

$$\frac{Z_2}{Z_1} \approx \alpha = \frac{V_{4B} - V_{4A}}{V_{1A} - V_{1B}}$$

5.1 Example calculations

This type of arrangement can be used for measuring the ratio of relatively high value resistors, usually for comparing transfer standards. I shall assume that Z_1 and Z_2 are 10kΩ transfer standard resistors with small phase errors. The current source is a generous 100μA (pk-pk) at 75Hz (ω = 471radians/s) resulting in voltage drops across the resistors of V_{Z1} and $V_{Z2} \approx$ 1Vpk-pk.

The power dissipation is 0.1mW and, with a little care taken in the thermal design, self heating is not an issue.

5.1.1 Cable phase and magnitude errors

A reasonable assumption for the cables is 1m long with ≈ 0.1Ω series resistance (both conductors) and 100pF of parallel capacitance. At low frequency each centimetre of cable can be approximated as a low pass filter with ≈1mΩ resistance and 1pF capacitance with a transfer function, for *N*cm, in the complex representation $(s = j\omega)$: -

$$T_N(s) \approx \frac{1}{(1+\tau s)^N} \approx 1 - N\tau s \quad \text{with} \quad \tau = RC < 1m\Omega \times 1pF \approx 10^{-15} s$$

The total phase error (quadrature) for each 1m of cable is, therefore: $N\tau\omega < 4.71\times10^{-11}$ **radians and negligible. The real part (in-phase) is second order with respect to frequency and even smaller.**

Also, electromagnetic signals travel along most types of cable at about two thirds the speed of light $(\approx 2\times10^8 ms^{-1})$. which, at 100Hz (period 10ms) represents a phase shift of 5 × 10^{-7} radians per metre of cable. **Time delays are negligible.**

Similarly, the loss factor ("tanδ") of the insulating dielectric of reasonably good quality cable is very small (typically <10^{-3} rad) and represents small phase errors in the currents $I_{C1} - I_{C7}$. **Cable phase and magnitude errors are truly negligible.**

5.1.2 Zero-Ohm junction errors

Type 1 zero-Ohm junctions are easily constructed and checked and, if necessary, adjusted to <10nΩ (with 1A measuring current the output is <10nV). **At an operating current of 100μA the error voltage is then negligible.**

5.1.3 HGB error voltage

The recommended HGB is a two-stage (type 1) for which the main error voltage is in-phase [1]. An open loop gain of 10^9 is readily achieved at 75Hz. With an output voltage of 1V (pk-pk) the input error voltage is less than 1nV (pk-pk). **The correction factor β is then very small and usually negligible: -**

The transfer function of the HGB is defined by: $$H_2(s) = \frac{V_{4A}}{V_{3B} - V_{3A}}$$

For an ideal inverting configuration: $$V_{4A} = -\frac{Z_2}{Z_1} V_{1A}$$

$$Z_2 \approx Z_1 \text{ and } V_{1B} \approx 0V \quad \Rightarrow \quad \beta = \frac{V_{3A} - V_{3B}}{V_{1A} - V_{1B}} \approx \frac{V_{3B} - V_{3A}}{V_{4A}} \left(\frac{Z_2}{Z_1}\right) \approx \frac{1}{H_2(s)} \left(\frac{Z_2}{Z_1}\right) < 1ppb$$

1. Part 4, monograph 1: "High gain blocks".

If one assumes negligible resistances in cable 4 the error voltage results in negligible currents through the capacitances of cables 3, 4 and 5: -

$$I_2 \approx I_4 \approx -\frac{V_{4B}}{Z_2} \quad \text{and} \quad I_{C3} + I_{C4} + I_{C5} = (V_{3A} - V_{3B})s(C_3 + C_4 + C_5)$$

$$\Rightarrow \quad \frac{I_{C3} + I_{C4} + I_{C5}}{I_2} = \frac{j\omega(C_3 + C_4 + C_5)Z_2}{H_2(s)} \quad \text{(quadrature)}$$

With a total cable capacitance of 300pF at 75Hz the error term is negligible (Z_2 = 10kΩ): -

$$\left|\frac{I_{C3} + I_{C4} + I_{C5}}{I_2}\right| < 1.4 \times 10^{-12}$$

5.1.4 Resistance in cable 4

The measuring current flows through cable 4 creating a small voltage gradient and currents through C_3 and distributed C_4. A basic model consists of half the conductor resistance either side of the actual capacitance of cable 4. Note that the measuring current does not flow through any part of cables 3 or 5: -

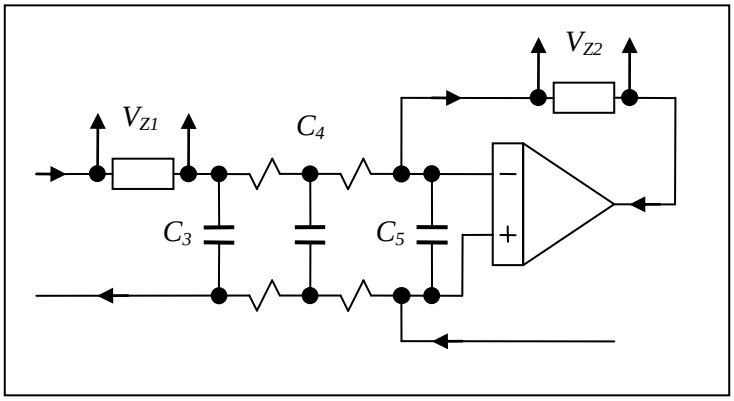


Fig. 5.1.4.1 A basic model

This is worth checking as the distribution of resistance and capacitance is continuous. If one assumes an ideal HGB: -

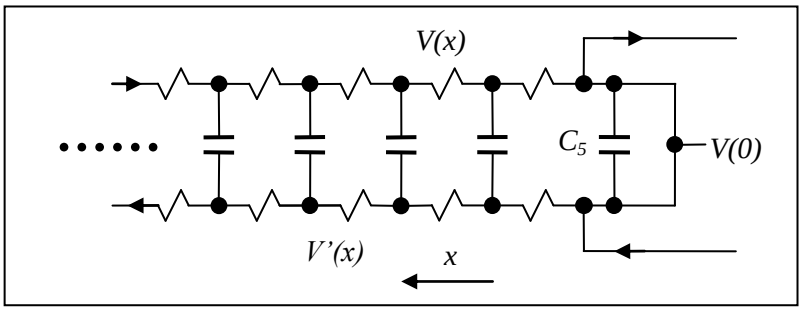


Fig. 5.1.4.2 A gradient model for cable 4

Define parameters: -

L = Length of cable 4 in metres

ρ = Resistance per unit length of the inner conductor (Ωm^{-1})

ρ' = Resistance per unit length of the outer conductor (Ωm^{-1})

σ = Capacitance per unit length (Fm^{-1})

ω = Operating frequency ($rads^{-1}$)

x = Distance from the HGB end in metres

According to Ohm's law the voltages along the inner and outer conductors of cable 4 vary according to: -

$$V(x) = V(0) + I_2 \rho x \quad \text{and} \quad V'(x) \approx V(0) - (I_2 + I_{C6} + I_{C7})\rho' x \quad \text{respectively}$$

The currents caused by the capacitance of cables 6 and 7 are very small compared to the measuring current so that the difference voltage is, to a sufficiently good approximation: -

$$|I_{C6} + I_{C7}| << I_2 \quad \Rightarrow \quad V - V' \approx I_2(\rho + \rho')x$$

The capacitive current over the infinitesimal distance *dx* is: $dI_{C4} = I_2(\rho + \rho')s\sigma x dx$

If one assumes that ρ and ρ' are constant (uniform resistance along the cable), being careful to employ good quality, low contact resistance connectors, the integral is simple: -

$$\frac{I_{C4}}{I_2} \approx (\rho + \rho')s\sigma \frac{L^2}{2} = \frac{R + R'}{2} sC_4$$

Where $R + R' =$ total resistance (inner and outer) of cable 4, confirming the basic model of fig. 5.1.4.1. The voltage across cable 3 also results in a current: -

$$\frac{I_{C3}}{I_2} = (R + R')sC_3$$

The voltage across cable 5 is zero so that: $\frac{I_{C5}}{I_2} = 0$

The correction factor is, therefore: $\frac{I_{C3} + I_{C4} + I_{C5}}{I_2} = (R + R')s\left(C_3 + \frac{C_4}{2}\right)$

With each cable 1m long (100pF) and a total resistance of 0.1Ω (inner plus outer) the correction factor is: -

$$\frac{I_{C3} + I_{C4} + I_{C5}}{I_2} = j \times 0.1 \times 471 \times 150 \times 10^{-12} \approx j7.1ppb \quad \text{(quadrature)}$$

This is easily rejected with a suitable quadrature null balance servo.

5.2 Calculable resistors

In recent years there has been a revival of interest in the link between the Farad (the SI unit of capacitance) and the Ohm (the SI unit of resistance) via a quadrature bridge [1]. The traceability chain is nicely summarised in a diagram: -

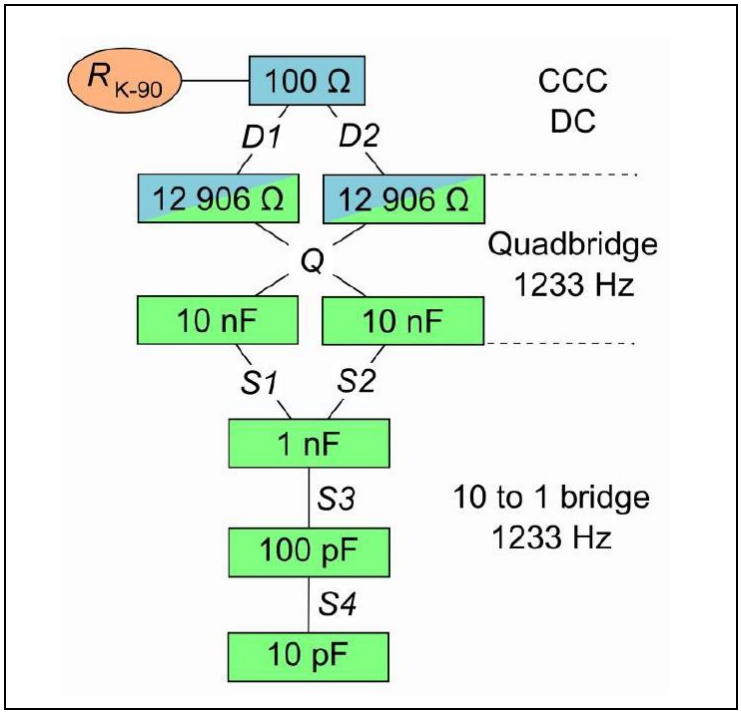


Fig. 5.2.1 Links in the traceability chain from Ohm to Farad (courtesy METAS [3])

The best QHR standards and cryogenic current comparators (CCC) employ alternating DC or very low frequency AC (<1Hz) for comparing resistors. The reasons include: -
a). Power dissipation within the cryostat and heat conduction down the connections must be very low. One consequence is long and very thin coaxial connections with significant series resistance and parallel capacitance. Some of the capacitive current inevitably flows through part of the series resistance resulting in heat dissipation and measurement errors which increase with frequency.
b). The very high magnetic field employed results in a force on any current carrying conductors. The result is unwanted vibration which manifests as increased series inductance as well as more heat dissipation.
c). Stray capacitance within and from the QHR to its surroundings also result in errors which increase with frequency [2].

The optimum frequency for measuring capacitance, on the other hand, is around 1-10 kHz. There is a need, therefore, for a type of transfer standard resistor with calculable magnitude and phase as a function of frequency (typically up to 10 kHz), based on a measured DC value and ("a priori") fundamental principles.

The AC characteristics of a more robust (portable) type of resistor (e.g. Vishay) can then be established, with a coaxial AC bridge, which is then passed around various NMIs for comparing with capacitors using a "quadrature" bridge [1].

For a comprehensive review see [3].

1. Part 3, monograph 10: "A quadrature bridge"
2. Ahlers F. J., Jeanneret B., Overney F., Schurr J. and Wood B. M.:"Compendium for precise ac measurements of the quantum Hall resistance." Metrologia (Oct 2009). DOI: 10.1088/0026-1394/46/5/R01
3. Schurr, J. et al: "Final Report of the Supplementary Comparison EURAMET.EM-S31: Comparison of capacitance and capacitance ratio (May 2017).

There are two main types of (a priori) calculable resistor. Both rely on relatively simple geometry so that the impedance can be modelled analytically (in terms of "lumped" components – resistance, inductance and stray capacitance) as well as calculated using finite element analysis:-
a) The coaxial or "Haddad" type [1].
b) The bifilar, quadrifilar and octofilar or "Gibbings" type [2].

Both are sensitive to shock and vibration and require very accurate temperature control. They must be moved, very gently, between the QHR/CCC and adjacent higher frequency apparatus and back again whilst maintaining a constant temperature. The more robust transfer standard resistors and/or capacitors can then be transported (usually by hand) to other calibration labs.

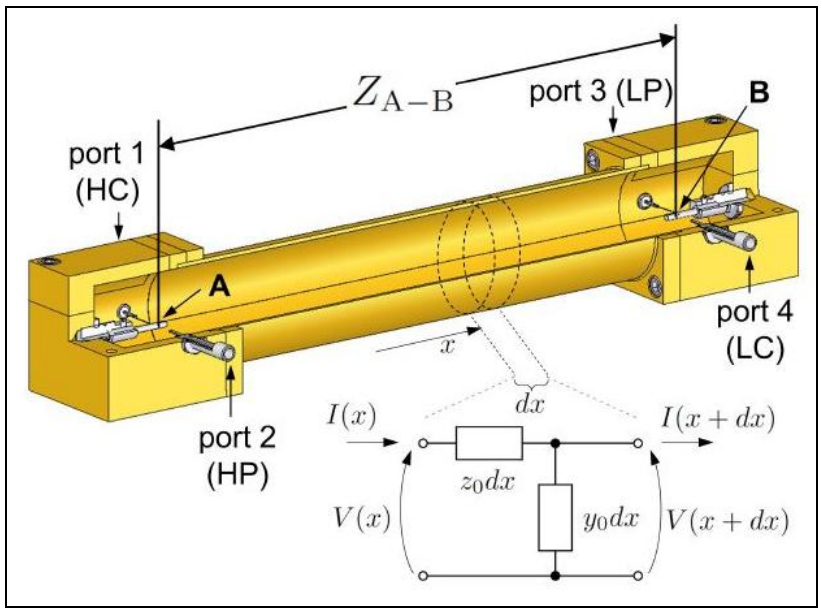


Fig. 5.2.2 A coaxial type calculable resistor (courtesy METAS)

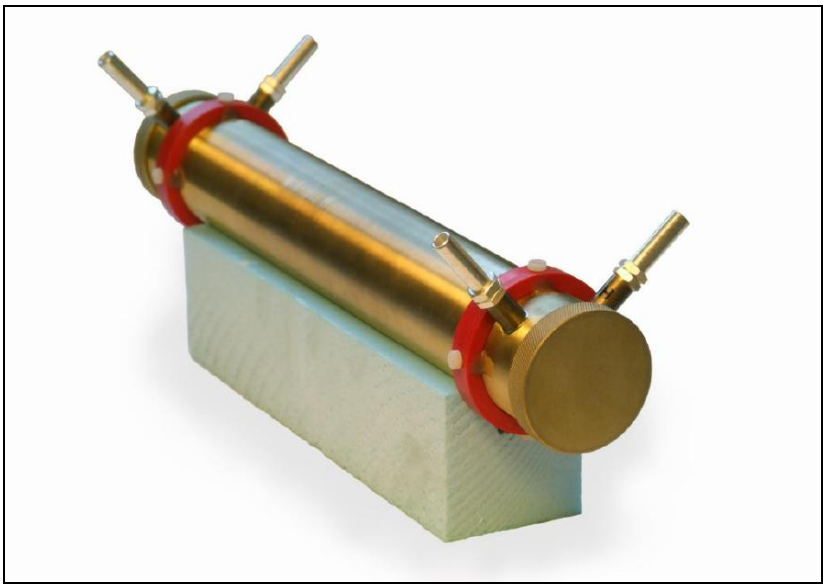


Fig. 5.2.3 Coaxial resistor – the final article (courtesy Kucera et al [3])

1. Haddad R. J.: "A resistor calculable from DC to ω = 10^5 rad/s," Sch. Eng. Appl. Sci., George Washington Univ., M. S. Thesis (Apr. 1969)
2. Gibbings, D. L. H.: "A design for resistors of calculable ac/dc resistance ratio" Proc. IEE vol. 110 pp. 335–347 (1963)
3. Kucera, J., Vollmer, E. and Schurr, J: "Precision Haddad-type Calculable Resistors". Presentation to CPEM/PTB 2008.

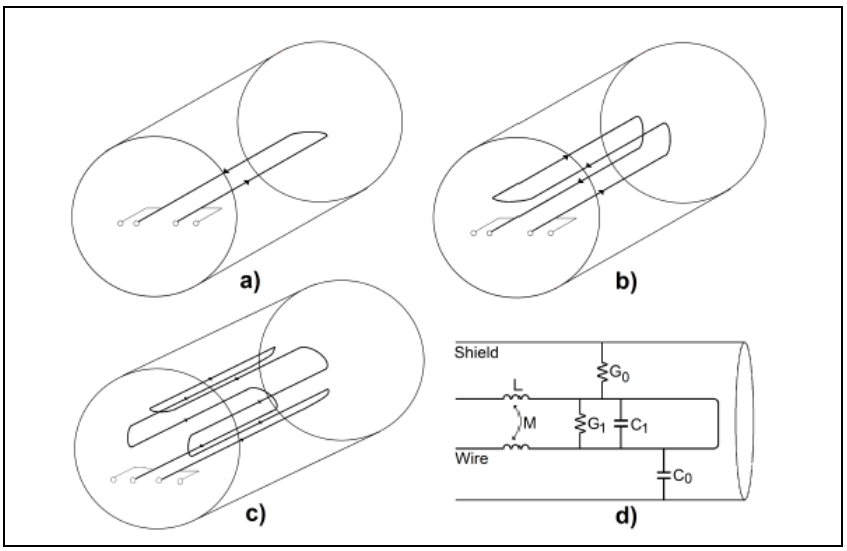


Fig. 5.2.4 Calculable resistors – basic construction and model (courtesy Pacheco-Estrada et al [1]): -

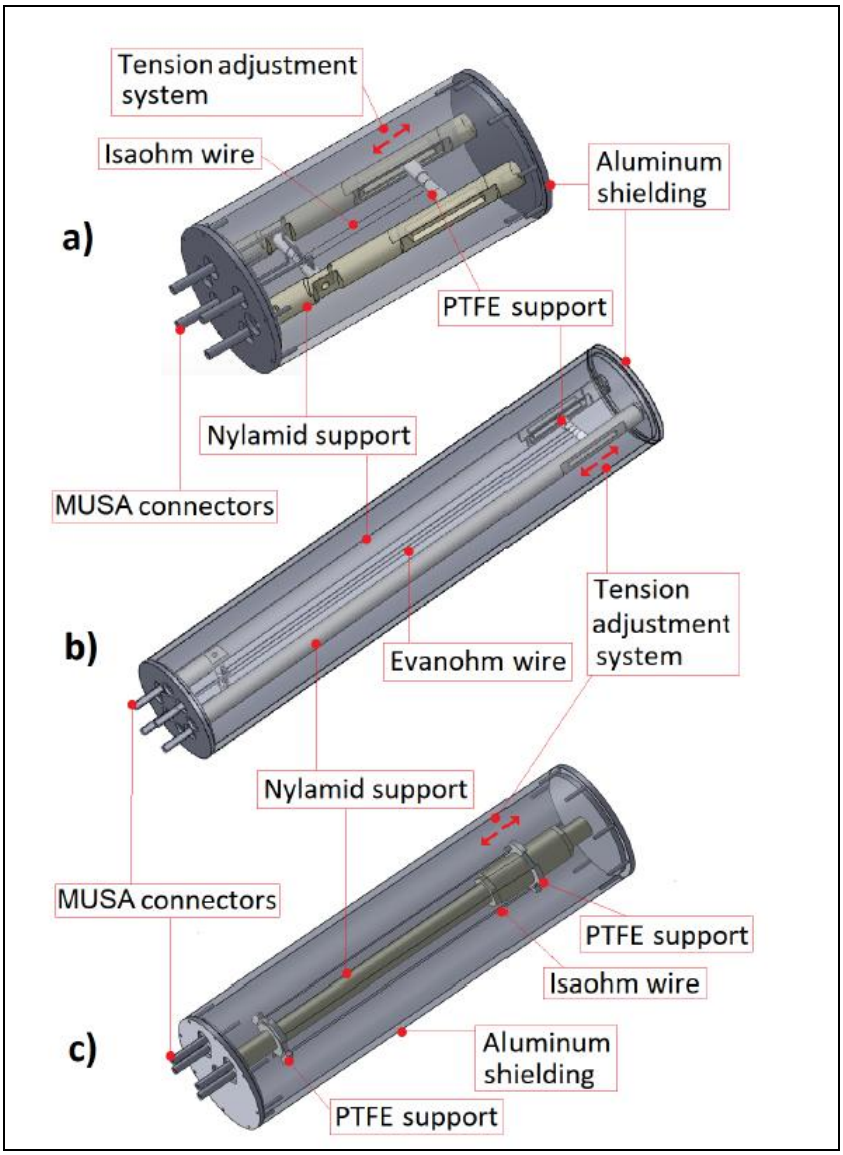


Fig. 5.2.5 Calculable resistors – construction details (courtesy Pacheco-Estrada et al [1]): -

1. Pacheco-Estrada, A. H. et al: "A Simple Methodology to Develop Bifilar, Quadrifilar and Octofilar Calculable Resistors". Appl. Sci. **2020**, 10, 1595.

6. Johnson Noise

One of the main disadvantages of resistors is that they generate random noise. The free electrons within the resistor behave much like a gas - flying about in random directions with a normal (i.e. Gaussian) energy distribution. With a low noise amplifier it is possible to measure a small random voltage at the terminals. The open circuit noise voltage (RMS) is: -

$$V_{JN} = \sqrt{4kTBR}$$

Where: -

$k = 1.38 \times 10^{-23} JK^{-1}$ is Boltzmann's constant
T = Absolute temperature (*K*)
B = Bandwidth in Hz
R = Resistance in Ohms

Similarly the closed circuit noise current (RMS) is: $$I_{JN} = \sqrt{\frac{4kTB}{R}}$$

The "noise power" is a useful concept: -

$$P_{JN} = V_{JN} I_{JN} = 4kTB$$

It is shown elsewhere [1] that the noise power is four times the heat energy which would flow down the conductors from a warm source resistor to a matching cold resistor.

Similarly, best noise performance is obtained when the first stage amplifier is matched to the source resistance [1]. This occurs when the noise resistance (noise voltage divided by noise current) of the amplifier input stage is equal to the source resistance. Bipolar transistors are best matched to resistances between 100Ω and 10kΩ [2]. Matching to lower source resistance can be achieved with a suitable transformer [3]. Matching to higher value resistance is typically improved by using a JFET transistor input stage (100kΩ – 10MΩ) [4].

1. Part 5, monograph 1: "Null detectors – the basics". See appendix A1.1 and A1.3
2. Part 5, monograph 2: "Low noise BJT pre-amps"
3. Part 3, monograph 5: "Noise matching transformers"
4. Part 5, monograph 3: "Low noise JFET pre-amps"

7. Picture gallery

Fig. 7.1 A transportable 10kΩ transfer standard resistor (courtesy IET Labs inc.)

An alternative to accurate temperature control is either by measuring the temperature of the oil bath with a separate thermometer (thermometer well provided) or by measuring the incorporated resistance thermometer (nominally 10kΩ) with a dedicated channel on a scanner. The calibration certificate provides the known dependence: -

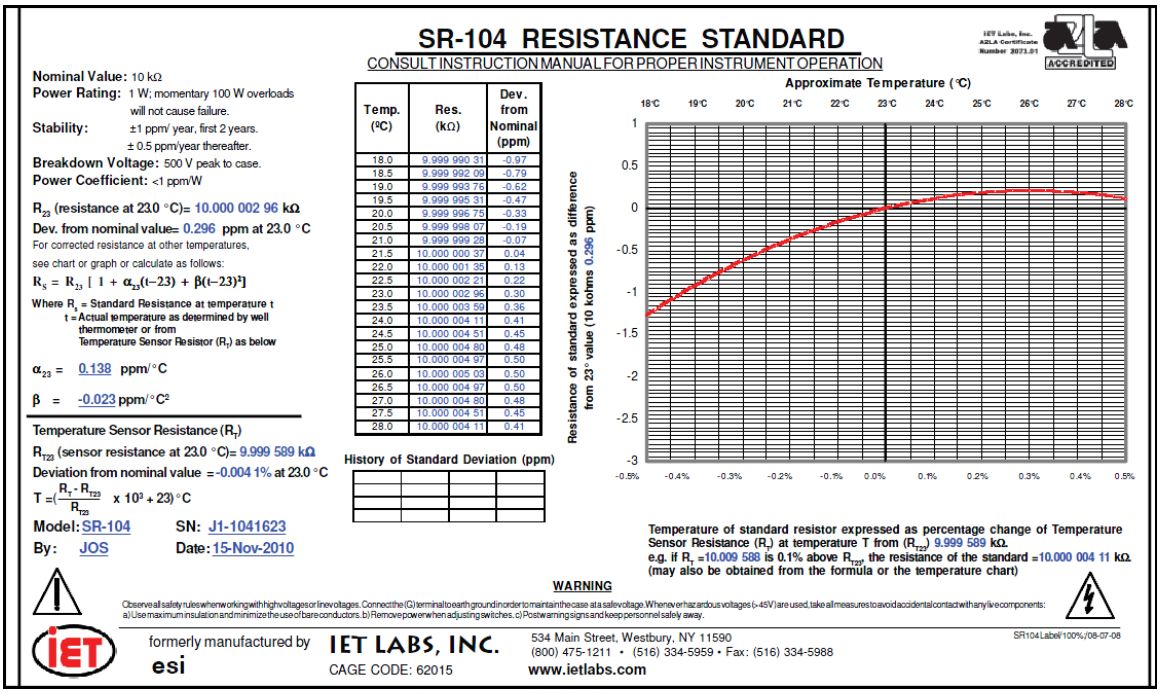


Fig. 7.2 Calibration certificate for a 10kΩ transfer standard resistor

High accuracy single capacitors

1. Introduction

High accuracy single capacitors are used as calibration transfer standards and convenient reference devices for capacitive sensors (e.g. variable gap, humidity and strain gauges). For some years a calculable capacitor was also a serious contender as a primary standard for the Ohm [1 and 2]. Attempts have also been made to construct capacitors with a fused quartz dielectric [3] but the most common types use air as a dielectric. The capacitance varies slightly with temperature, pressure and humidity. Temperature sensitivity can be mitigated by choosing materials and relative dimensions whereby the expansion coefficient is self-compensating, at least over normal lab temperatures and with stable temperature control. An empirical formula can be used to compensate for variations in atmospheric pressure and humidity [4]. For the very highest accuracy the capacitor can be flushed through with dry air and then sealed.

According to Marina Santo Zarnik and Darko Belavic: "The permittivity of the air as a function of the relative humidity can be calculated from the empirical relation [4]: -

$$\varepsilon_{air} = \varepsilon_0 \cdot \left[1 + \frac{211}{T} \cdot \left(P + \frac{48 \cdot P_S}{T} \cdot RH \right) \cdot 10^{-6} \right]$$

where ε_0 is the permittivity of vacuum, T is the absolute temperature (K), RH is the relative humidity (%), P (mm Hg) is the pressure of the air, and P_S (mm Hg) is the pressure of saturated water vapour at the temperature T".

Fig. 1.1 A 1pF reference capacitor (picture courtesy ASL Ltd)

1. Raynor, G. H.: "NPL Calculable Capacitor." IEEE Trans. Instrum. Meas. Vol IM-21. Pp 361 – 365, 1972.
2. Thomson, A. M. 1958.: "The precise measurement of small capacitances."
IRE Trans. Instrum. I7 245-53.
3. Cutkosky, R. D. and Lee, L. H.: "Improved ten-picofarad fused silica dielectric capacitor." NBS J. Res., vol 69C, pp 173 – 179, July – Sept 1965.
4. Santo Zarnik, M. and Belavic, D.: "An Experimental and Numerical Study of the Humidity Effect on the Stability of a Capacitive Ceramic Pressure Sensor". Radioengineering, Vol. 21, No. 1, April 2012 201. http://www.radioeng.cz/fulltexts/2012/12_01_0201_0206.pdf

2. Transfer standard air capacitors

Transfer standard air capacitors are usually fabricated as concentric cylinders, machined to high precision and mounted on ceramic insulators inside a screened metal enclosure.

Fig. 2.1 A 10pF transfer standard capacitor (picture courtesy ASL Ltd)

The electrode that connects to the signal generator output is the outer cylinder and acts as extra screening for the (high impedance) output electrode. As the name implies the dielectric medium is simply air as this is adequate for most applications.
The connections to the electrodes are usually BNC with the outer screen of only one coax cable connected to the enclosure. The screen of the input BNC must be earthed at the signal generator but not to the enclosure. The enclosure and screen of the output BNC must be earthed at the charge amplifier, hence the "third terminal". This ensures that the current due to the cable capacitance on the input side flows back to the signal generator and not down the screen of the output cable. See fig 2.3.

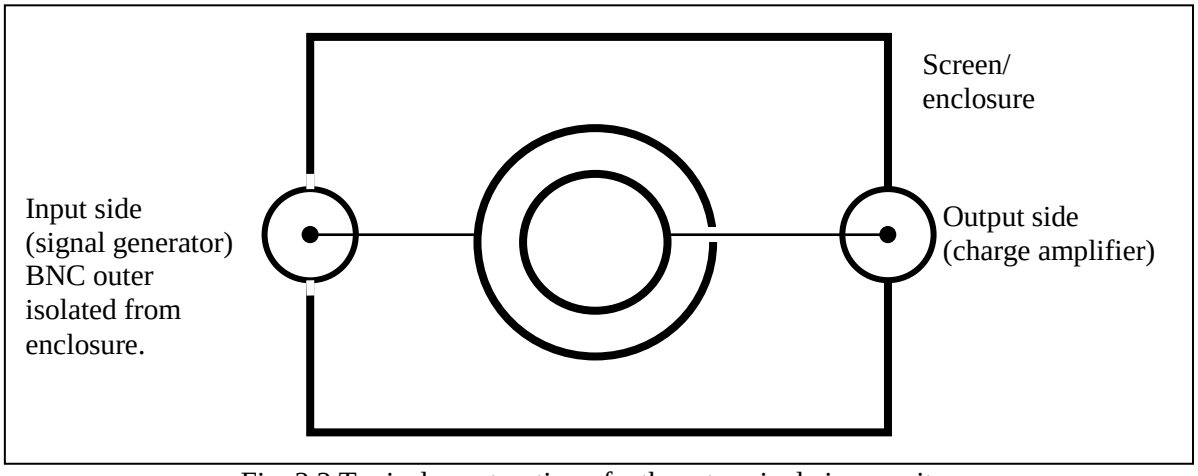


Fig. 2.2 Typical construction of a three terminal air capacitor

The main characteristics are: -

- Moderately accurate: usually specified to ±0.1% or ±0.01%
- Very stable with time and temperature (a few ppm).
- Very low loss tangent (typically less than 10^{-5} at 1.6kHz)

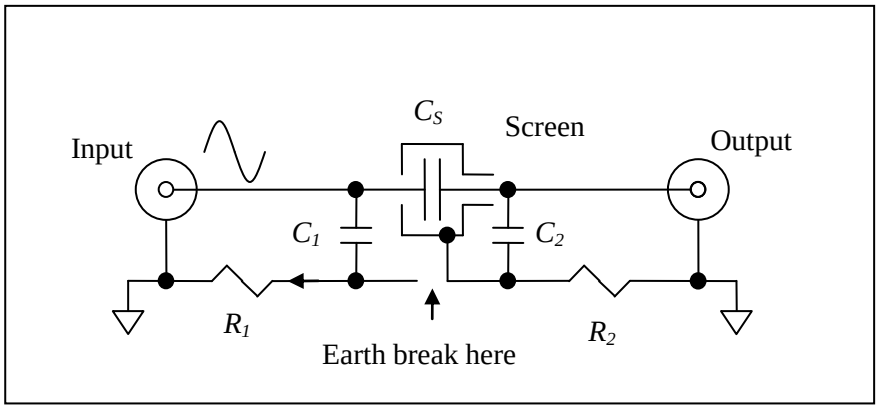


Fig. 2.3 Simplified equivalent circuit of a three terminal capacitor

A high voltage AC signal is applied to the input side while the output side is usually held at a virtual earth. The current flowing through cable capacitance C_1 must be routed back to earth through the small (but not negligible) resistance R_1. If the earth connection is not broken some of that current would flow down the screen of the output cable, setting up a voltage across R_2. This signal then produces a current flowing through the relatively large cable capacitance C_2. Whereas this error signal is substantially quadrature it can be sufficiently large to be a problem.

A more significant problem, at the very highest levels of accuracy, is the inductive component of the outer screens which would, if the earth is not broken, introduce an in-phase component via C_2.

3. Basic principles

Calculating capacitance from basic principles (e.g. Coulomb's law) is relatively simple and demonstrates that capacitance really is a fundamental concept (unlike resistance). Consider, for example, a conducting sphere with uniformly distributed charge. From the symmetry one would expect Coulomb's law to apply so that the electric force field is uniformly radial outside the sphere - exactly the same as that produced by a point charge at the centre (the "inverse square law"): -

$$\vec{E} = \frac{Q_S}{4\pi\varepsilon_0}\frac{\vec{r}}{r^3} \qquad r \geq a$$

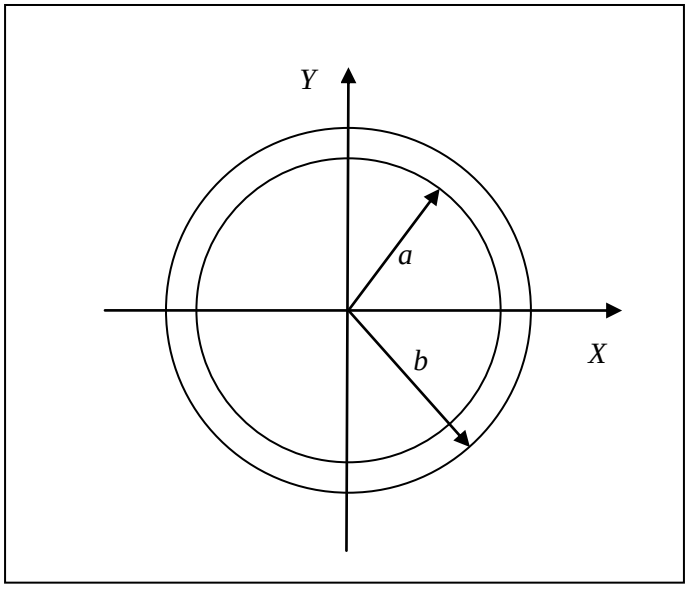


Fig. 3.1 Concentric spherical conductors

Consider now a second concentric sphere. The work done in moving an infinitesimal test charge from the outer sphere to the inner sphere along any radial is, therefore: -

$$W = \int F dr = \int_b^a \frac{Q_S Q_T}{4\pi\varepsilon_0 r^2} dr = \frac{Q_S Q_T}{4\pi\varepsilon_0}\left[\frac{-1}{r}\right]_b^a = \frac{Q_S Q_T}{4\pi\varepsilon_0}\left[\frac{1}{b} - \frac{1}{a}\right]$$

The voltage difference is defined as the increase in potential energy per unit of charge and, therefore: -

$$V = \frac{W}{Q_T} = \frac{Q_S}{4\pi\varepsilon_0}\left[\frac{1}{b} - \frac{1}{a}\right]$$

From the definition of capacitance: $$C = \frac{Q_S}{V} = \frac{4\pi\varepsilon_0 ab}{(b-a)}$$

In the limit that the radii become infinite but with a fixed gap of $g = b - a$ then the capacitance is: -

$$C = \frac{4\pi\varepsilon_0 ab}{g}$$

The total surface area of the spheres tends to the limit: $A_S = 4\pi r^2 = 4\pi ab$

The capacitance per unit area is, therefore: $$\frac{C}{A_S} = \frac{\varepsilon_0}{g}$$

If one takes a section of the spheres, small compared to the radius, then the plates, in the limit, become flat and parallel and the capacitance is proportional to the area of the plates according to the familiar expression for the capacitance of an "infinite parallel plate capacitor": -

$$C_{PP} = \varepsilon_0 \frac{A_{PP}}{g}$$

In practice the capacitance can be a little more than this due to the fringe field at the edge of the plates. This can be overcome, however, by the use of a guard electrode. This technique is often used in single capacitors and capacitance transducers [1]. Similar analysis can be applied to other geometries (e.g. cylindrical electrodes) with remarkably close agreement between calculated and measured capacitance.

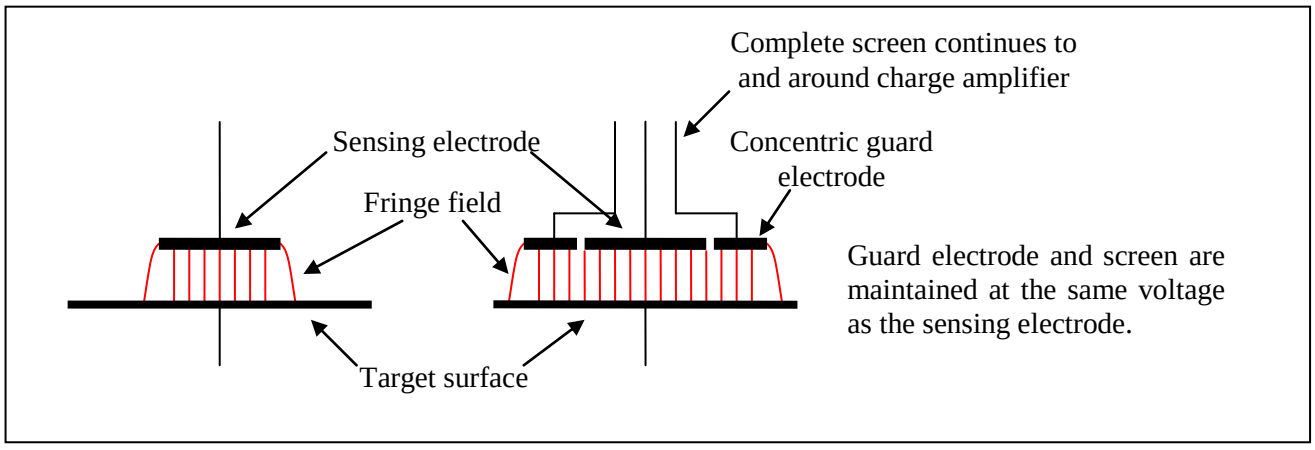


Fig. 3.2 The principle of the guard electrode

1. Part 1, monograph 5: "Variable gap capacitive displacement transducers".

4. The Thompson Lampard calculable capacitor

In the 1950s Thompson and Lampard discovered a mathematical property of parallel cylinders which makes it possible to calculate the capacitance per metre, based only on a fundamental constant (the permittivity of vacuum) [1].

This led to a practical device which could be used as a primary standard for capacitance [2].

A comparison between capacitors and resistors (including calculable resistors) at a known frequency (with a "quadrature" bridge) also provides an independent way of checking the consistency of the definition and practical implementation of the Farad and the Ohm [3]. At the time of writing (2022) there is a resurgence of interest and research in this area [4].

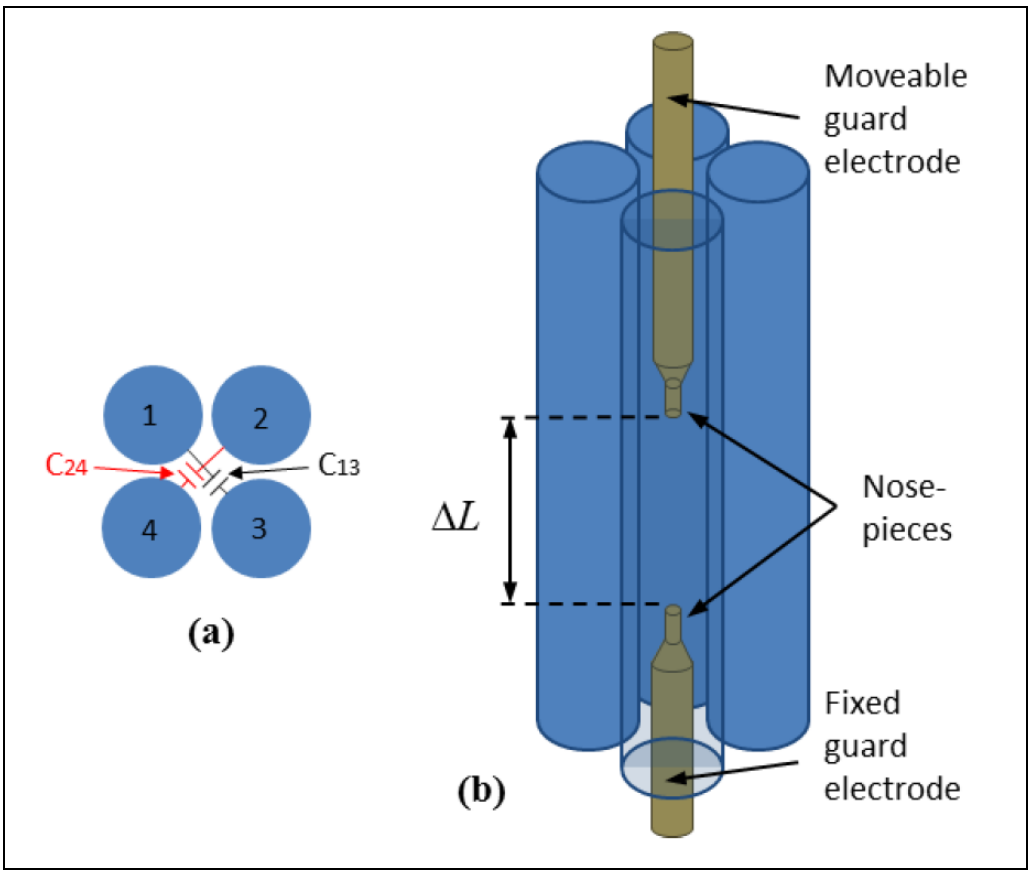


Fig. 4.1 The basic geometry of a calculable capacitor
(image courtesy Gournay et al)

1. Thompson A.M., Lampard D.G.: "A New Theorem in Electrostatics and its Application to Calculable Standards of Capacitance". Nature, 1956, **177**, 888
2. Clothier W.K.: "A Calculable Standard of Capacitance". Metrologia, 1965, **1**, 36
3. Thompson A.M.: "An Absolute determination of Resistance Based on a Calculable Standard of Capacitance". Metrologia, 1968, **4**, 1
4. Gournay, P., Fletcher, N., Robertsson L., Stock M.: "Progress on the Thompson-Lampard calculable capacitor project at BIPM". 17th International Congress of Metrology, 12001 (2015)

5. Industry standards

One of the most widely employed transfer standards is the IET type 1404

Fig. 5.1 Transfer standard gas capacitors (courtesy IET Labs Inc.)

SPECIFICATIONS

Calibration: An A2LA certificate of calibration is supplied with each capacitor, giving the measured direct capacitance at 1 kHz, 30 Vac and 23° ±1°C.

Model	Calibration Uncertainty	Adjustment Accuracy	Short Term Drift	Long Term Drift	Max Change with Orientation	Maximum Voltage	Dissipation Factor
1404-A 1404-B 1404-C	< ±6 ppm	±5 ppm to a capacitance about 5 ppm above the nominal value	less than 2 ppm	less than 20 ppm per year	10 ppm and completely reversible	750 V peak	< 10^{-5} at 1 kHz
1404-5 nF 1404-10 nF	±100 ppm	±500 ppm	less than 12 ppm	less than 40 ppm per year	15 ppm and completely reversible	100V peak	< 10^{-5} at 1 kHz

Temperature Coefficient of Capacitance: A measured value with and accuracy of ±1 ppm/°C is given on the certificate.

Model	Temperature Coefficient
1404-A 1404-B	2 ±2 ppm/°C from -20°C to +65°C
1404-C	5 ±2 ppm/°C from -20°C to +65°C
1404-5 nF 1404-10 nF	4 ±6 ppm/°C from -20°C to +65°C

Temperature Cycling: For temperature cycling over range from -20°C to +65°C , hysteresis (retraceable) is less than 20 ppm at 23°C.

Residual Impedance:

Model	C_H	C_L	L
1404-A 1404-B 1404-C	30 pF	28 pF	0.05 µH
1404-5 nF 1404-10 nF	130 pF	127 pF	0.1 µH

Terminals: Two BNC coaxial connectors (legacy locking G874 coaxial connectors are available). Outer shell of one connector is ungrounded to permit capacitor to be used with external resistor as a dissipation factor standard.

Mechanical:

Model	Dimensions	Weight
1404-A 1404-B 1404-C	(16.9 cm H x 17.2 cm W x 20.4 cm D (6.63" x 6.75" x 8")	3.9 kg 8.5 lb 6.4 kg 14 lb Shipping
1404-5 nF 1404-10 nF	(23 cm W x 23 cm H x 22 cm D) 9" x 9" x 8.5"	16 kg 35 lb 18 kg 40 lb Shipping

Fig. 5.2 An extract from the datasheet

Fig. 5.3 The internal construction of a 1nF transfer standard

Rotary capacitive displacement transducers

1. Introduction

Capacitive type rotary displacement transducers can be made with accuracies approaching 1ppm of range and sub-arc second resolution and repeatability. The component parts need to be made with reasonably high precision and assembled accurately but, otherwise, the technology required is well within the capabilities of a normal precision engineering workshop, equipped for turning, milling and grinding. The metal components (usually stainless steel) can be held together with epoxy resin or, for high temperature applications, with ceramic spacers for insulation.

Good quality ball bearings should be used, though these are readily available and not expensive.

N.B. Most other techniques for accurate measurement of angular displacement are incremental (e.g. incremental optical encoders). Capacitive transducers measure absolute angle.

Fig. 1.1 A capacitive type rotary displacement transducer (picture courtesy ASL Ltd)

1.1 The two main types

There are two main types of variable capacitance transducer for the accurate measurement of angular displacement: cylindrical and planar (flat plates). The cylindrical type has a diameter which is approximately the same as its length (see fig. 1.1). The flat plate type can be much thinner (axially) but has a larger diameter (radially). The basic transducer has limited angular range (up to ±45deg) but unlimited angle can be achieved by combining two transducers, with overlapping ranges, and switching between them.

1.2 The basic principle of operation

In both cases the stator usually consists of four energised electrodes, in symmetrically opposed pairs, and a single sensor electrode attached to the stator (the stationary element). The rotor (the rotating element) consists of a screen electrode, connected to 0V, usually via a rotating friction contact or slip ring. As the screen rotates it varies the area of the sensor electrodes exposed to the energised electrodes. The capacitance is proportional to the area of exposure and, therefore, varies in direct proportion to the angle. The symmetry of the technique makes both types insensitive to non-concentricity of the rotor and stator. Stator and rotor can be supplied separately and fitted to a shaft, which already has a bearing, without the need for special jigs, fixtures or gauges.

2. Ratiometric operation

The ratio $n = \dfrac{C_1 - C_2}{C_1 + C_2}$ is accurately proportional to angular displacement for both types of transducer. This ratio can be measured directly with a ratio transformer/inductive voltage divider bridge for the highest accuracy and resolution. In the following, for example, the centre of a centre-tapped transformer is driven with a voltage, V_C, which is derived from a reference voltage $\left(V_R = \dfrac{V_1 - V_2}{2}\right)$ using an inductive voltage divider (IVD).

At null balance: $$V_1 C_1 + V_2 C_2 = 0$$

$$\Rightarrow \quad (V_C + V_R)C_1 + (V_C - V_R)C_2 = 0$$

With a little algebra one finds: $$\frac{V_C}{V_R} = n = \frac{C_1 - C_2}{C_1 + C_2}$$

For more detail see the monograph “A 20 bit binary differential capacitance bridge” by the same author.

Reduced performance is possible with a low cost signal conditioner (see section 6).

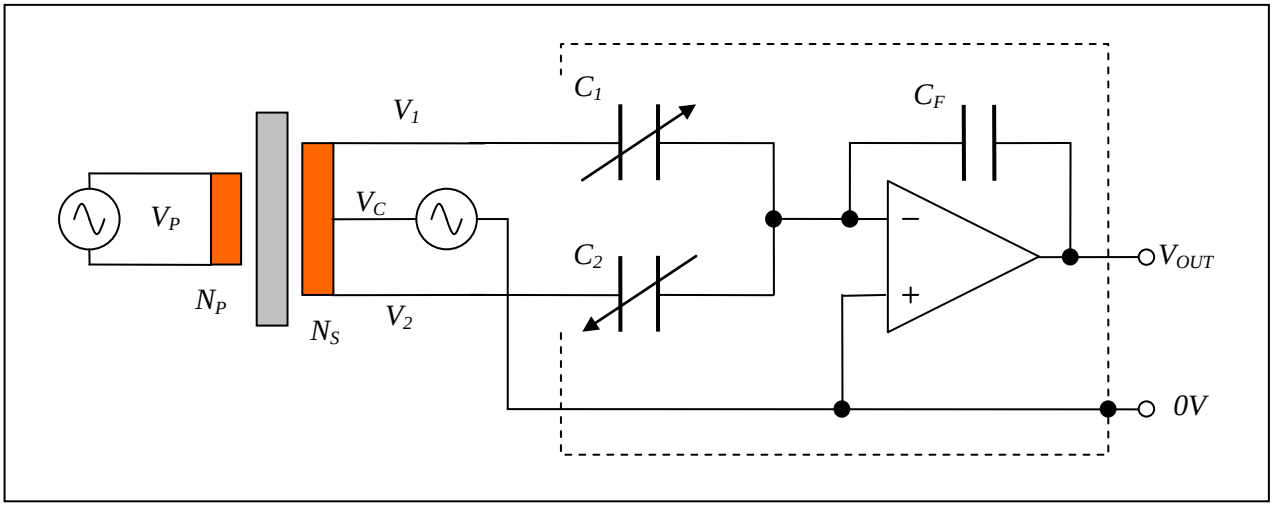


Fig. 2.1 A typical bridge configuration

Ratiometric operation has a number of advantages. Specifically: -

- The ratio of capacitance and, therefore, the overall measurement accuracy, is independent of certain manufacturing tolerances, specifically: actual dimensions.
- The ratio is independent of the dielectric medium, between the electrodes, as long as it is uniform. The effect due to small changes in the dielectric constant of air, due to humidity, is negligible. The transducer will also function satisfactorily while submersed in non-conducting liquid with a uniform dielectric constant.
- The ratio is independent of temperature, as long as it is uniform throughout the body of the transducer.
- Linearity depends on the uniformity of the cylindricity, flatness and/or circularity of the electrode surfaces. In practice these are quite easy to achieve using finish grinding and honing techniques.
- The rotor and stator must be reasonably accurately coaxial, though the design affords some self-compensation for small errors. Fairly basic jigs and fixtures usually suffice.

N.B. repeatability depends mainly on the quality of the ball bearing. With a really good bearing this can be down to the sub arc-second level (micro-radians).

Stability over time is also excellent so that any non-linearity can be measured and compensated.

3. The charge amplifier

Accurate measurement of the capacitance ratio relies on the sensor electrode being maintained at 0V by the action of feedback of an operational amplifier (aka "virtual earth"). See fig. 3.1.

N.B. This is called a charge amplifier because the output voltage is proportional to the electrical charge on the feedback capacitor.

The sensor electrodes, charge amplifier and the connection between them must be completely surrounded by a conducting screen connected to local 0V. This is to prevent electrical interference from external sources and also to prevent disturbance of the electric field and, therefore, the capacitances, due to the proximity of external objects. This is the concept of a "three terminal capacitance" – even though there is a considerable stray capacitance between the screen and the sensor electrode/virtual earth, there is no voltage difference and no current flows. The capacitances C_1 and C_2 are thereby accurately defined.

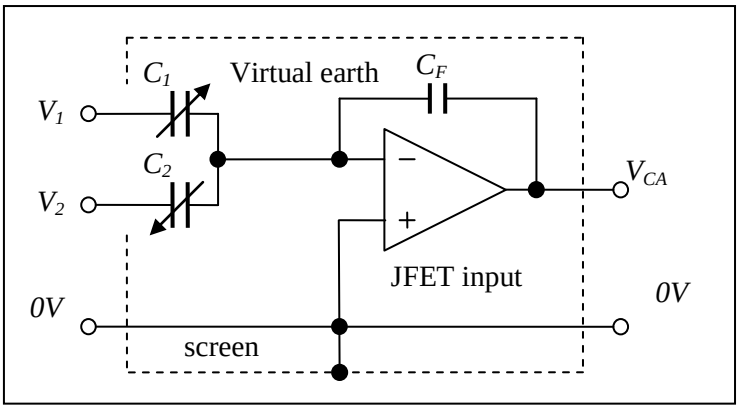


Fig. 3.1 A differential capacitive transducer and charge amplifier

At null balance: -

$$V_{CA} = \frac{V_1 C_1 + V_2 C_2}{C_F} = 0$$

$$\Rightarrow \quad V_1 C_1 + V_2 C_2 = 0$$

Lowest noise is best achieved with a composite op-amp – with a dual matched JFET pair front end [1].

1. Part 5, monograph 3: "Low noise JFET pre-amplifiers

4. The cylindrical type transducer

The following design was invented and patented by Peter Wolfendale. The first application was as part of a very high performance servo-mechanism, using a high speed limited angle torque motor to position a small mirror.

The sensing electrode is a simple cylinder, mounted on the stator body member (SBM). The energised electrodes also start life as a simple cylinder but end up as four electrically isolated segments of approximately 90 degrees each. Electrical connections are made via holes in the SBM with studs, spot welded onto the electrodes. The rotating screen electrode is part of a cylinder and is mounted on the rotating shaft. Reliable electrical connection to 0V is made via a rotating precious metal contact at the end of the shaft.

Although the manufacture of the cylindrical type is quite involved (i.e. expensive in low volumes), its high performance and robustness makes it an excellent choice for demanding, high value applications.

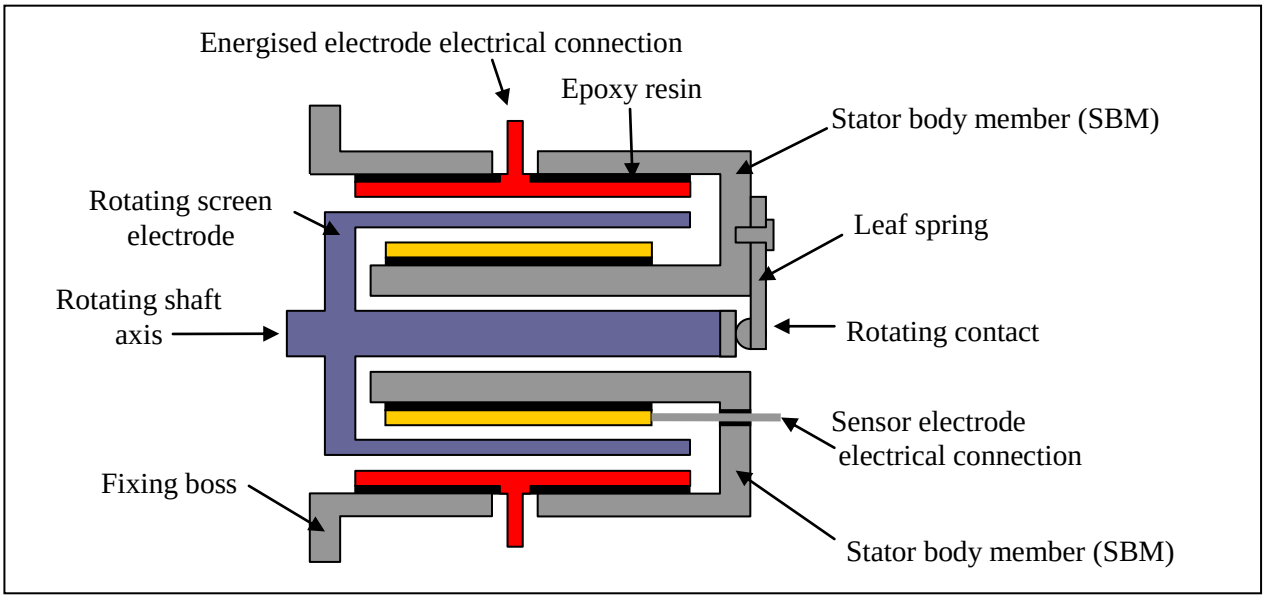


Fig. 4.1 Cylinder type transducer (radial view cross section)

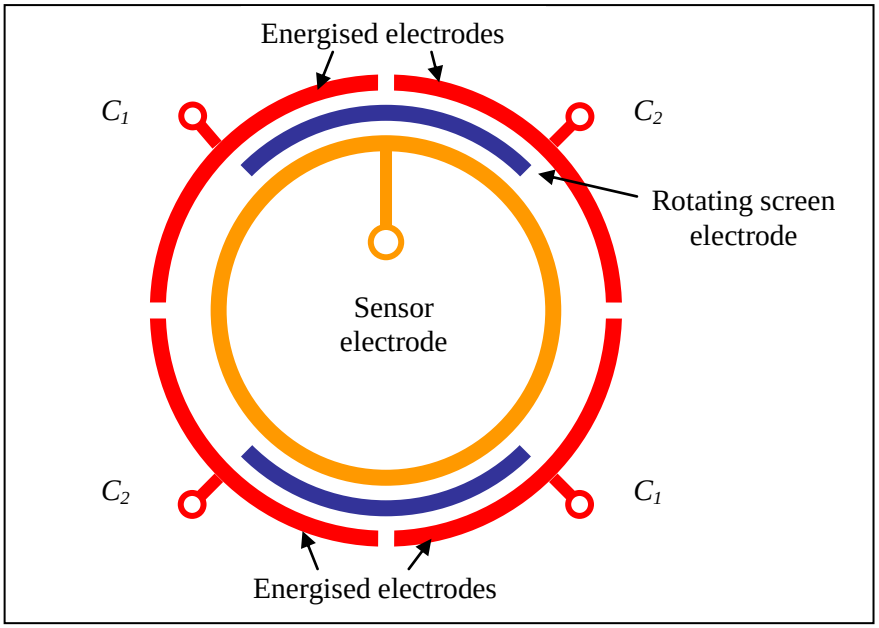


Fig. 4.2 Cylinder type transducer (axial view)

Opposing electrodes are connected in parallel to form two variable capacitors. As the rotor screen electrode rotates clockwise (see fig. 4.2) capacitance C_1 increases and C_2 decreases.

5. Plate type transducers

Plate types can be subdivided into quadrant types, like the cylindrical example above, or multi-segmented types.

5.1 A low cost limited angle example

This type is ideal for one-off projects or low volume applications as it can be fabricated from readily available materials (e.g. double-sided copper clad PCB material). If high accuracy is required then precisely machined metal parts with epoxy resin, ceramic and/or glass insulators may be used.

The following example, constructed from double sided PCB material, consists of two fixed plates for the stator (energised electrodes and sensor electrode) and a symmetrically cut rotating screen electrode between the two (the rotor). As with the cylinder type the screen varies the area between the energised electrodes and the sensor electrode.

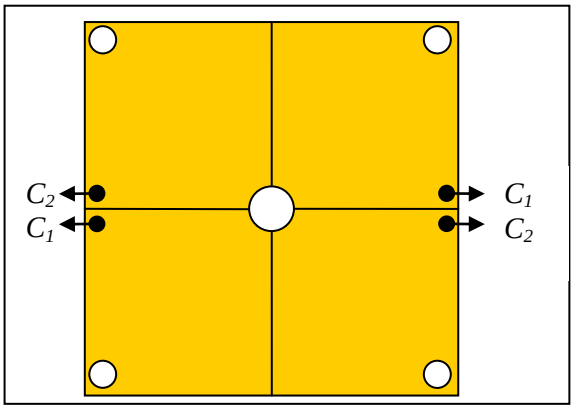


Fig. 5.1.1 Energised electrode plate

The energised electrode plate is chemically etched to leave four conducting quadrants. The dots indicate electrical connection to form the two capacitors. The hole in the centre allows the shaft to pass through. The holes on the periphery are for fixing.

The thickness, uniformity and straightness of the etching is not critical but the plate surfaces need to be flat.

The copper on the reverse side is usually left and connected to 0V to provide screening of the energised electrodes.

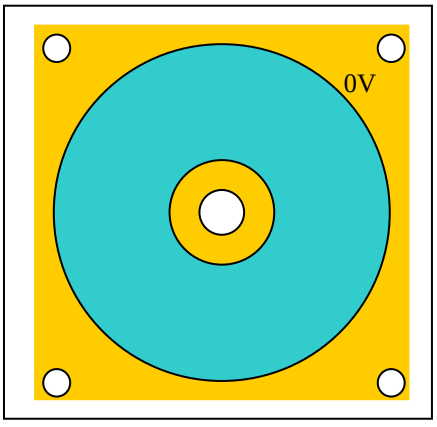


Fig. 5.1.2 Sensor electrode plate

The sensor electrode plate is also chemically etched to remove two thin circles of copper, leaving two concentric annular rings.
The inner ring and the copper remaining on the periphery are connected to 0V to act as “guard electrodes”.
The larger annular ring is the sensing electrode and is connected to the preamplifier. This can be done with a through-hole-plated “via” or a fine wire soldered to the surface which remains under the rotor screen. The copper on the reverse side is left and connected to 0V to provide screening.

It is advantageous if the preamp is mounted on the reverse side – as near to the sensing electrode as possible, to minimise ground capacitance.

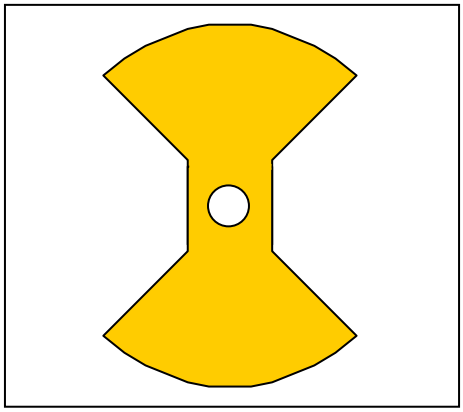


Fig. 5.1.3 The rotor plate

The rotor can also be fabricated from PCB material or thin sheet metal.

The shape and thickness do not have to be precise, as long as it is sufficiently flat and stiff.

The rotor must be reliably connected to 0V, either through the shaft and a good quality metallic ball bearing or a rotating contact as with the cylinder type.

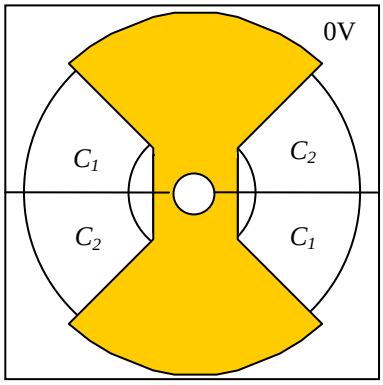


Fig. 5.1.4 The assembled transducer

The range of the transducer is limited to just less than the angle of the rotor screen due to the fringing effect - when the edge of the rotor approaches the boundary between C_1 and C_2. The maximum range is, therefore, a little less than ± 45 degrees, depending on the width of the gap. As a rule of thumb the fringe effect is negligible when the edge of the screen electrode is five times the gap width from the energised electrode boundary.

If one defines clockwise as the positive direction for angle (see fig. 5.1.4) then for an angle +45 degrees C_2 is totally obscured and C_1 totally exposed. The resulting ratio is: -

$$n(\theta = 45) = \frac{C_1 - C_2}{C_1 + C_2} = \frac{C_1 - 0}{C_1 + 0} = +1$$

Similarly for a counter-clockwise angle of -45 degrees, C_1 is totally obscured and C_2 exposed.

$$n(\theta = -45) = \frac{C_1 - C_2}{C_1 + C_2} = \frac{0 - C_2}{0 + C_2} = -1$$

The sensitivity of the transducer is, therefore: -

$$\frac{n(45) - n(-45)}{90} = \frac{2}{90} = \frac{1}{45} \text{ per degree} \quad \Rightarrow \quad n(\theta) = \frac{C_1 - C_2}{C_1 + C_2}(\theta) = \frac{\theta}{45}$$

Clearly the cylindrical type has the same slope.

The width of the rotor does not affect the sensitivity. Fig. 5.1.5 illustrates two cases: with rotor widths of 90 degrees and 45 degrees. The slope is the same but the range is reduced for the latter (maximum ratio of ±0.5). It is often useful to reduce the size of the rotor if the maximum range is not required – to minimise its moment of inertia.

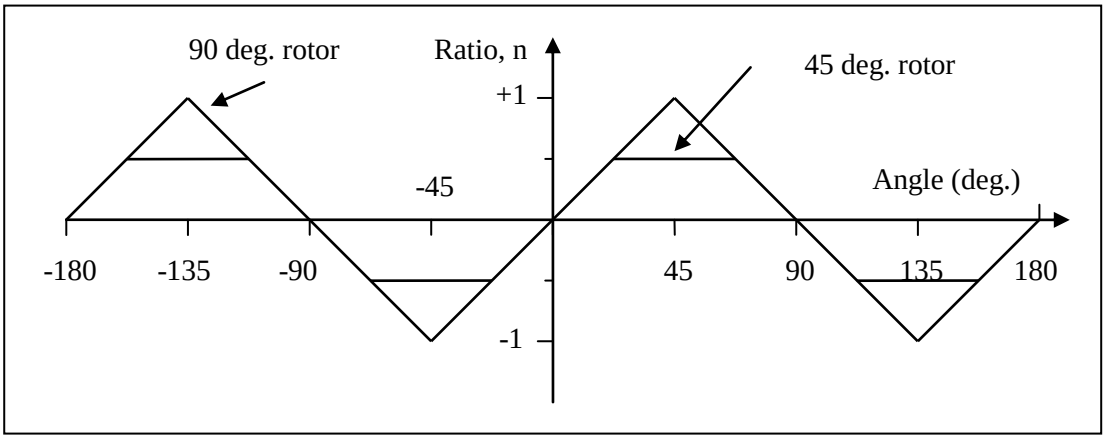


Fig. 5.1.5 Ratio versus angle in degrees for a quadrant type transducer (plate or cylindrical)

Note that the ratio does not depend on the actual diameter of the electrodes or the size of the gap. In practice the insulating gaps between the electrodes need to be as thin as possible and the sensor electrode as circular as possible. Deviations from the ideal formula are small and highly repeatable and so the transducer can be calibrated and any non-linearity compensated.

5.2 A multi-segmented design

It is possible to reduce the size of the segments and increase their number while retaining the balanced structure. Using more segments reduces the accuracy required of the ratio measurement. Fig. 5.2.1 is an example of a multi-segmented type. An absolute range of 360 degree can be achieved with two transducers, separated by one quarter of a cycle.

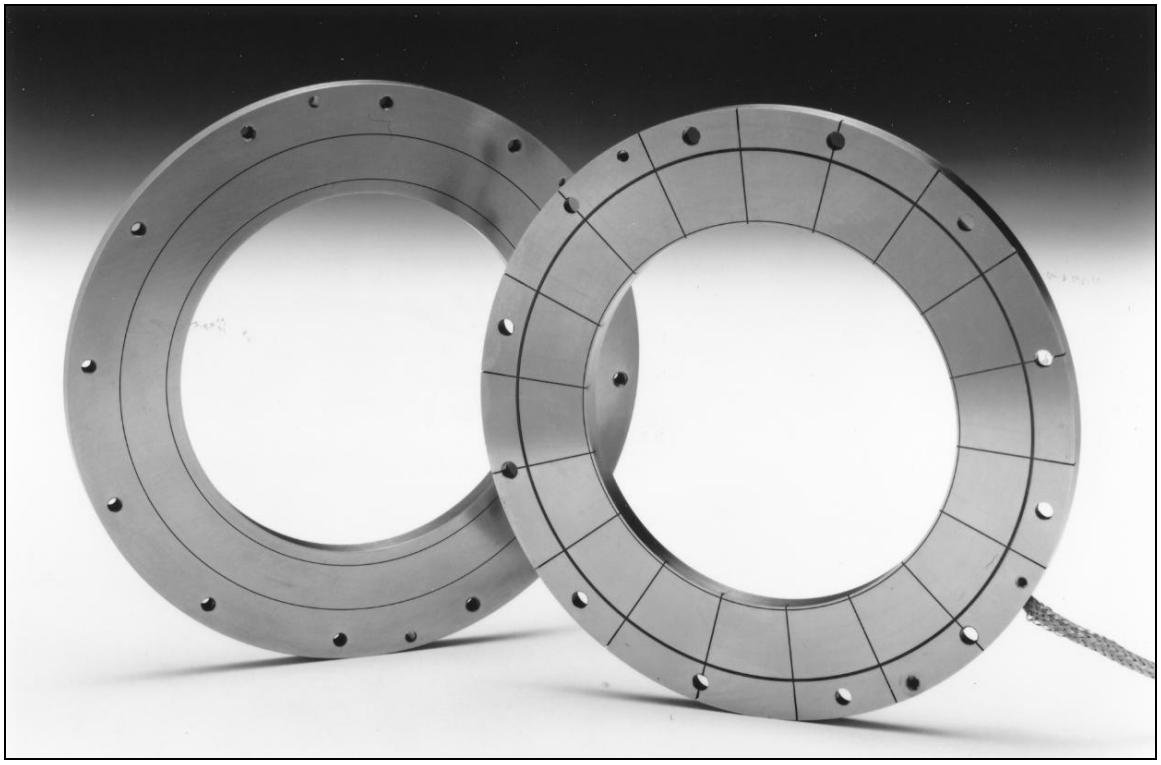


Fig. 5.2.1 A high accuracy multi-segment transducer (picture courtesy ASL Ltd)

There are 18 electrode segments of 20 degrees each (9 pairs with one cycle per pair)). A rotor screen of the same size (each side) provides a maximum range of ±10 degrees. When transducer 1 gets to, say, +6 deg. the measurement is switched to transducer 2, which is then at -4 degrees of its range. As the rotor continues the second transducer gets to 6 deg. of its range (16 deg. absolute) and the measurement is switched back. The same overlap is applied in the reverse direction. The hysteresis is necessary to avoid rapid switching back and forth on the boundary. Hysteresis is easily implemented using a microcontroller.

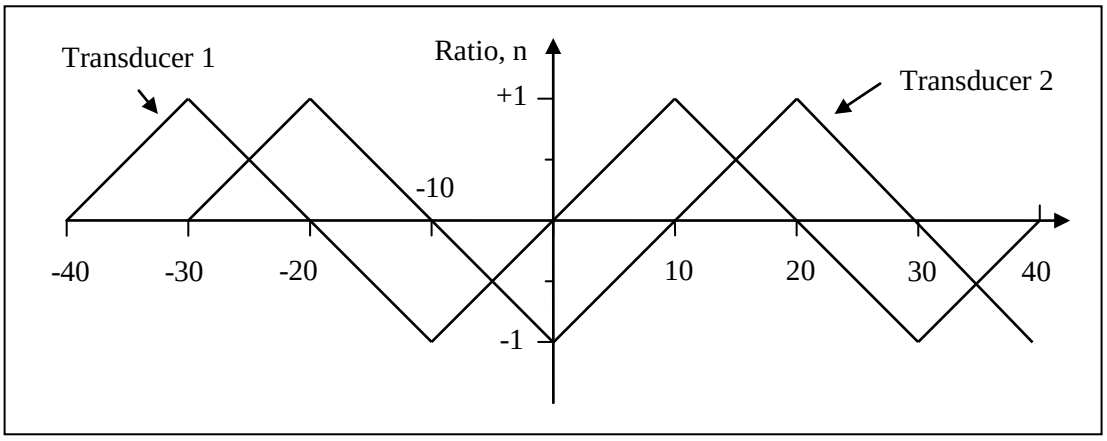


Fig. 5.2.2 The principle of overlapping ranges

Each transducer covers 9 cycles of 20 degrees per cycle (180 deg. each). The technique is inherently stable and repeatable with very high resolution and it is possible, therefore, to calibrate such a device to achieve an accuracy of sub-arc second (about 1ppm) with a modest ratio transformer.

6. Low cost signal conditioning

Remarkably high accuracy can be achieved without large and expensive IVD or ratio transformer techniques. The signal conditioner, fig. 6.1, for example, employs negative feedback in a simple and elegant null balance circuit.

Semiconductor switches (MOSFETs) convert the output DC voltage and a reference DC voltage into square wave signals which drive the energised electrodes (this usually referred to as a "modulator"). The output from the charge amplifier is amplified, filtered and converted to DC by the synchronous rectifier [1] (also MOSFET switches). The high gain block (HGB [2]) is an integrator (and, possibly, one or more one-plus-integrators in series). The output ramps up or down until the output of the synchronous rectifier is zero. The action of the negative feedback, therefore, is to achieve a null balance. The result is a high level DC output voltage, related to capacitance ratio as required: -

$$\frac{V_{OUT}}{V_R} = \frac{C_1 - C_2}{C_1 + C_2}$$

The only critical parts of the circuit are the modulator and the synchronous rectifier. The overall gain inside the loop affects the closed loop bandwidth and stability but can vary by 10% with no problem. The effect of any DC error due to the HGB is reduced by the AC gain before it.

The modulator switches and synchronous rectifier can operate accurately up to a frequency of 100kHz or more, offering high speed as well as high accuracy (there is a trade-off here). An operating bandwidth of 1kHz is readily achievable.

The closed loop bandwidth and frequency response are determined by the type of HGB and its frequency response. A simple integrator, for example, results in a first order low pass response. A two-stage HGB results in a response the same as a two-stage low pass filter (i.e. lower phase error). A three-stage HGB results in a response the same as a three-stage low pass filter (i.e. low phase and magnitude error).

The energising voltage, V_R, may be derived from a very stable reference (e.g. a 10V regulator chip) for a simple DC output. Alternatively it is possible to use a ratiometric type digital voltmeter to measure the ratio $\frac{V_{OUT}}{V_R}$ directly, thus eliminating any variations in V_R over time (including low frequency noise) and temperature. An accuracy of 0.001% (10ppm) can be achieved using a readily available high quality DC voltmeter.

For more details on signal conditioning see part 5.

1. Part 5, monograph 3: "Low noise JFET pre-amplifiers".
2. Part 4, monograph 1: "High gain blocks"

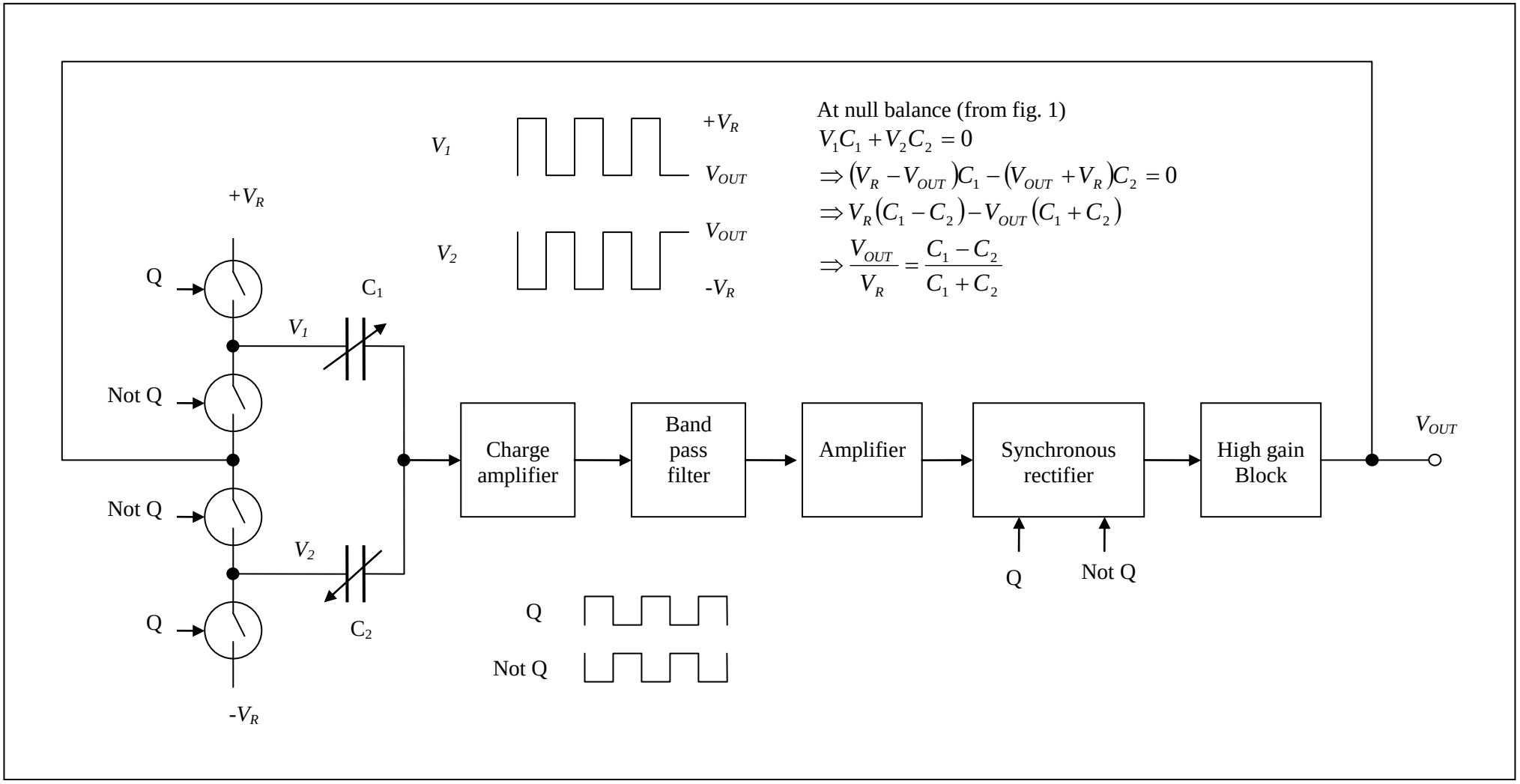


Fig. 6.1 Ratiometric signal conditioner employing feedback

7. Noise analysis

The resolution (smallest detectable change in angle) is usually limited by the noise voltage, V_N, generated at the input of the charge amplifier. The effect of this noise is increased by the "noise gain" of the charge amp, G_N. One can choose to place the noise voltage in series with either of the inputs. When in series with the non-inverting input it is easy to see that the noise gain is that of a non-inverting amplifier (see fig. 7.1 and imagine V_1 and V_2 connected to 0V). The noise gain is, therefore, largely determined by the stray capacitance to ground (in the transducer, the charge amp and, most significantly, any interconnecting cable). It is important, therefore, to keep this to a minimum - hence the reason for locating the charge amp as near as possible to the transducer. If the contribution from the noise current [1] is negligible (usually the case) the output of the charge amplifier is: -

$$V_{CA} = \frac{V_1 C_1 + V_2 C_2}{C_F} + V_N G_N$$

With: $$G_N = 1 + \frac{C_1 + C_2 + C_G}{C_F} \approx \frac{C_G}{C_F}$$

The approximation is valid in most cases as the ground capacitance is usually much larger than the transducer capacitance.

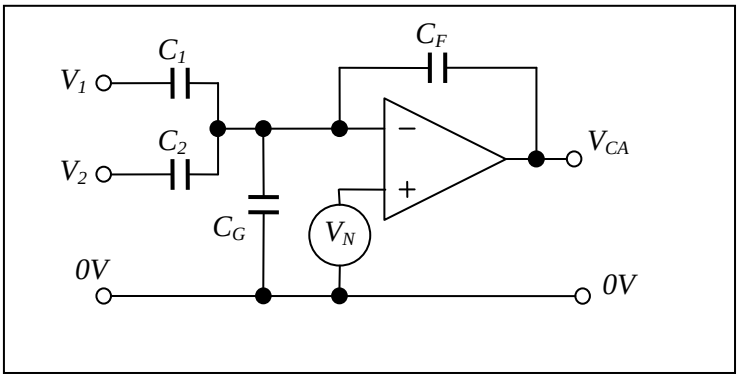


Fig. 7.1 Noise analysis

If one employs the approximation the algebra is simplified and the result will be sufficiently accurate in most cases. The action of the feedback loop is to adjust V_1 and V_2 until the output is zero, therefore: -

At null balance: $$V_{CA} = \frac{V_1 C_1 + V_2 C_2}{C_F} + V_N \frac{C_G}{C_F} = 0$$

$$\Rightarrow V_R (C_1 - C_2) - V_{OUT} (C_1 + C_2) + V_N C_G = 0$$

The output is, therefore: -

$$V_{OUT} = V_R \frac{C_1 - C_2}{C_1 + C_2} + V_N \frac{C_G}{C_1 + C_2}$$

The first component is due to the variation in angle (as before) but the second element is due to the noise.

This appears as random fluctuations in the output signal and is usually specified in terms of root mean square (RMS) voltage.

1. Part 5, monograph 3: "Low noise JFET pre-amplifiers".

The resolution in terms of angular displacement with the known value of *k* is: -

$$V_{OUT(NOISE)} = V_N \frac{C_G}{C_1 + C_2} = V_R k \theta_N \qquad \text{with} \qquad k = \frac{1}{45}\,\text{deg.}^{-1} \quad \text{or} \qquad k = \frac{4}{\pi}\,\text{rad}^{-1}$$

The noise, in terms of angle (in radians) is, therefore: -

$$\theta_N = \frac{\pi}{4}\frac{V_N}{V_R}\frac{C_G}{C_1 + C_2}$$

Example: -
Sensor electrode: inner radius 2cm and outer radius 5cm operating with a gap of 2mm.
Charge amp with input noise level of 5nV (RMS) in 1Hz of bandwidth.
Ground capacitance; C_G = 100pF
Reference voltage: V_R = 10V
Permittivity constant: $\varepsilon_0 = 8.85 \times 10^{-12}$ Fm^{-1}

For a parallel plate capacitor in the form of an annular ring: $C = \frac{\varepsilon_0 A}{d} = \frac{\varepsilon_0}{d}\left(\pi r_1^2 - \pi r_2^2\right)$

Half of the capacitance is obscured by the screen so: -

$$C_1 + C_2 = \frac{\pi}{2}\frac{\varepsilon_0}{d}\left(r_1^2 - r_2^2\right) = \frac{\pi}{2}\frac{8.85\times10^{-12}}{0.002}\frac{(25-4)}{10^4} \approx 15\,pF$$

The RMS angular resolution in 1Hz of bandwidth is, therefore: -

$$\theta_N = \frac{\pi}{4}\frac{V_N}{V_R}\frac{C_G}{C_1 + C_2} = \frac{\pi}{4}\times\frac{5\times10^{-9}}{10}\times\frac{100}{15} \approx 2.6\times10^{-9}\,\text{radians (RMS)}$$

This is about 0.5 milli-arc seconds! One could easily achieve ten times better using a ratio transformer, with ten times the operating voltage

Linear Capacitive Displacement Transducers

1. Introduction

Capacitive type linear displacement transducers can be made with an accuracy approaching 1ppm of range and sub-micron resolution and repeatability. The component parts need to be made and assembled with reasonably high precision but the technology required is well within the capabilities of a normal precision engineering workshop, equipped for turning, milling grinding and honing. The metal components (usually stainless steel) can be held together with epoxy resin or, for high temperature applications, with ceramic spacers for insulation. The most critical component is the linear ball bushing.

N.B. Most other techniques for accurate measurement of linear displacement are incremental (e.g. Laser interferometry and Moiré fringe). Capacitive transducers measure absolute displacement.

Fig. 1.1 A Linear capacitive displacement transducer

1.1 The principle of operation

The following design was invented and patented by Peter Wolfendale. Its principle of operation is very simple and elegant. Fig. 1.1.1 illustrates the main elements

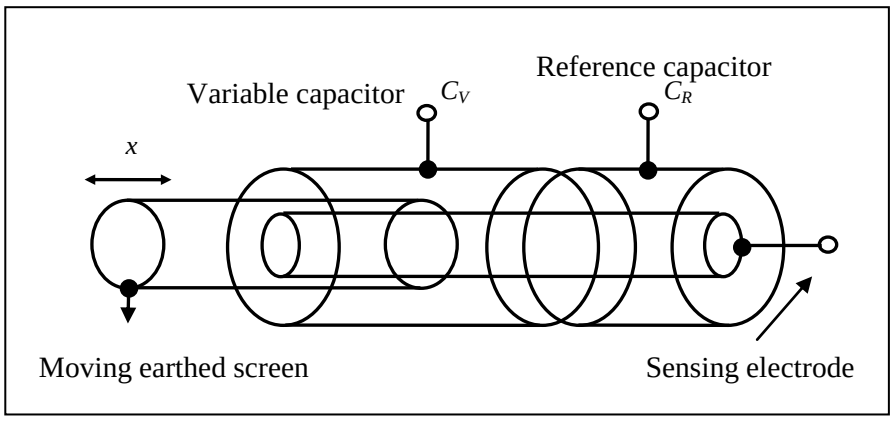


Fig. 1.1.1 The principle of operation

Two cylindrical capacitors are formed with concentric cylinders with a common co-axial central cylinder. A fourth concentric cylinder acts as an earthed screen to vary the area of one of the capacitors. The ratio of capacitance is accurately related to the linear displacement over a useful range: -

$$\frac{C_V}{C_R} = kx \qquad \text{with } k \text{ a constant.}$$

Linearity falls off when the edge of the screen electrode approaches the ends of the variable capacitor due to the fringe field effect. As a rule of thumb the fringe field effect is negligible when the edge is five times the gap width from the nominal limit.

The slope depends on the value of the reference capacitor and, therefore, the length of the reference capacitor electrode, with a small but stable difference due to fringe field effects. The other dimensions are the same, however, and the value of k is, therefore, approximately the reciprocal of the length: -

$$\frac{C_V}{C_R} \approx \frac{x}{x_R}$$

N.B. The extra length required due to the fringe field effect, the reference capacitor, the bearing and, often, the space at the end for connections and the charge amplifier, tend to make the overall length of the transducer considerably more than incremental types (e.g. Moiré fringe). This is, perhaps the one and only disadvantage of this type of transducer.

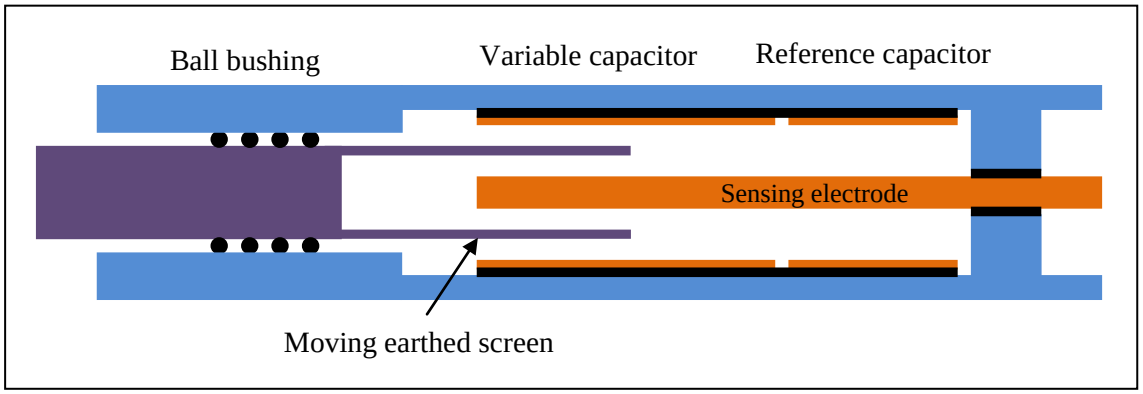


Fig.1.1.2 Cross section illustrating the construction (gaps exaggerated)

2. Ratiometric operation

Ratiometric operation has a number of advantages. Specifically: -

- The ratio of capacitance and, therefore, the overall measurement accuracy, is independent of certain manufacturing tolerances, specifically: actual dimensions. For interchangeability, at the level of 0.01%, it is sufficient for the internal diameters of the variable and reference electrodes to be accurately the same and the external diameter of the sensing electrode to be uniform and accurately co-axial. The first condition is easily achieved if the two electrodes are first machined as one piece and then divided.
- The ratio is independent of temperature, as long as it is uniform throughout the body of the transducer.
- Linearity depends on the uniformity of the cylindricity of the electrode surfaces. They must be reasonably accurately coaxial, though the design affords some self-compensation for small errors.
- The ratio is independent of the dielectric medium, between the electrodes, as long as it is uniform. The effect due to small changes in dielectric constant of air, due to humidity, is negligible. The transducer will also function satisfactorily while submersed in non-conducting liquid.

N.B. repeatability depends mainly on the quality of the ball bushing. With a really good bearing this can be down to the sub-micron level.

For the highest accuracy the ratio can be measured using a ratio transformer bridge technique.

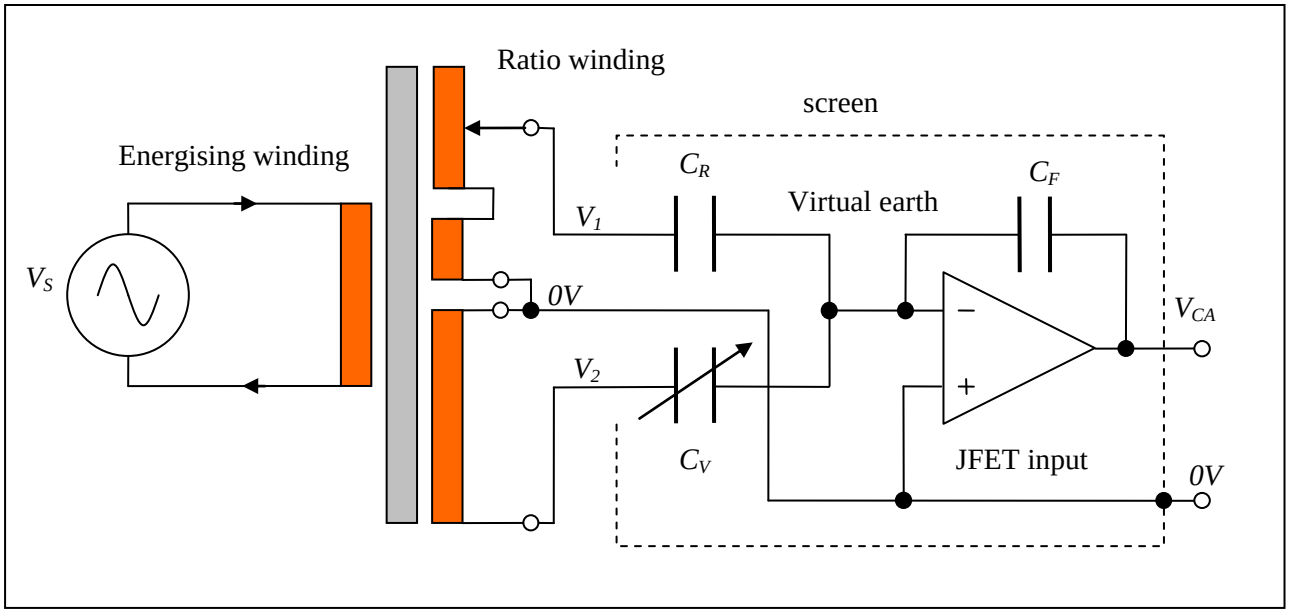


Fig. 2.1 Outline schematic of a typical ratio transformer/capacitance bridge

At null balance:

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} = 0$$

$$\Rightarrow n = -\frac{V_1}{V_2} = \frac{C_V}{C_R} = kx$$

N.B. The secondary windings are connected in anti-phase so that the ratio is numerically positive.

The linearity of the transducer falls significantly as C_V approaches zero and maximum, due to the fringe field effect. It is usual, therefore, to provide a small, fixed offset in order to make best use of the range of the variable winding. The accuracy of ratio from primary to secondary is not critical but the secondary and subsequent decades are. For more detail on basic ratio transformers see the monograph “Single stage inductors and transformers” by the same author.

3. The charge amplifier

Accurate measurement of the capacitance relies on the sensor electrode being maintained at 0V by the action of feedback of an operational amplifier (aka “virtual earth”).

N.B. This is called a charge amplifier because the output voltage is proportional to the electrical charge on the feedback capacitor.

The sensor electrode, charge amplifier and the connection between them must be completely surrounded by a conducting screen connected to 0V. This is to prevent electrical interference from external sources and also to prevent disturbance of the electric field and, therefore, the capacitances, due to the proximity of external objects. This is the concept of a “three terminal capacitance” – even though there is a considerable stray capacitance between the screen and the sensor electrode/virtual earth, there is no voltage difference and no current flows. The capacitances C_V and C_R are thereby accurately defined.

An effective screen could be maintained by connecting the transducer to the charge amplifier with a good quality co-axial cable, though it is better to place the charge amplifier (or at least the first stage - usually a low noise JFET) as close as possible to the sensor electrode. Best noise performance is achieved with minimum ground capacitance. See section 5 for more details.

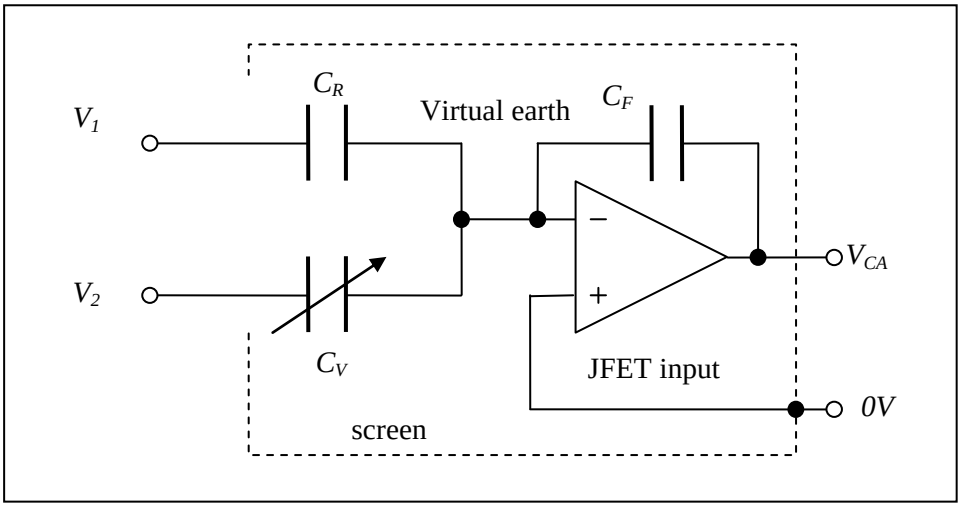


Fig. 3.1 A ratiometric capacitive transducer and charge amplifier

At null balance: $$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} = 0$$

4. A low cost signal conditioner

Remarkably high accuracy can be achieved without large and expensive IVD or ratio transformer techniques. The signal conditioner, fig. 4.1, for example, employs negative feedback in a simple and elegant null balance circuit.

Semiconductor switches (MOSFETs) convert the output DC voltage and a reference DC voltage into square wave signals which drive the energised electrodes (this usually referred to as a "modulator"). The output from the charge amplifier is amplified, filtered and converted to DC by the synchronous rectifier (also usually MOSFET switches). The high gain block (HGB), consisting of one or more integrators in series, ramps up or down until the DC error is zero. The action of the negative feedback is to achieve a null balance. The result is a high level DC output voltage, related to capacitance ratio as required: -

$$\frac{V_{OUT}}{V_R} = \frac{C_V}{C_R}$$

The only critical parts of the circuit are the modulator and the synchronous rectifier. The overall gain inside the loop affects the closed loop bandwidth and stability but can vary by 10% with no problem. The effect of any DC error due to the HGB is reduced by the AC gain before it.

The modulator switches and synchronous rectifier can operate accurately up to a frequency of 10kHz or more, offering high speed as well as high accuracy (there is a trade-off here). An operating bandwidth of 1kHz is readily achievable.

The closed loop bandwidth and frequency response are determined by the type of HGB and its frequency response. A simple integrator, for example, results in a first order low pass response. A two-stage HGB results in a response the same as a two-stage low pass filter (i.e. lower phase error). A three-stage HGB results in a response the same as a three-stage low pass filter (i.e. low phase and magnitude error).

The energising voltage, V_R, may be derived from a very stable reference (e.g. a 10V regulator chip) for a simple DC output. Alternatively it is possible to use a ratiometric type digital voltmeter to measure the ratio $\frac{V_{OUT}}{V_R}$ directly, thus eliminating any variations in V_R over time (including low frequency noise) and temperature. An accuracy of 0.001% (10ppm) can be achieved using a readily available high quality DC voltmeter.

For more details on signal conditioning see part 5.

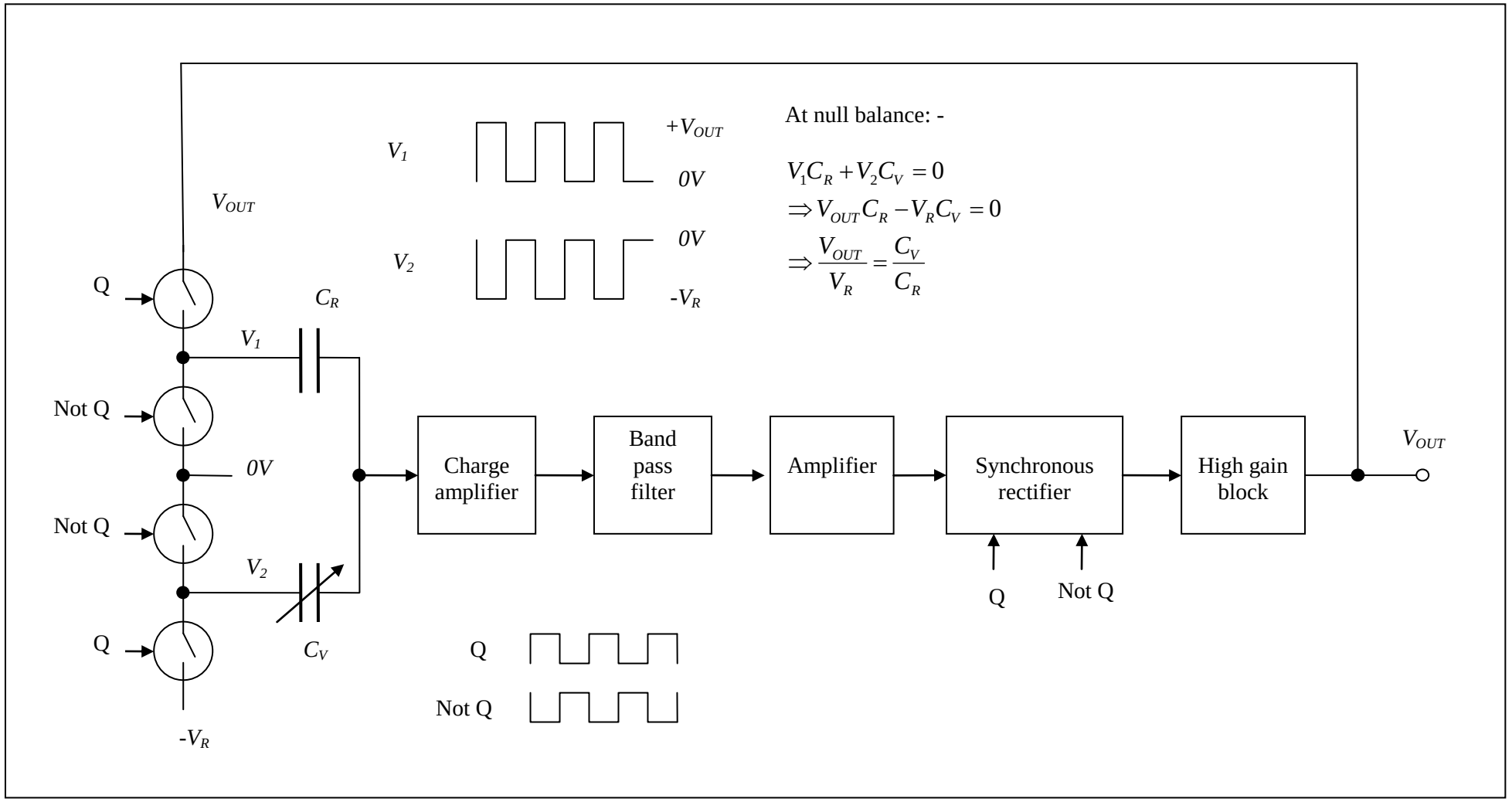


Fig. 4.1 A simple, low cost signal conditioner

5. Noise analysis

The resolution (smallest detectable change in displacement) is usually limited by the noise voltage, V_N, generated at the input of the charge amplifier. The effect of this noise is increased by the "noise gain" of the charge amp, G_N. One can choose to place the noise voltage in series with either of the inputs. When in series with the non-inverting input it is easy to see that the noise gain is that of a non-inverting amplifier (imagine V_1 and V_2 connected to 0V). The noise gain is, therefore, largely determined by the stray capacitance to ground (in the transducer, the charge amp and, most significantly, any interconnecting cable). It is important, therefore, to keep this to a minimum - hence the reason for locating the charge amp as near as possible to the transducer. If the contribution from the noise current is negligible the output of the charge amplifier is, therefore: -

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} + V_N G_N$$

$$\text{With:} \quad G_N = 1 + \frac{C_V + C_R + C_G}{C_F} \approx \frac{C_G}{C_F}$$

The approximation is valid in most cases as the ground capacitance is usually much larger than the transducer capacitance.

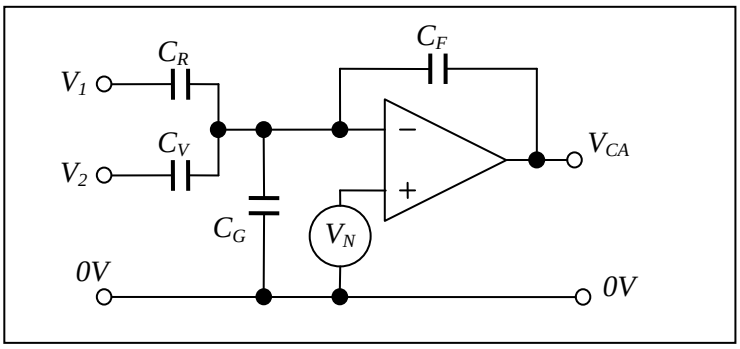


Fig. 5.1 Noise analysis

At null balance, therefore:

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} + V_N \frac{C_G}{C_F} = 0$$

$$\Rightarrow V_1 C_R + V_2 C_V + V_N C_G = 0$$

Divide by C_R and V_2: -

$$\Rightarrow \frac{V_1}{V_2} + \frac{C_V}{C_R} + \frac{V_N}{V_2}\frac{C_G}{C_R} = 0$$

$$\Rightarrow n = \frac{x}{x_R} + \frac{V_N}{V_2}\frac{C_G}{C_R} = 0$$

The second component is clearly the random error in the ratio measurement. In terms of displacement the noise is, therefore: -

$$x_N = x_R \frac{V_N}{V_2}\frac{C_G}{C_R}$$

The factor $\frac{C_R}{x_R} = \frac{4\pi\varepsilon_0}{\ln\left(\frac{b}{a}\right)}$ is the capacitance per unit length, for a cylindrical capacitor. The noise, in terms of displacement is, therefore:

$$x_N = \frac{V_N}{V_2}\frac{C_G}{C_R/x_R} = \frac{V_N}{V_2}\frac{C_G}{4\pi\varepsilon_0}\ln\left(\frac{b}{a}\right)$$

Example calculation: -
Sensor electrode: outer diameter a = 6mm
Variable and reference electrodes: inner diameter b = 10mm
Reference electrode length: x_R = 20mm
Charge amp with input noise level of 5nV (RMS) in 1Hz of bandwidth.
Ground capacitance; C_G = 100pF
Reference voltage: V_2 = 10V
Permittivity constant: $\varepsilon_0 = 8.85 \times 10^{-12}$ Fm^{-1}

For a cylindrical capacitor the capacitance per unit length is:

$$\frac{C_R}{x_R} = \frac{4\pi\varepsilon_0}{\ln\left(\frac{b}{a}\right)} = 218\,pFm^{-1}$$

The RMS angular resolution in 1Hz of bandwidth is, therefore: -

$$x_N = \frac{5\times10^{-9}}{10}\times\frac{100\,pF}{218\,pFm^{-1}} \approx 0.23nm \quad \text{(RMS)}$$

One could easily achieve ten times better using a ratio transformer, with ten times the operating voltage.

The contribution due to noise current of the charge amplifier is small because of the relatively large ground capacitance. Suitable JFETs with the required noise resistance are readily available. If the charge amplifier is incorporated into the transducer (small CG) then a higher noise resistance charge amplifier is recommended. For more detail see the monographs “Low noise JFET pre-amplifiers” and “An ultra-high input impedance high pass filter” by the same author.

Variable gap capacitive displacement transducers

1. Introduction

Whereas variable gap capacitive displacement transducers are not inherently accurate, they are frequently used as part of very accurate measurement and/or control systems. The capacitance is usually small (around a picofarad) but can be measured to the nearest attofarad (10^{-18}F). Sub-nanometre resolution is possible with stable mechanics.

Capacitance also allows non-contacting operation at the point of measurement and the transducer may be fabricated with robust materials (stainless steel or nimonic alloys, epoxy resin or ceramic insulation etc). This can be very useful in certain applications and especially in hostile environments (e.g. high temperature and corrosive atmosphere).

Fig. 1.1 Variable gap capacitive sensors (picture courtesy Physik Instrumente GmbH)

In the example above the transducer in the middle has a linearity better than ±0.1% over a range of 25-75µm and a resolution of 1nm (RMS) in a bandwidth of 10kHz with their ordinary signal conditioner (E-852.10).

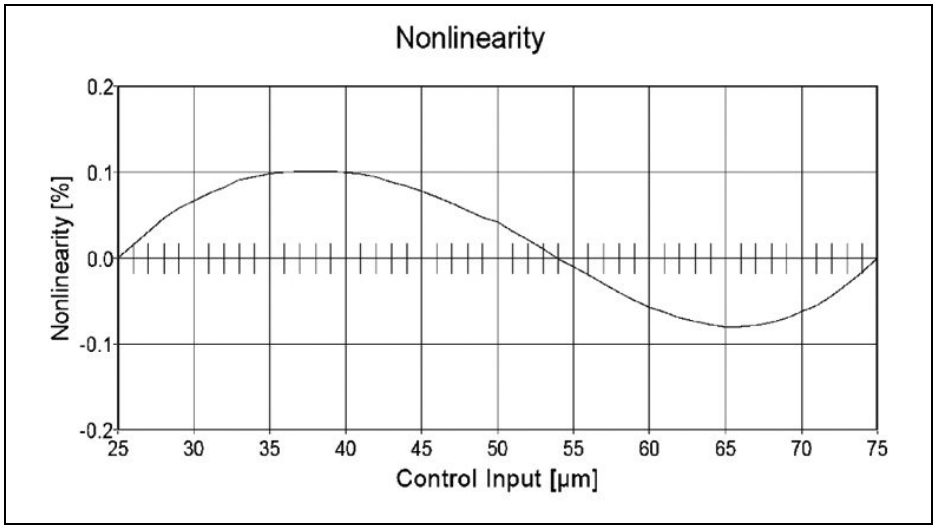


Fig. 1.2 A typical non-linearity curve (courtesy Physik Instrumente GmbH)

1.1 Profile measurement

A number of variable gap sensors may be incorporated in a single transducer for measuring small deviations from a simple geometric profile. Fig 1.1.1, for example, is a bore gauge for measuring internal diameters or cylindricity.

Fig.1.1.1 Four sensors arranged as a bore gauge

The sensors can be made very small. In fig. 1.1.2, for example, chemical milling of thin sheet metal was employed to construct a cone gauge for the precisely ground seat of a diesel fuel injector.

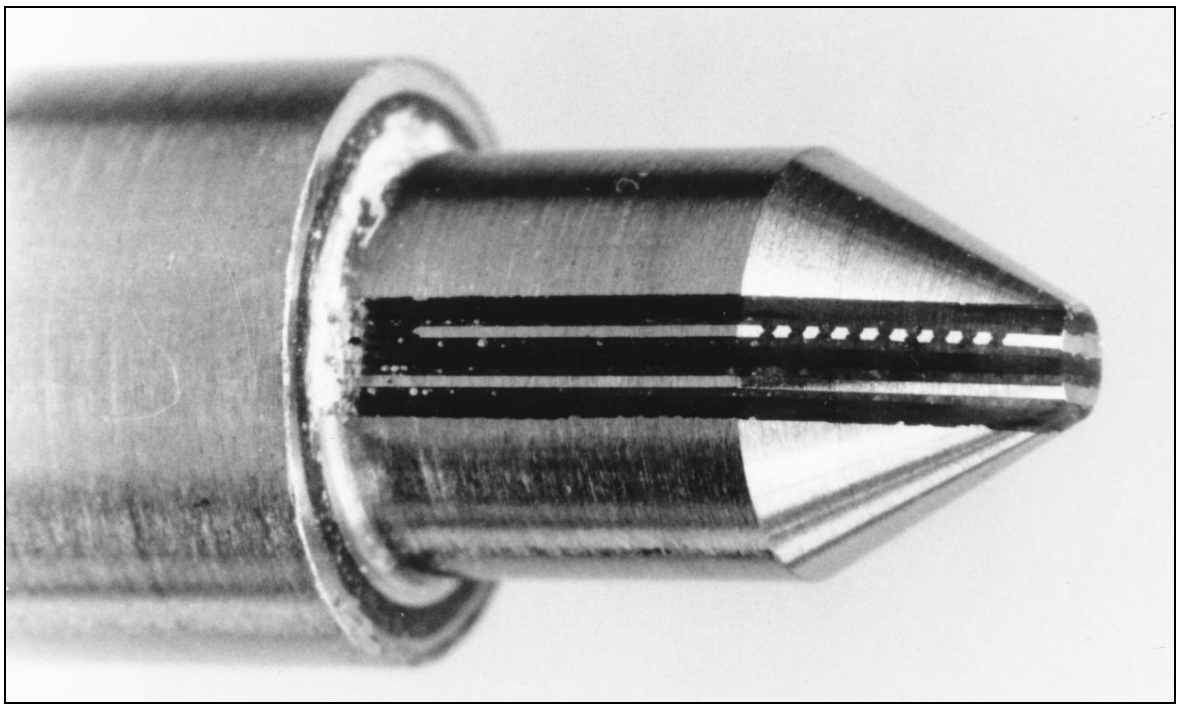


Fig. 1.1.2 Sixteen tiny senors (eight the other side) for gauging a conical profile

2. The guard electrode and charge amplifier

An active guard electrode is essential for high accuracy and stability. It is held at the same voltage as the sensing electrode in order to substantially remove the effect of the fringe field (to the edge of the guard electrode) and to shield the sensor from nearby moving objects. The action of the feedback in a charge amplifier performs this function, as well as providing a voltage output with low source impedance (no longer sensitive to interference). The following diagram depicts the effect of the fringe field:-

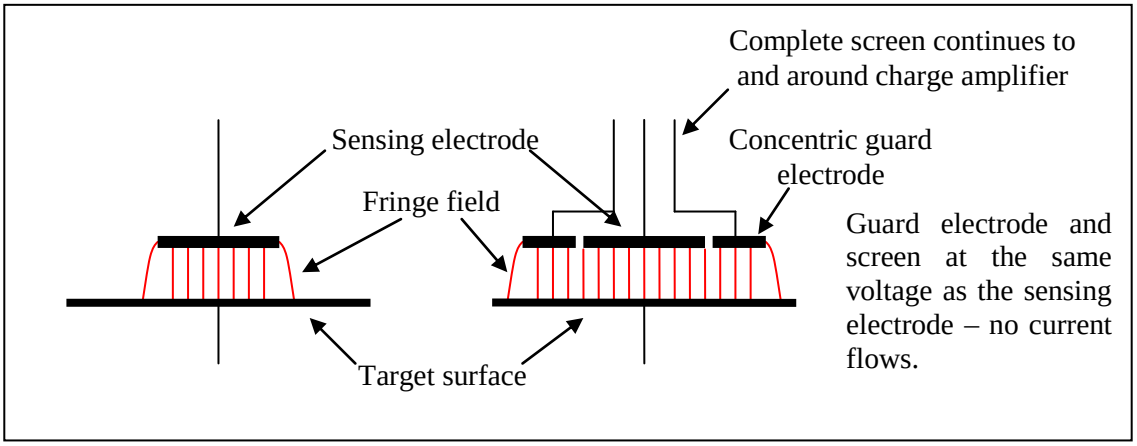


Fig. 2.1 Capacitive transducer with active guard electrode and screen

The target surface needs to be a reasonably good electrical conductor and it is necessary to make electrical connection to it, possibly through a bearing.

The capacitance of a parallel plate capacitor, with active guard electrode, is accurately related to the area and gap, thanks to the removal of the fringe capacitance [1]: -

$$C = \varepsilon_R \varepsilon_0 \frac{A}{x}$$

Where: -

$\varepsilon_0 = 8.85 \times 10^{-12} Fm^{-1}$

$\varepsilon_R =$ the relative permittivity of the dielectric medium, usually air.

A = the area in m^2

$x =$ the gap in m.

If one assumes that the area is constant then the sensitivity is: -

$$\frac{dC}{dx} = \varepsilon_0 A \frac{d}{dx}\left(\frac{1}{x}\right) = -\varepsilon_0 A \frac{1}{x^2} = -\frac{C}{x}$$

Alternatively, a more useful expression is: $$\frac{dC}{C} = -\frac{dx}{x}$$

I.e. a 1% increase in gap results in a 1% decrease in capacitance.

1. Part 1, monograph 2: “High accuracy single capacitors”. For a derivation of this formula see section 3.

2.1 The charge amplifier

The high gain of the op-amp and feedback operates to maintain zero volts difference between its two inputs [1]. If the target surface can be isolated from earth and driven with an AC signal the guard electrode is connected to earth. The sensor electrode is, therefore, maintained at earth potential without a direct connection, hence the commonly used term “virtual earth”. The rest of the metalwork around the transducer is also earthed though in a way to avoid earth loops [2]. Any current caused by energising the target surface must be carefully managed and routed back to the supply in a controlled way.

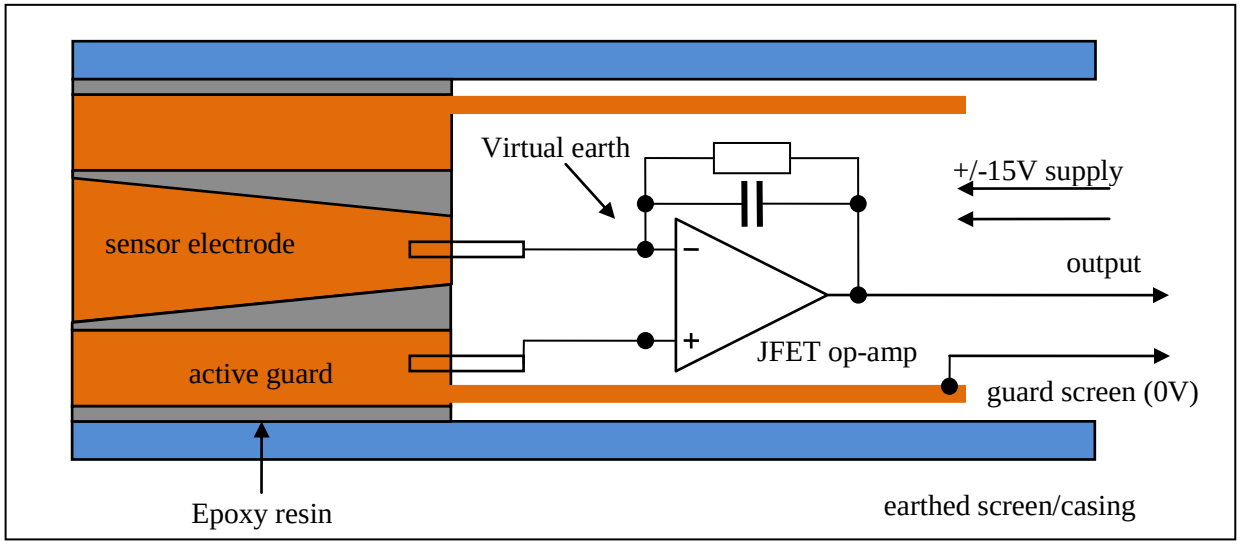


Fig 2.1.1 Typical construction of a variable gap capacitative displacement transducer with integral pre-amp

The main features include: -

- Overall earthed screen for mounting (often threaded) and EMC.
- Active guard with thin glue layer for accuracy and reduction of side influences.
- JFET pre-amp for low noise/high resolution.
- Conical sensor electrode for ease of manufacture and to minimize ground capacitance.
- Overall separate screen, which can be earthed if the guard electrode is driven (earthed target)

Analysis shows [1] that best noise performance and, therefore, resolution is achieved when the charge amplifier is mounted as near as possible to the transducer (minimum capacitance to ground from the virtual earth). In most cases there is sufficient space to incorporate a small op-amp or a single JFET transistor in the transducer housing. If high temperature is expected, however, the transducer needs to be remote and connected via good quality co-axial cable with 100% effective screening. A good choice is semi-rigid PTFE insulated co-axial cable often used by RF engineers.

1. Part 5, monograph 3: “Low noise JFET pre-amps”.
2. Part 1, monograph 2: “High accuracy single capacitors”. For the avoidance of earth loops see section 2.

3. The Ratio transformer/capacitance bridge

A ratio transformer bridge provides the most accurate and stable measurement of capacitance ratio. The reference capacitor is best positioned near to the transducer so that it shares the same air dielectric. Better still is to incorporate it as part of the mechanics in such a way as to provide compensation for temperature variations. Alternatively, materials with very low coefficients of thermal expansion (e.g. Zerodur® with ≈0.02ppm/degC) can be used. A pair of transducers may also be operated in opposing pairs and the differential ratio measured. For more detail on differential ratio measurement see the relevant monographs [1 and 2]. Repeatability, stability and, therefore, accuracy are potentially better than one part per million (1ppm) of range.

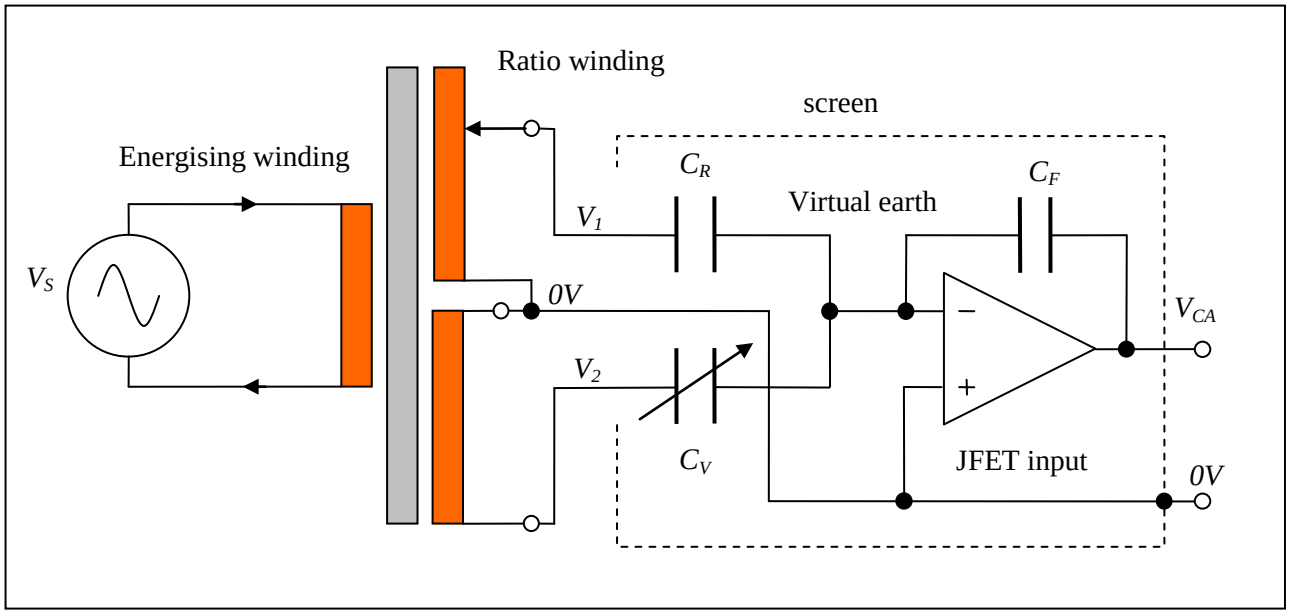


Fig. 3.1 Outline schematic of a typical ratio transformer/capacitance bridge

The analysis of the circuit is quite simple. The output of the charge amplifier is: -

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F}$$

At null balance, therefore:

$$\frac{V_1}{V_2} = \frac{C_V}{C_R}$$

The connections could be reversed so that the ratio is inversely proportional to transducer capacitance and, therefore, proportional to displacement. The only disadvantage is the variation in the gain which complicates the bridge balance algorithm.

In the above configuration: $V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} \Rightarrow \frac{dV_{CA}}{dC_V} = \frac{V_2}{C_F}$ which is constant.

With reference and variable interchanged: $V_{CA} = \frac{V_1 C_V + V_2 C_R}{C_F} \Rightarrow \frac{dV_{CA}}{dC_V} = \frac{V_1}{C_F}$ which varies.

In practice the displacement is easily calculated, including any corrections for non-linearity, using a digital computer**.**

1. Part 1, monograph 3: “Rotary capacitive displacement transducers”.
2. Part 3, monograph 8: “A 16 bit binary differential capacitance bridge”.

3.2 Earthing the target surface

It is often unavoidable that the target surface is at earth potential. This is not a major problem and a simple solution is to provide the charge amplifier with a separate floating power supply and a transformer coupled output. The interwinding capacitance of the output transformer needs to be kept to a minimum to ensure a high level of common-mode rejection.

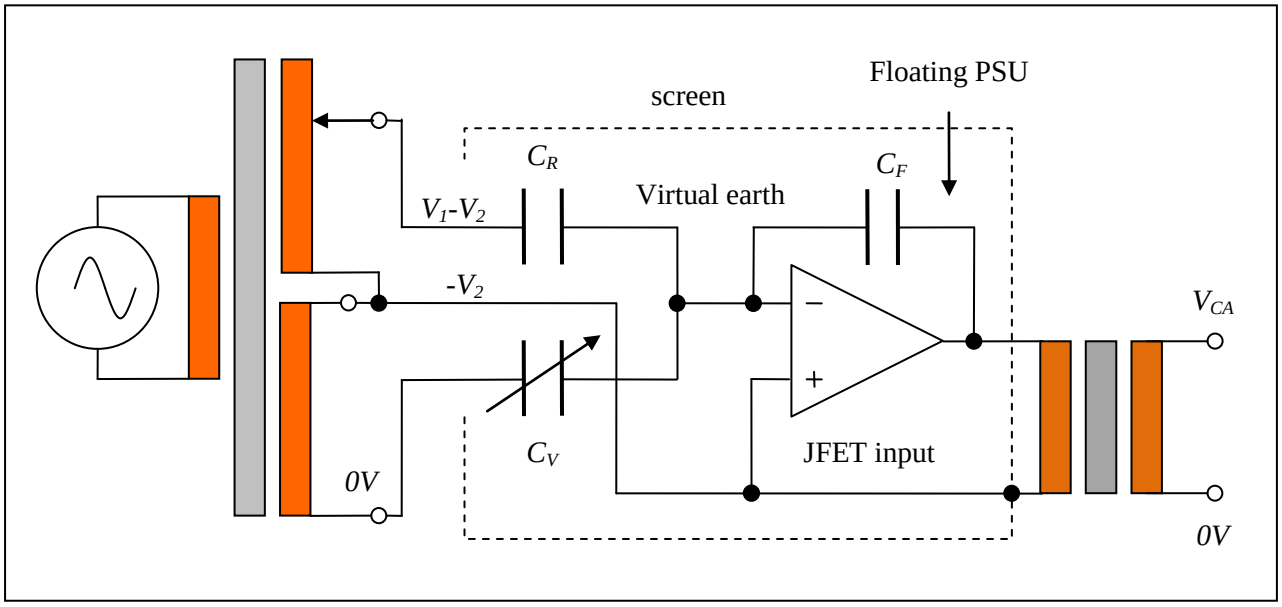


Fig. 3.2.1 A capacitance bridge with the target surface earthed

With a 1:1 output transformer the output is exactly the same: -

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F}$$

4. Noise analysis

The resolution (smallest detectable change in gap) is usually limited by the noise voltage, V_N, generated at the input of the charge amplifier. The effect of this noise is increased by the "noise gain" of the charge amp, G_N. One can choose to place the noise voltage in series with either of the inputs. When in series with the non-inverting input it is easy to see that the noise gain is that of a non-inverting amplifier (see fig. 4.1 and imagine V_1 and V_2 connected to 0V). The noise gain is, therefore, largely determined by the stray capacitance to ground (in the transducer, the charge amp and, most significantly, any interconnecting cable). It is important, therefore, to keep this to a minimum - hence the reason for locating the charge amp as near as possible to the transducer. If the contribution from the noise current is negligible the output of the charge amplifier is [1 and 2]: -

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} + V_N G_N$$

$$\text{With:} \quad G_N = 1 + \frac{C_V + C_R + C_G}{C_F} \approx \frac{C_G}{C_F}$$

The approximation is valid in most cases as the ground capacitance is usually much larger than the transducer and reference capacitance.

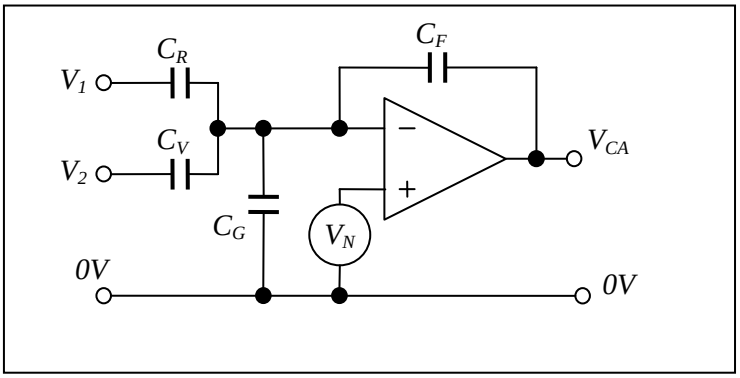


Fig. 4.1 Noise analysis

At null balance, therefore:

$$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} + V_N \frac{C_G}{C_F} = 0$$

$$\Rightarrow V_1 C_R + V_2 C_V + V_N C_G = 0$$

Divide by C_R and V_2: -

$$\Rightarrow \frac{V_1}{V_2} + \frac{C_V}{C_R} + \frac{V_N}{V_2}\frac{C_G}{C_R} = 0$$

The ratio $\left(n = -\frac{V_1}{V_2} \right)$ is defined so that it is numerically positive (V_1 and V_2 have opposite polarity): -

$$\Rightarrow n = \frac{C_V}{C_R} + \frac{V_N}{V_2}\frac{C_G}{C_R}$$

1. Part 5, monograph 3: "Low noise JFET pre-amplifiers".
2. Part 2, monograph 3: "An ultra-high input impedance high-pass filter"

The first component is the capacitance ratio, as expected, and inversely proportional to the gap. The second component is clearly the random error in the ratio measurement. In terms of transducer capacitance: -

$$\Rightarrow n = \frac{C_V + \delta C_V}{C_R} \qquad \text{with} \qquad \delta C_V = \frac{V_N}{V_2} C_G$$

The transducer capacitance is inversely proportional to the displacement and so, for small variations (see section 2 for the derivation): -

$$\frac{\delta C_V}{C_V} = -\frac{\delta x}{x}$$

Expressed in terms of displacement the random component is, therefore: -

$$\delta x = -x \frac{C_G}{C_V} \frac{V_N}{V_2}$$

V_N is usually specified as the RMS value and so the magnitude of the resolution is: -

$$\delta x = x \frac{C_G}{C_V} \frac{V_N}{V_2}$$

Example calculation: -

Transducer capacitance: C_V = 1pF
Operating gap: x = 1mm
Total ground capacitance: C_G = 50pF
Reference voltage : V_2 = 10V_{RMS}.
Charge amplifier noise voltage : $V_N = 5nV_{RMS}/\sqrt{Hz}$

$$\delta x = x \frac{C_G}{C_V} \frac{V_N}{V_2} \approx 10^{-3} m \times \frac{50\,pF}{1pF} \times \frac{5 \times 10^{-9} V_{RMS}}{10 V_{RMS}} \approx 25\,pm \qquad \text{(RMS in 1Hz)}$$

If the ground capacitance is very small the noise current may become an issue. For details see [1] and [2].

1. Part 5, monograph 3: “Low noise JFET pre-amplifiers”.
2. Part 2, monograph 3: “An ultra-high input impedance high-pass filter”

Low phase error capacitors and inductors

1. Introduction

Low phase error capacitors and inductors have a number of applications in high accuracy electronics. One important component in null balance bridges, for example, is used to generate a phase shift of precisely 90 degrees as part of a quadrature servo [1]. Other applications include high accuracy passive filters [2], active filters [3] and oscillators [3]. The best capacitors and inductors have phase errors of around 0.2mradians (2×10^{-4}). Whereas one might expect a capacitor to be the obvious (cost-effective) choice, with a wide range of values available, the most practical for the quadrature servo application turns out to be a low phase error ferrite pot core (with air gap) [4].

2. Polypropylene capacitors

The ideal capacitor has purely negative imaginary impedance. In practice, however, even high quality polymer dielectric capacitors have impedance with a real part, which can be modelled as a small series or large parallel resistance, which varies with frequency. The series model is used in this monograph as it simplifies the algebra: -

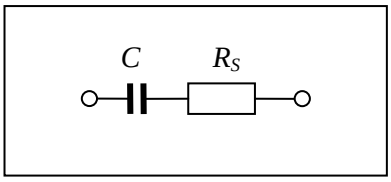


Fig. 2.1 Equivalent circuit for a practical capacitor

The impedance is, in the complex representation: $$Z = R_S + \frac{1}{j\omega C}$$

The ratio of the real part to the imaginary part is known as the "tanδ": the tangent of the acute angle to the negative imaginary axis of the complex number which represents the impedance: -

$$\tan\delta = \omega R_S C$$

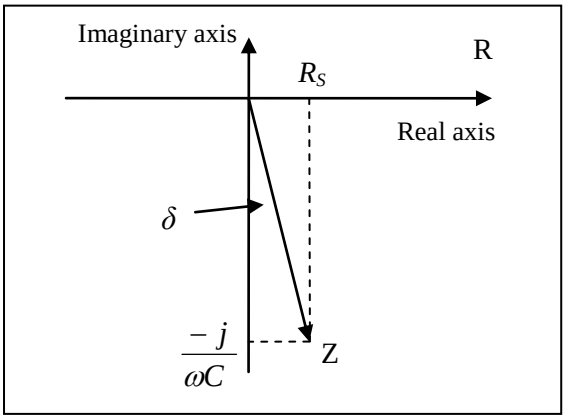


Fig. 1.3 The complex representation of a capacitor (real part exaggerated)

The angle is very small so that, to a very good approximation (in radians): $\tan\delta \approx \delta$

1. Part 5, monograph 1: "Null detectors – the basics".
2. Part 2, monographs 1 and 2: "Two-stage filters" and "Three-stage filters"
3. Part 6, monograph 3: "A circuit for measuring tanδ".
4. Part 4, monograph 4: "High accuracy amplifiers, integrators and differentiators". See section 5.

The phase error is constant at low frequency - consistent with a lower limit due to the intrinsic characteristic of the dielectric material. The mechanism is believed to be energy loss as the molecules are rotated, stretched and compressed (c.f. friction in a spring) [1]. At higher frequency the impedance of the capacitor drops so that the actual resistance of the connections and thin foil electrodes becomes significant. This hypothesis fits the available evidence: -

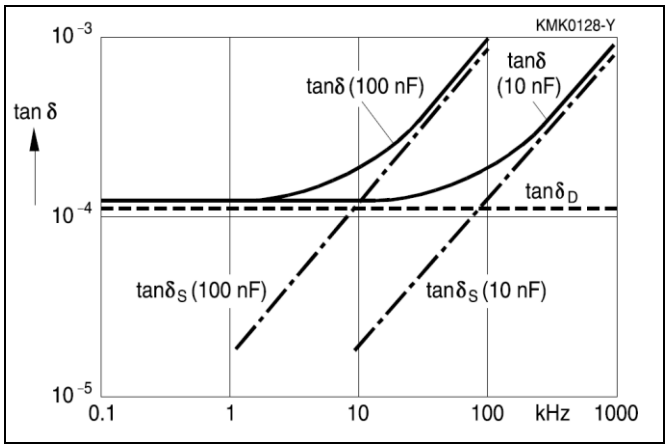


Fig. 2.2 Tanδ for polypropylene dielectric capacitors (courtesy EPCOS AG 2015 [1])

We are concerned, primarily, with low frequency behaviour and a suitable model is a complex capacitance: -

$$f < 1kHz \;\Rightarrow\; \delta = \omega R_S C \text{ is constant} \;\Rightarrow\; R_S = \frac{\delta}{\omega C} \;\Rightarrow\; Z = \frac{\delta}{\omega C} + \frac{1}{j\omega C} = \frac{1 + j\delta}{j\omega C}$$

In a typical circuit analysis one can, therefore, replace any symbol for an ideal capacitor, *C*, with the complex equivalent: -

$$C \to \frac{C}{(1 + j\delta)} \approx C(1 - j\delta)$$

1. Film Capacitors: General technical information, Sept 2015. See section 2.3.2

3. Low phase error ferrite transformers

Despite the extra labour cost of winding a transformer a ferrite pot core can provide a cost-effective means of implementing an accurate 90 degree phase shift of a sinusoidal voltage signal. A differentiator or integrator based on a low phase error capacitor and resistor would be possible but, for a quadrature servo, it would require a low phase error isolating transformer (i.e. two-stage) and be less cost-effective. The best choice is a gapped ferrite RM type pot core made with a manganese-zinc type material. For a brief review of magnetic circuit theory see the appendix.

Fig. 3.1 The RM type pot core and accessories

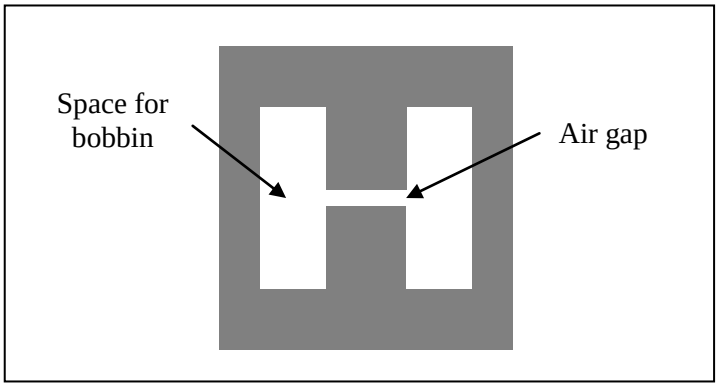


Fig. 3.2 The magnetic circuit of a gapped RM core

Whereas the intrinsic phase error of the best ferrite material is of the order 2mradians (2×10^{-3}) the presence of the air gap results in a reduction, by an order of magnitude, in the phase error at the expense of reducing the effective permeability of the magnetic circuit. The reluctance of each part of a magnetic circuit depends on the length, cross-sectional area and permeability [1]: -

$$R = \frac{L}{\mu_0 \mu_R A}$$

$\mu_0 =$ Permeability of the vacuum $4\pi \times 10^{-7}\, Hm^{-1}$

$\mu_R =$ Relative permeability

The total reluctance of the magnetic circuit is the sum of reluctances of each section: -

$$R_T = R_F + R_G = \frac{1}{\mu_0 \mu_R} \sum \frac{L_F}{A_F} + \frac{1}{\mu_0} \frac{L_G}{A_G} = \frac{1}{\mu_0 \mu_E} \frac{L_E}{A_E}$$

Where: -

R_F and R_G are the reluctances of the ferrite sections and air gap respectively.

L_F and A_F are the lengths and areas of the ferrite sections.

L_G and A_G are the length and area of the gap section.

μ_E, L_E and A_E are the effective permeability, length and area of the equivalent uniform magnetic circuit [2].

1. Part 3, monograph 2: “Single stage inductors and transformers”. See section 2.
2. E.g. a large diameter, small cross-section toroid resulting in uniform magnetic flux.

The reluctance allows one to calculate the magnetic flux for the given magneto-motive force (ampere-turns): -

$$MMF = NI = R_T \Phi$$

In practice the parameters specified and tested in production include the reciprocal of the reluctance (the "permittance"), the effective relative permeability of each ferrite component and the loss factor for each type of material. The permittance allows one to calculate the number of turns required for a given inductance and the loss factor provides, albeit approximately, the tanδ of the resulting inductance: -

$$A_L = \frac{1}{R_T} \quad \text{with} \quad L = N^2 A_L \quad \text{and} \quad \tan\delta_E = \frac{\tan\delta_I}{\mu_I}\mu_E$$

$\tan\delta_I$ and μ_I are the intrinsic phase error and relative permeability of the ferrite material.
The ratio $\tan\delta_I / \mu_I$ is called the "loss factor" by most manufacturers.

The validity of the latter equation can be demonstrated as follows: -

The phase error can be modelled as a complex permeability: $\mu_I \to \mu_I(1 + j\tan\delta_I)$

The effective reluctance is: $$\frac{1}{\mu_0\mu_E}\frac{L_E}{A_E} = \frac{1}{\mu_0\mu_R}\sum\frac{L_F}{A_F} + \frac{1}{\mu_0}\frac{L_G}{A_G}$$

The cross-sectional area of the circuit is approximately the same around the circuit and the length of the air gap is much smaller than the length of the ferrite part of the circuit: -

$$A_E \approx A \approx A_G \quad \text{and} \quad L_E \approx \sum L_F \quad \Rightarrow \quad \frac{1}{\mu_E} \approx \frac{1}{\mu_R} + \frac{L_G}{L_E} \quad \Rightarrow \quad \mu_E \approx \mu_R\left(1 + \mu_R\frac{L_G}{L_E}\right)^{-1}$$

The multiplying factor depends on the choice of gap and typically varies from 0.02 to 0.1. The tanδ is reduced by the same factor as the effective permeability: -

$$\left(1 + \mu_R\frac{L_G}{L_E}\right)^{-1} \approx k \quad \Rightarrow \quad \mu_E(1 + j\tan\delta_E) = k\mu_I(1 + j\tan\delta_I)$$

3.1 Example calculation

A widely available component is the RM10 core in P11 material (MMG: Magnetic Materials Group) with an air gap of 0.21mm and the following characteristics: -

$$A_L = 400\,nH/turn^2 \quad \mu_I = 2250 \quad \mu_E = 160 \quad \Rightarrow \quad k = \frac{\mu_E}{\mu_I} = \frac{160}{2250} \approx 0.071 \quad \text{and} \quad \frac{\tan\delta_I}{\mu_I} \approx 1.5\times10^{-6}$$

The effective tanδ is comparable to the best available polypropylene capacitors: $\tan\delta_E \approx 2.4\times10^{-4}$

The bobbin can comfortably accept 50 turns for both primary and secondary (on separate sections to minimise capacitance). Each section has ≈ 20mm^2 so that 0.3mm diameter enamelled copper wire would be satisfactory.

The self inductance of the primary and mutual inductance (primary to the secondary) are: -

$$L = N_P^2 A_L = 1mH \quad \text{and} \quad L_M = N_P N_S A_L = 1mH \quad \Rightarrow \quad \omega L = 0.157\Omega \quad \text{at 25Hz}$$

3.2 A practical differentiator circuit

For the detailed analysis see the monograph "High accuracy amplifiers, integrators and differentiators" [1].

The basic idea is to employ the transformer primary as the feedback element in an inverting configuration with a high gain block so that the input (sine wave) voltage is converted into a current.

The transformer output voltage is proportional to the rate of change of the flux/current and the mutual inductance of the transformer (Faraday's law). It is also, therefore, proportional to the rate of change of the input voltage (i.e. differentiation). The resistance of the primary winding is relatively unimportant as the current is accurately in-phase with the input voltage. The error contribution due to the high gain block is negligible and the phase accuracy is largely determined by the pot core [1].

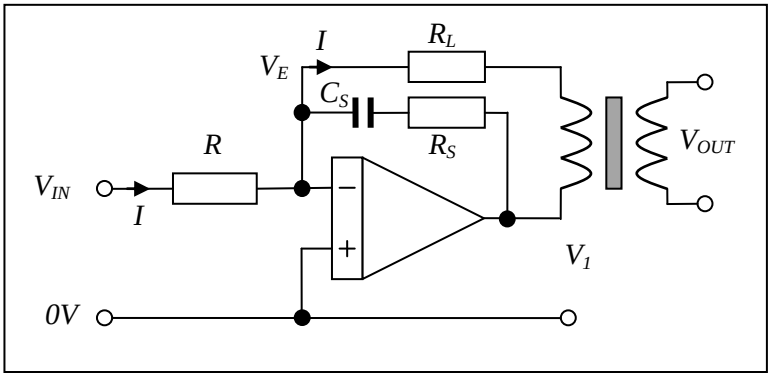


Fig. 3.2.1 A differentiator circuit

As frequency increases the impedance of transformer primary increases as does the gain, limited only by the open loop gain of the high gain block (HGB). This could result in significant noise and interference at the output. The snubber (capacitor in series with a resistor) in parallel with the winding has little effect at low frequency but limits the gain at higher frequency.

The transfer function, in the complex representation $(s = j\omega)$ is [1]: -

$$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{sL_M}{R}$$

Where L_M is the mutual inductance of the transformer.

With a mutual inductance of 1mH (see previous example, section 3.1) and a 10kΩ input resistor the transfer function is: -

$$T(s) = j1.57 \times 10^{-5}$$

If the input voltage is the (in-phase) reference voltage in a resistance bridge this represents the maximum range of the quadrature servo [2]. This is insufficient in many cases and the range can be increased with a differential amplifier between the reference voltage and the differentiator [3]. This is more practicable than increasing the mutual inductance or reducing the input resistance.

1. Part 4, monograph 4: "High accuracy amplifiers, integrators and differentiators". See section 5.
2. Part 5, monograph 1: "Null detectors – the basics".
3. Part 4, monograph 4: "High accuracy amplifiers, integrators and differentiators". See section 6.

Appendix: Magnetic circuit theory (basics)

In the following analysis I shall assume that the core material is linear with a constant value of permeability. I also assume that the effect of inter-winding capacitance is negligible and that the applied voltage is a sine wave. A more in-depth analysis of particular RT and IVD designs, including inter-winding capacitance is undertaken in the relevant monographs.

Fig. 1 defines the main parameters of a basic inductor with winding resistance: -

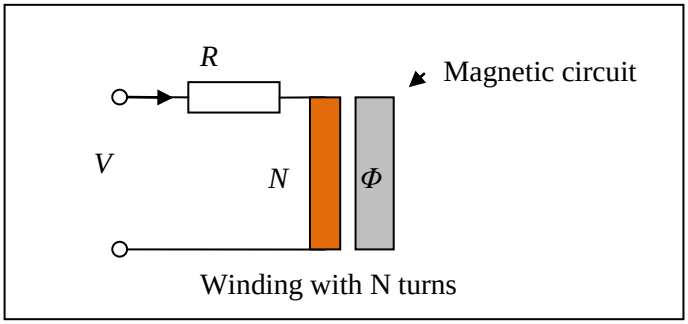


Fig. 1 An inductor with winding resistance

The magneto-motive force (MMF) generated by the current flowing through the winding is defined as:-

$$MMF = NI \qquad \text{(units: "Ampere-turns")}$$

The resulting magnetic flux is determined by the reluctance R_L of the magnetic circuit.
(c.f. Ohm's law: V = IR): -

$$MMF = \Phi R_L$$

Reluctance depends on the cross-sectional area, length and permeability of the magnetic circuit: -

$$R_L = \frac{L_E}{\mu_0 \mu_R A_E}$$

N.B. The formula for reluctance can be compared with that for the resistance of a conductor with length L_E, cross-sectional area, A_E, and conductivity, σ: -

$$R = \frac{L_E}{\sigma A_E}$$

Clearly the permeability is analogous to conductivity.

In practice the parameter provided in most data sheets for soft magnetic components is the reciprocal of reluctance. I call this the "permittance" [1]: -

$$A_L = \frac{1}{R_L} = \frac{\mu_0 \mu_R A_E}{L_E}$$

The flux is, therefore: $\Phi = NIA_L$

1. According to Wikipedia this term was originally used by Oliver Heaviside for the imaginary part of admittance, the reciprocal of impedance. That parameter is now widely referred to as the "susceptance".

The changing flux induces a "back EMF", according to Faraday's Law: -

$$V_B = N\frac{d\Phi}{dt} = N^2 A_L \frac{dI}{dt}$$

If one assumes that the applied voltage is sinusoidal then using the complex representation: $\frac{dI}{dt} = j\omega I = sI$

$$V_B = N^2 A_L sI = sLI$$

Where L is the inductance: -

$$L = N^2 A_L$$

The applied voltage is balanced by the back-EMF plus the resistive voltage drop: -

$$V = IR + V_B = I(R + sL)$$

The total impedance is, therefore, the sum of winding resistance and inductive impedance: -

$$Z = \frac{V}{I} = R + sL$$

The reader will not be surprised that a key parameter is the relative value of the resistance and inductive impedance. A useful measure of this is the natural frequency: -

$$\omega_N = 2\pi f_N = \frac{R}{L}$$

In virtually all cases a lower natural frequency is better – lots of turns of thick copper wire. The art of inductor/transformer design, in this context, is to find the best practical compromise – minimum size and number of turns that is guaranteed to achieve the accuracy required.

High levels of flux density and core saturation

The effective cross-sectional area of the magnetic circuit, A_E, and the number of turns, N, determine the maximum voltage that can be applied at a given frequency. The limitation is set by the maximum flux density, B_{MAX}, of the magnetic material. Above this the permeability of the material drops quite sharply, resulting in "saturation". The inductance drops accordingly and the current increases. If B_{MAX} is the maximum flux density then the maximum amount of flux is $A_E B_{MAX}$.

The maximum (sine wave) voltage, at frequency, ω, is, therefore: -

$$V_{MAX} = N\frac{d\Phi_{MAX}}{dt} = \omega N A_E B_{MAX}\ .$$

In practical RT and IVD design the B_{MAX} is often set quite low (though not in all cases), compared to power transformer applications. As a rule of thumb a B_{MAX} of 500mT is the most one can tolerate for mumetal and similar alloys.

Two-stage filters

1. Introduction

High accuracy in this context refers to low phase error, at a frequency sufficiently below (low pass) or above (high pass) the cut-off frequency of the filter.
The problem with conventional (single-stage) filters is that the phase error, in the pass-band, is first order with respect to frequency. For a low-pass filter, for example, in the complex representation $(s = j\omega)$: -

$$T(s) = \frac{1}{1+\tau s} \quad \text{and} \quad |\tau s| << 1 \quad \Rightarrow \quad T(s) \approx 1 - \tau s + \tau^2 s^2$$

Whereas the in-phase error (the real component) is proportional to frequency squared (very small) the phase error (imaginary and often referred to as "quadrature") is directly proportional to frequency often a problem.

Low phase error is achieved by using the "bootstrap" principle: one filter sits upon and is bootstrapped by another. Two stage low-pass filters, for example, are based on the following structure: -

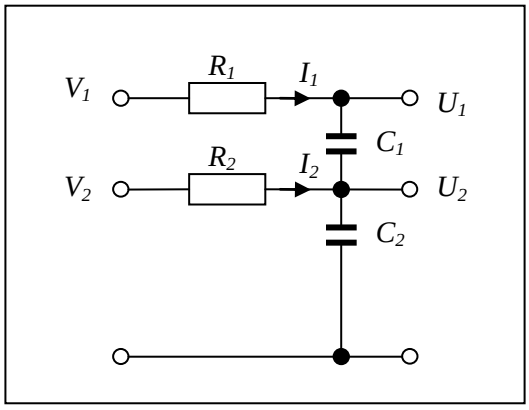


Fig. 1.1 A generalised two-stage low-pass filter network

The resistors and capacitors can be interchanged to provide a high pass filter: -

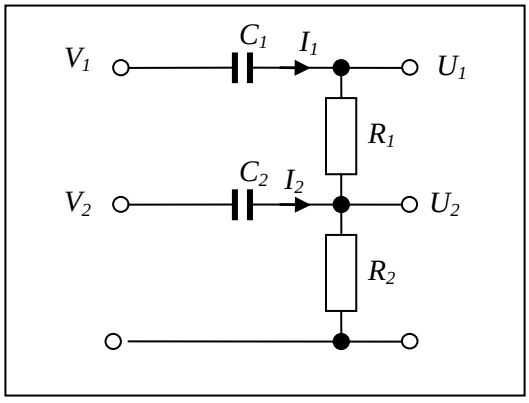


Fig. 1.2 A generalised two-stage high-pass filter network

The most basic filters are when both resistors and capacitors are the same and both inputs are connected to the same low impedance source. The second input (V_2) can, however, be driven from the main output (U_1) via a voltage follower to provide higher input impedance and low output impedance, both useful in some circumstances.

2. Two-stage low-pass filters

2.1 Network analysis

According to Ohm's and Kirchhoff's Laws the output voltages are (see fig. 1.1): -

$$U_2 = \frac{1}{sC_2}(I_1 + I_2) \qquad U_1 = \frac{1}{sC_2}(I_1 + I_2) + \frac{I_1}{sC_1}$$

From this one can construct the matrix equivalent of Ohm's Law. To stop the algebra becoming too messy one can also make a substitution for the impedance of the capacitors $(X = 1/sC)$, restricting the equations to single or, at most, double-deckers: -

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 + X_2 & X_2 \\ X_2 & X_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \text{ i.e. } \mathbf{U} = \mathbf{XI}$$

This defines the output impedance matrix $\mathbf{X}$. It also proves useful when one replaces the capacitors with inductors in the analysis of two-stage inductors (see part 3: "Two-stage IVDs and RTs" [1]).

The input impedance matrix equation can also be constructed by noting: -

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} + \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} R_1 + X_1 + X_2 & X_2 \\ X_2 & R_2 + X_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \qquad \text{i.e. } \mathbf{V} = \mathbf{ZI}$$

This defines the input impedance matrix $\mathbf{Z}$. The inverse of this matrix can be calculated resulting in the matrix equation for the currents. This is also useful for calculating input impedances: -

$$\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$$

Combine with the output impedance matrix to obtain the matrix equivalent of the transfer function equation (output versus input): -

$$\mathbf{U} = \mathbf{XI} = \mathbf{XZ}^{-1}\mathbf{V} = \mathbf{TV}$$

Where $\mathbf{T}$ is the transfer function matrix. Define the matrix $\mathbf{Y}$ so that: $\mathbf{Y} = |\mathbf{Z}|\mathbf{Z}^{-1}$

The inverse of a 2×2 matrix is easy to remember (from "A" level days?): -

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

The result is: $$\mathbf{Y} = \begin{pmatrix} R_2 + X_2 & -X_2 \\ -X_2 & R_1 + X_1 + X_2 \end{pmatrix}$$

Define the matrix $\mathbf{\Omega}$ so that: $\mathbf{\Omega} = |\mathbf{Z}|\mathbf{T} = \mathbf{XY}$ so that $\mathbf{U} = \frac{1}{|\mathbf{Z}|}\mathbf{\Omega V}$

To find the output voltages, in terms of the inputs, it is now a simple case of calculating only the required components of $\mathbf{\Omega}$ and the determinant $|\mathbf{Z}|$.

1. Part 3, monograph 3: "Two-stage inductive voltage dividers and ratio transformers"

2.2 A basic two-stage low-pass filter

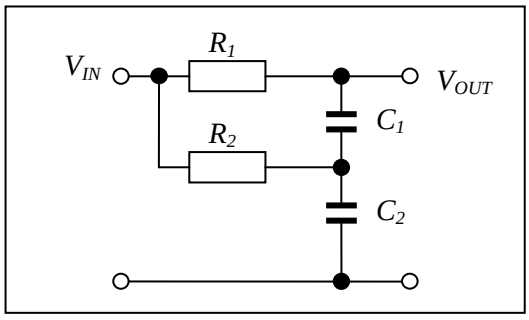


Fig. 2.2.1 A basic two-stage low-pass filter

If, for example, both inputs are connected to the same low resistance source $(V_1 = V_2 = V_{IN})$ the output is: -

$$V_{OUT} = U_1 = \frac{\Omega_{11}V_1 + \Omega_{12}V_2}{|\mathbf{Z}|} = \frac{\Omega_{11} + \Omega_{12}}{|\mathbf{Z}|}V_{IN}$$

From above it is not difficult (unlike larger matrices) to calculate all four components: -

$$\mathbf{\Omega} = \begin{pmatrix} X_1 + X_2 & X_2 \\ X_2 & X_2 \end{pmatrix}\begin{pmatrix} R_2 + X_2 & -X_2 \\ -X_2 & R_1 + X_1 + X_2 \end{pmatrix}$$

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} X_1R_2 + X_2R_2 + X_1X_2 & X_2R_1 \\ X_2R_2 & X_2R_1 + X_1X_2 \end{pmatrix}$$

The determinant is: $$|\mathbf{Z}| = \begin{vmatrix} R_1 + X_1 + X_2 & X_2 \\ X_2 & R_2 + X_2 \end{vmatrix} = R_1R_2 + R_1X_2 + X_1R_2 + X_1X_2 + X_2R_2 + X_2X_2 - X_2X_2$$

Note that the last two terms cancel. The transfer function is, therefore: -

$$T(s) = \frac{(\Omega_{11} + \Omega_{12})}{|\mathbf{Z}|} = \frac{X_1R_2 + X_2R_2 + X_1X_2 + X_2R_1}{R_1R_2 + R_1X_2 + X_1R_2 + X_1X_2 + X_2R_2} \quad \text{with} \quad X_1 = {}^{1}\!/\!_{sC_1} \text{ etc.}$$

Divide top and bottom by X_1X_2 and it is easy to see: -

$$T(s) = \frac{1 + (R_2C_2 + R_2C_1 + R_1C_1)s}{1 + (R_2C_2 + R_2C_1 + R_1C_1)s + R_1R_2C_1C_2s^2}$$

In normalised form $(s = j\,\omega/\omega_N)$: $$T(s) = \frac{1 + as}{1 + as + s^2}$$

With: $$\omega_N = \frac{1}{\sqrt{R_1R_2C_1C_2}} \qquad a = (R_2C_2 + R_2C_1 + R_1C_1)\omega_N$$

If both resistors and capacitors are the same: $a = 3$ and $$T(s) = \frac{3s + 1}{s^2 + 3s + 1}$$

This is the most basic two-stage low pass filter which is useful in many applications.

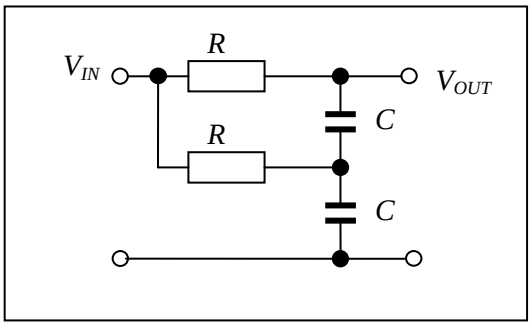


Fig. 2.2.2 A basic two-stage bootstrapped filter

The result is similar to a conventional low pass filter (first order) but with a slight peaking near to the natural frequency and a very much reduced phase shift at low frequency. See section 2.3 for examples.

2.2.1 Phase shift at low frequency $(\omega << \omega_N)$

From above in normalised form $(s = j\omega/\omega_N)$: $$T(s) = \frac{1+as}{1+as+s^2} = 1 - \frac{s^2}{1+as+s^2}$$

To a good approximation, therefore, retaining both real and imaginary error components [1]: -

$$T(s) \approx 1 - s^2 + as^3$$

In more convenient form, with frequency in Hz: -

$$f << f_N \Rightarrow \quad T(f) \approx 1 + \left(\frac{f}{f_N}\right)^2 - ja\left(\frac{f}{f_N}\right)^3 \qquad \text{with} \qquad f_N = \frac{1}{2\pi\sqrt{R_1C_1R_2C_2}}$$

Whereas the real component is second order with respect to frequency (the same as a single stage filter) the imaginary component is now third order and, therefore, very small. This is the phase error, in radians.

One may ask if the phase error of the capacitors is significant. Typical capacitors (polypropylene or polyester dielectric) have "tanδ" of 1-5mrad. Similarly, stray capacitance in parallel with large resistors would introduce a phase error. Either would be equivalent to adding a small imaginary component to the parameter *a*. The algebra remains valid for complex numbers and the result would be a very small contribution to the real (in phase) error component. Consider, for example, the transformation: -

$$ja\left(\frac{f}{f_N}\right)^3 \rightarrow ja(1+\delta j)\left(\frac{f}{f_N}\right)^3 = ja\left(\frac{f}{f_N}\right)^3 - \delta a\left(\frac{f}{f_N}\right)^3$$

In practice, therefore, the operating frequency need only be a factor of 20 or 30 below the natural frequency for the phase shift to be negligible (< 0.5mrad).

1. The method of approximation is: $$\frac{1}{1+z} = \frac{1-z}{1-z^2} = \frac{(1-z)(1+z^2)}{1-z^4} \ldots \text{etc}$$

Now $$|z| << 1 \Rightarrow \qquad |z^4| <<< 1 \qquad \Rightarrow \frac{1}{1+z} \approx 1 - z + z^2$$

This is usually a sufficiently close approximation to calculate, accurately, both real (in-phase) and imaginary (quadrature) components of the error.

2.3 Different values of resistors and capacitors

The parameter *a* determines the degree of resonant peaking. A lower value of a results in a higher peak (c.f. the "damping ratio" in conventional second order systems): -

At the natural frequency $(s = j)$: $$T(s) = \frac{1}{aj} + 1$$

A higher value results in higher phase error (see previous section) and less attenuation at high frequency: -

At high frequency $(\omega >> \omega_N)$: $$T(s) \approx \frac{a}{s}$$

The issue of different *R* and *C* values reduces, therefore, to the question of how much resonant peaking can be tolerated. In practice a value between 2 and 4 is found to be a good compromise. The following plots are for values of *a* = 4 (lowest peak), 2, 1, 0.5 and 0.12 (highest peak): -

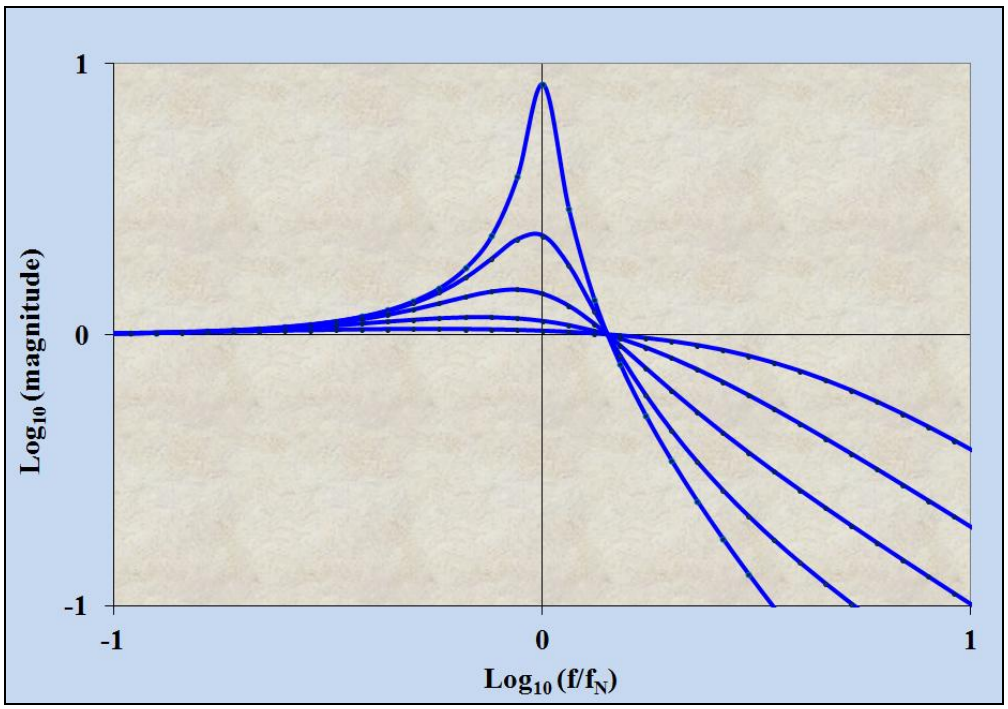


Fig. 2.3.1 Examples of peak response versus high frequency attenuation (for dB ×20).

Given the component values the transfer function, natural frequency and accuracy are easily calculated. In the design process the problem is the other way round. Given the (highest) operating frequency the required accuracy can be obtained by first selecting the natural frequency. One may be a bit conservative but the obvious trade-off is the level of attenuation provided at high frequency. The choice of parameter *a* depends on the application. This author has employed a value of 2 or 3 with no problems The choice of practical capacitor values is also a consideration as the range of values is best kept limited (or else hold stock of unusual values): -

Typical capacitor values (each decade): 1 1.5 2.2 3.3 4.7 6.8 10 etc…

Whereas good quality capacitors are available in only a limited range of standard values (up to a maximum of 2.2 µF or possibly 4.7µF in extremis) the range of standard resistor values is much greater: -

Standard resistor values (E24 rage): -

1	1.1	1.2	1.3	1.5	1.6	1.8	2	2.2	2.4	2.7	3
3.3	3.9	4.3	4.7	5.1	5.6	6.2	6.8	7.5	8.2	and	9.1

The trick is, therefore, to first choose a practical combination of capacitors, usually with a bit of "trial and error", and then calculate the resistances required which give the chosen value for natural frequency and *a*. The resistors in the E24 range are about 5% apart and should get one sufficiently close to the target in the vast majority of cases. For greater precision choose resistors from the E96 range.

Selecting the best combination can also be made easier with a "slide rule component calculator" (see appendix 1). It is also worth noting that from the available capacitors, the following pairs can make selection much simpler: -

$0.1 \approx \frac{1}{10}$ $0.22 \approx \frac{1}{4.7}$ $0.33 \approx \frac{1}{3.3}$ $0.68 \approx \frac{1}{1.5}$ and vice versa. The worst case error is about 10%.

The following method is one way to avoid too much "trial and error": -

Define a baseline capacitor, *C*, and resistor value, *R*, so that: $f_N = \frac{1}{2\pi RC}$

If possible choose cardinal values (e.g. $f_N = 1.6kHz; R = 10k\Omega;\ C = 10nF$) or else use the slide rule calculator.

Define parameters $\alpha\ \beta\ \gamma$ and δ (all >0) such that: -

$$R_1 = \alpha R \qquad R_2 = \beta R \qquad C_1 = \gamma C \qquad C_2 = \delta C \qquad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

Substitution into the equation for a results in a much simpler expression (the baseline values *R*, *C* cancel): -

$$a = (R_2C_2 + R_1C_1 + R_2C_1)\omega_N = \beta\delta + \alpha\gamma + \beta\gamma \qquad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

One can remove any explicit involvement of δ by noting: -

$$\beta\delta = \frac{1}{\alpha\gamma} \quad \Rightarrow \quad a = \frac{1}{\alpha\gamma} + \alpha\gamma + \beta\gamma \qquad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

This proves that one can choose δ as small as one likes (a sufficiently small capacitor for C_2) so that the other parameters are as big as one likes and there is no upper limit to the value of parameter a. The usual problem, however, is finding a combination of practical values for a nearer to the minimum.

2.3.1 Example calculation: The minimum $(a = 2)$: -

The terms in $\alpha\gamma$ occur in such a way that the minimum contribution and, therefore, the minimum value is $a = 2$. This is easily shown by basic calculus with a change of variable $(\alpha\gamma = x)$: differentiate and solve for zero (slope):

$$y = \frac{1}{x} + x \qquad \Rightarrow \frac{\partial y}{\partial x} = -\frac{1}{x^2} + 1 = 0 \qquad \text{(at the minimum)}$$

$$\Rightarrow x = \pm 1 \qquad \Rightarrow \alpha\gamma = 1 \qquad \text{and} \qquad \frac{1}{\alpha\gamma} + \alpha\gamma = 2 \qquad \text{(at the minimum)} \qquad \text{QED.}$$

Clearly the positive result is the only sensible one as all the parameters must be numerically positive. This uses up one more constraint but the remaining constraint allows us to choose a practical combination of α and γ for almost any value of a (above 2): -

$$\text{choose } \alpha = \frac{1}{\gamma} \quad \Rightarrow \quad a = 2 + \beta\gamma \qquad \text{and} \qquad \delta = \frac{1}{\alpha\beta\gamma} = \frac{1}{\beta}$$

For very nearly the minimum choose, therefore: -

$$\alpha = \gamma = 1 \qquad \text{and} \qquad \beta << 1 \qquad \Rightarrow \quad \delta >> 1$$

Example: $\beta = 0.01 \quad \Rightarrow \quad \delta = 100$ (easy decade values) $\quad \Rightarrow \quad a = 2.01$

For a natural frequency of 10kHz choose, therefore: -

$$C = 1nF \qquad R = 16k\Omega \qquad R_1 = 16k\Omega \qquad R_2 = 160\Omega \qquad C_1 = 1nF \qquad C_2 = 100nF$$

2.3.2 Example calculation 2: For a relatively high value $(a = 4)$

One can still retain the simplification which comes from choosing: -

$$\alpha = \frac{1}{\gamma} \quad \Rightarrow \quad \beta\gamma = 2 \quad \text{and} \quad \delta = \frac{1}{\alpha\beta\gamma} = \frac{1}{\beta}$$

Now $\alpha\gamma = 1$ so choose $\gamma = 2.2$ (standard value capacitor) for reasons which will become apparent.

$$\gamma = 2.2 \quad \Rightarrow \quad \alpha = \frac{1}{2.2} = 0.47 \quad \beta = \frac{2}{\gamma} = 0.91 \quad \delta = \frac{1}{0.91} \approx 1 \quad \text{(near enough)}$$

For a natural frequency of 16kHz choose nearest preferred value resistors (E24 range) based on the same baseline values as before: -

$$C = 1nF \qquad \Rightarrow R = 10k\Omega \qquad C_1 = 2.2nF \qquad C_2 = 1nF$$

$$R_1 = 0.47 \times 10k\Omega = 4.7k\Omega \qquad R_2 = 0.91 \times 10k\Omega \approx 9.1k\Omega$$

The natural frequency is slightly out (by 2.5%): -

$$f_N = \frac{1}{2\pi\sqrt{R_1C_1R_2C_2}} = 16.4kHz$$

The resulting value for a is also very close to target. The most convenient way to calculate it is with the actual parameter values chosen (in order to coincide with preferred values): -

$$\gamma = 2.2 \qquad \alpha = 0.47 \qquad \beta = 0.91 \qquad \delta = 1$$

$$\Rightarrow a = \beta\delta + \alpha\gamma + \beta\gamma = 0.91 + 1 + 0.91 \times 2.2 = 3.9$$

2.3.3 Widely different values for resistors and capacitors

It is interesting to note the practical consequence of section 2.3.1. The interaction between the bootstrap stage (R_2 and C_2) and the upper stage is minimised by choosing a much larger value for C_2 and a correspondingly lower value for R_2 to arrive at the required time constant: -

$$a = (R_2C_2 + R_1C_1 + R_2C_1)\omega_N \quad \text{and} \quad C_2 >> C_1 \Rightarrow \quad a \approx (R_2C_2 + R_1C_1)\omega_N = \sqrt{\frac{R_2C_2}{R_1C_1}} + \sqrt{\frac{R_1C_1}{R_2C_2}}$$

The parameter, *a*, is seen to simplify and be a function of the ratio of time constants, in the way (as above) so that the minimum value must be *a* = 2. This becomes a useful rule of thumb when it comes to actively driven filters…

3. Actively driven low-pass filters

3.1 Network analysis

A lower value for the parameter a (< 2) can be achieved by employing an active drive for V_2. A lower value reduces the phase error at the expense of a higher peak in the frequency response near to the natural frequency. This approach also has the advantage of increasing the input impedance and provides very low output impedance. It is also easier to calculate component values. The general circuit is as follows: -

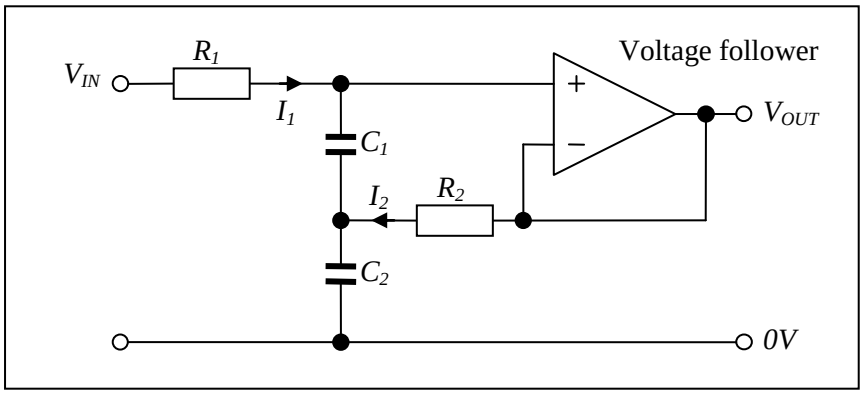


Fig. 3.1 A two-stage actively driven Low-pass filter

The input is: $V_{IN} = V_1$ but the other network input is now connected to the follower output: $V_2 = V_{OUT}$

$$V_{OUT} = U_1 = \frac{\Omega_{11}V_1 + \Omega_{12}V_{OUT}}{|\mathbf{Z}|} \qquad \Rightarrow \qquad T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{\Omega_{11}}{|\mathbf{Z}| - \Omega_{12}}$$

The basic network analysis remains the same, with the matrix elements: -

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} X_1R_2 + X_2R_2 + X_1X_2 & X_2R_1 \\ X_2R_2 & X_2R_1 + X_1X_2 \end{pmatrix}$$

And $$|\mathbf{Z}| = R_1R_2 + R_1X_2 + X_1R_2 + X_1X_2 + X_2R_2$$

$$\Rightarrow T(s) = \frac{X_1R_2 + X_2R_2 + X_1X_2}{R_1R_2 + X_1R_2 + X_1X_2 + X_2R_2} \qquad \text{with} \qquad X_1 = 1/sC_1 \text{ etc.}$$

Divide top and bottom by X_1X_2 and it is easy to see: -

$$T(s) = \frac{1 + (R_2C_2 + R_2C_1)s}{1 + (R_2C_2 + R_2C_1)s + R_1R_2C_1C_2s^2}$$

In normalised form $(s = j\omega/\omega_N)$ the transfer function is of the same form as before: $T(s) = \dfrac{1 + as}{1 + as + s^2}$

With the same natural frequency: $\omega_N = \dfrac{1}{\sqrt{R_1R_2C_1C_2}}$ but this time: $a = R_2(C_2 + C_1)\omega_N$

3.2 Equal resistor and capacitor values

If the resistors and capacitors are the same the result is a lower value $(a = 2)$ which proves to be satisfactory in many applications.

3.3 Different values of resistors and capacitors

A two-stage filter can also be used as a feedback element for which the natural frequency and the value of a need to be optimised for reasons of stability. See [1] for example. The method for calculating component values is similar but much simplified. Define parameters α β γ and δ (all > 0) such that: -

$$R_1 = \alpha R \qquad R_2 = \beta R \qquad C_1 = \gamma C \qquad C_2 = \delta C \qquad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

$$\Rightarrow \qquad a = \beta(\gamma + \delta) \quad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

The parameter α is not included in the first expression so that one can choose any value of R_1 and suitable values for R_2, C_1 and C_2 for any value of natural frequency. There is no longer a minimum value for parameter a. One may as well choose cardinal values for the capacitors and then the resistors are in inverse proportion (and there are lots to choose from).

$$\gamma = \delta = 1 \qquad \Rightarrow \qquad a = 2\beta \qquad \text{and} \qquad \alpha = \frac{1}{\beta}$$

I.e. given a value of a it is easy to calculate β and then α.

3.3.1 Example calculation: $a = 5$ and $\gamma = \delta = 1 \quad \Rightarrow \quad \beta = 2.5 \quad$ and $\quad \alpha = 0.4$

For a natural frequency of 1kHz choose, therefore: -

$$C = 10nF \;\Rightarrow\; R = 16k\Omega \qquad \Rightarrow \qquad R_1 = 6.4k\Omega \qquad R_2 = 40k\Omega \qquad \text{and} \qquad C_1 = C_2 = 10nF$$

The nearest preferred E24 values are 6k2 and 39k for which: -

$$f_N = \frac{1}{2\pi\sqrt{R_1 R_2 C_1 C_2}} = 995Hz \qquad \text{and} \qquad a = R_2(C_2 + C_1)\omega_N = 5.0$$

3.4 Widely different values of resistors and capacitors

N.B. One can understand the reason why there is no lower limit to the value of parameter *a* in terms of positive feedback around the follower. If the input is connected to 0V it will be noted that the feedback network consists of a low-pass filter followed by a high-pass filter, with some interaction between the two. The combination is a band-pass filter. If the cut-off frequency of the high-pass filter is lower than the cut-off frequency of the low-pass filter the peak loop gain approaches unity (i.e. 100% positive feedback) resulting in low stability margin.

As a rule of thumb, therefore: $\qquad R_1 C_1 \le R_2 C_2$

A practical approach, if one requires a low value for parameter *a*, is to simplify the calculation by choosing a much higher value for C_2 and a correspondingly lower value for R_2. The interaction between the first and second stage is much reduced: -

$$\delta >> \gamma \Rightarrow \qquad a \approx \beta\delta \qquad \text{and} \qquad \beta\delta = \frac{1}{\alpha\gamma} \qquad \Rightarrow \qquad a \approx \sqrt{\frac{R_2 C_2}{R_1 C_1}} \qquad \text{and} \qquad R_1 C_1 \approx R_2 C_2$$

If the resistors and capacitors have the same preferred values other than, say, a factor of 10 the result is a low value for parameter *a*: -

$$C_2 = 10C_1 \qquad \text{and} \qquad R_1 = 10R_2 \qquad \Rightarrow \qquad a \approx 1$$

Such pairs are easily obtained for any natural frequency with the slide rule calculator (see appendix 1).

1. Part 6, monograph 2: "A simulated negative capacitor circuit"

4. Two-stage high-pass filters

4.1 Network analysis

Swap resistors and capacitors and employ the same matrix method as above: -

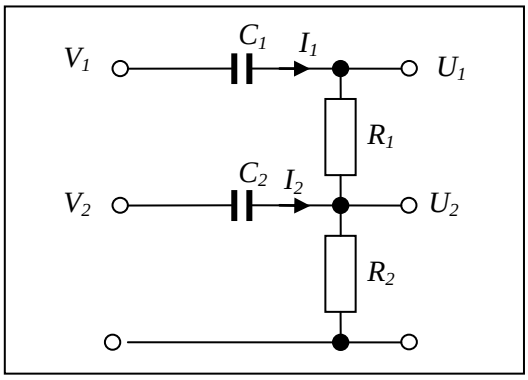


Fig. 3.2 A generalised two-stage high-pass filter network

According to Ohm's and Kirchhoff's Laws the output voltages are: -

$$U_2 = (I_1 + I_2)R_2 \qquad U_1 = (I_1 + I_2)R_2 + I_1R_1$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} R_1 + R_2 & R_2 \\ R_2 & R_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \text{ i.e. } \mathbf{U} = \mathbf{XI}$$

This defines the output impedance matrix $\mathbf{X}$. The input impedance matrix equation can also be constructed. The algebra is simplified with $(X_1 = 1/sC_1)$ etc: -

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} + \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} X_1 + R_1 + R_2 & R_2 \\ R_2 & X_2 + R_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \qquad \text{i.e. } \mathbf{V} = \mathbf{ZI}$$

This defines the input impedance matrix $\mathbf{Z}$. The inverse of this matrix can be calculated resulting in the matrix equation for the currents. This is also useful for calculating input impedances: -

$$\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$$

Combine with the output impedance matrix equation to obtain the matrix equivalent of the transfer function equation (output versus input): -

$$\mathbf{U} = \mathbf{XI} = \mathbf{XZ}^{-1}\mathbf{V} = \mathbf{TV}$$

Where $\mathbf{T}$ is the transfer function matrix. Define the matrix $\mathbf{Y}$ so that: $\mathbf{Y} = |\mathbf{Z}|\mathbf{Z}^{-1}$

The inverse of a 2×2 matrix is easy to remember: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

The result is: $\mathbf{Y} = \begin{pmatrix} R_2 + X_2 & -R_2 \\ -R_2 & R_1 + R_2 + X_1 \end{pmatrix}$

$$|\mathbf{Z}| = R_2R_1 + R_2R_2 + R_2X_1 + X_2R_1 + X_2R_2 + X_2X_1 - R_2R_2$$

Define the matrix $\mathbf{\Omega}$ so that: $\mathbf{\Omega} = |\mathbf{Z}|\mathbf{T} = \mathbf{XY}$ so that $\mathbf{U} = \frac{1}{|\mathbf{Z}|}\mathbf{\Omega V}$

then $\mathbf{\Omega} = \begin{pmatrix} R_1 + R_2 & R_2 \\ R_2 & R_2 \end{pmatrix}\begin{pmatrix} R_2 + X_2 & -R_2 \\ -R_2 & R_1 + R_2 + X_1 \end{pmatrix}$

$$\mathbf{\Omega} = \begin{pmatrix} R_1R_2 + R_1X_2 + R_2X_2 & R_2X_1 \\ R_2X_2 & R_2R_1 + R_2X_1 \end{pmatrix}$$

To find any of the output voltages, in terms of the inputs, it is now a simple case of calculating only the required components of $\mathbf{\Omega}$ and the determinant $|\mathbf{Z}|$.

Note that the result is identical to that for the two-stage low pass filter network with the substitutions: -

$$X_1 \leftrightarrow R_1 \quad \text{etc.}$$

One can employ this shortcut, at least as a check, when it comes to the more complicated 3 × 3 matrices [1].

4.2 A basic two-stage high-pass filter

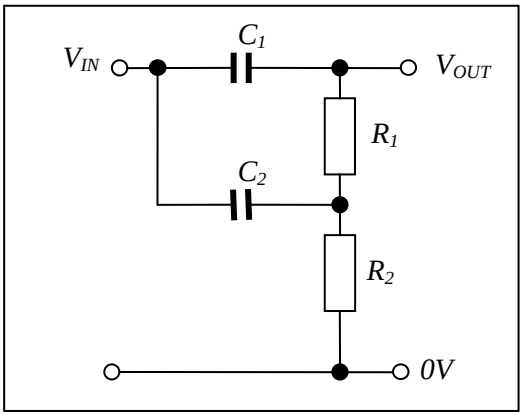


Fig. 4.2.1 A basic two-stage high-pass filter

If, for example, both inputs are connected to the same low resistance source $(V_1 = V_2 = V_{IN})$ the output is: -

$$V_{OUT} = U_1 = \frac{(\Omega_{11}V_1 + \Omega_{12}V_2)}{|\mathbf{Z}|} = \frac{(\Omega_{11} + \Omega_{12})}{|\mathbf{Z}|}V_{IN}$$

The transfer function is, therefore: -

$$T(s) = \frac{R_1R_2 + R_1X_2 + R_2X_2 + R_2X_1}{R_1R_2 + R_1X_2 + R_2X_2 + R_2X_1 + X_1X_2} \quad \text{with} \quad (X_1 = 1/sC_1) \quad \text{etc.}$$

Divide top and bottom by X_1X_2: $T(s) = \frac{s^2C_1C_2R_1R_2 + sR_1C_1 + sR_2C_1 + sR_2C_2}{s^2C_1C_2R_1R_2 + sR_1C_1 + sR_2C_1 + sR_2C_2 + 1}$

$$\Rightarrow \quad T(s) = \frac{as + s^2}{1 + as + s^2} \quad \text{with} \quad a = (R_1C_1 + R_2C_1 + R_2C_2)\omega_N \quad \text{and} \quad \omega_N = \frac{1}{\sqrt{R_1C_1R_2C_2}}$$

1. Part 2, monograph 2: “Three-stage filters”

4.2.1 Phase shift at high frequency $(\omega >> \omega_N)$

One can employ the same method as before (see footnote in section 2.2.1). From above in normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{as + s^2}{1 + as + s^2} = 1 - \frac{1}{1 + as + s^2}$$

To a good approximation, therefore: $T(s) = 1 - \dfrac{1}{s^2\left(1 + \dfrac{a}{s} + \dfrac{1}{s^2}\right)} \approx 1 - \dfrac{1}{s^2} + \dfrac{a}{s^3}$

In more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \qquad \text{with} \qquad f_N = \frac{1}{2\pi\sqrt{R_1C_1R_2C_2}}$$

Whereas the real component is second order with respect to frequency (the same as a single stage filter) the imaginary component is now third order and, therefore, very small. This is the phase error, in radians.

4.3 Equal resistor and capacitor values

If the resistors and capacitors are the same the result is a basic two-stage high pass filter with $a = 3$

4.4 Different values of resistors and capacitors

The expressions for natural frequency and the parameter *a* are exactly the same as for the basic two-stage low pass filter and the same method may be employed for calculating component values. See section 2.3 and its subsections. The trade-off between resonant peaking and attenuation are similar: -

At the natural frequency $(s = j)$: $T(s) = 1 + \dfrac{j}{a}$ At high frequency $(\omega >> \omega_N)$: $T(s) \approx \dfrac{a}{s}$

The following plots are for values of *a* = 4 (lowest peak), 2, 1, 0.5 and 0.12 (highest peak): -

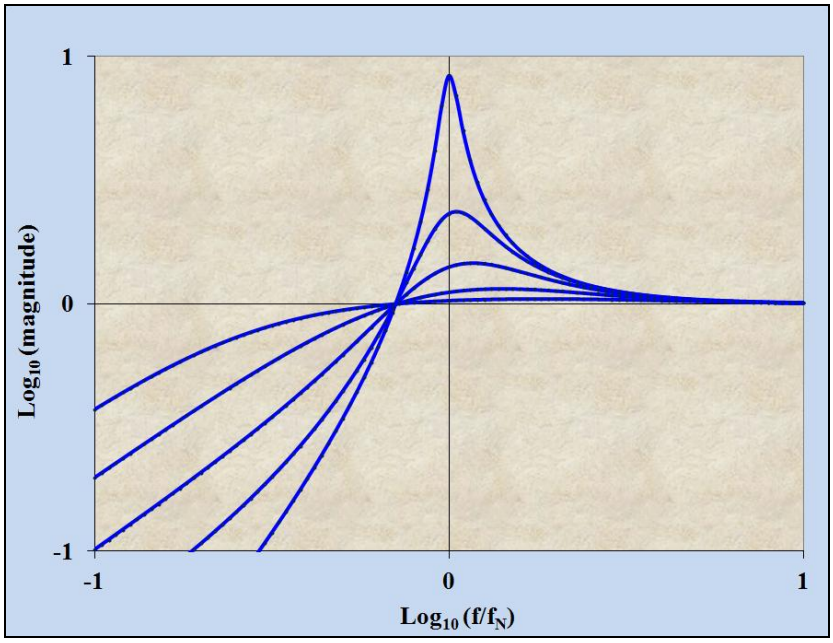


Fig. 4.3.1 Examples of peaking versus low frequency attenuation (for dB ×20)

5. Actively driven two-stage high-pass filters

5.1 Circuit analysis

Active drive affords the same advantages as with the actively driven low pass filter. Specifically: -
a). An unlimited range of values for the parameter a (especially a value lower than 2).
b). High input impedance and low output impedance.
c). Easier to calculate component values.

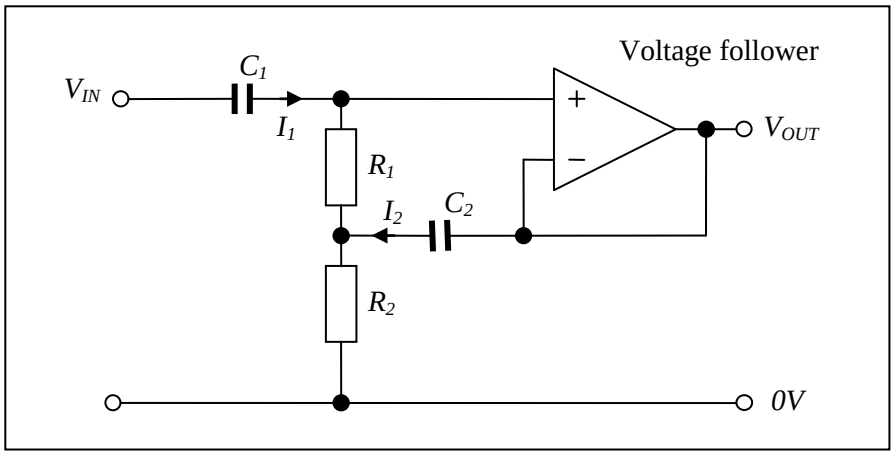


Fig. 5.1.1 A two-stage actively driven high pass filter

The input is: $V_{IN} = V_1$ but the other network input is now connected to the follower output: $V_2 = V_{OUT}$

$$V_{OUT} = U_1 = \frac{\Omega_{11}V_1 + \Omega_{12}V_{OUT}}{|\mathbf{Z}|} \quad \Rightarrow \quad T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{\Omega_{11}}{|\mathbf{Z}| - \Omega_{12}}$$

The basic network analysis remains the same, with the matrix elements (see section 4.1): -

$$\mathbf{\Omega} = \begin{pmatrix} R_1R_2 + R_1X_2 + R_2X_2 & R_2X_1 \\ R_2X_2 & R_2R_1 + R_2X_1 \end{pmatrix} \quad \text{with} \quad X_1 = \frac{1}{sC_1} \text{ etc: -}$$

The determinant is: $$|\mathbf{Z}| = R_2R_1 + R_2X_1 + X_2R_1 + X_2R_2 + X_2X_1$$

$$T(s) = \frac{R_1R_2 + R_1X_2 + R_2X_2}{R_2R_1 + R_2X_1 + X_2R_1 + X_2R_2 + X_2X_1 - R_2X_1}$$

One of the terms in the denominator cancels (R_2X_1). Divide top and bottom by X_2X_1 with $(X_1 = 1/sC_1)$ etc.

$$T(s) = \frac{s^2C_1C_2R_1R_2 + sR_1C_1 + sR_2C_1}{s^2C_1C_2R_1R_2 + sR_1C_1 + sR_2C_1 + 1}$$

In normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{as + s^2}{1 + as + s^2} \quad \text{with} \quad a = (R_1 + R_2)C_1 \quad \text{and} \quad \omega_N = \frac{1}{\sqrt{R_1C_1R_2C_2}}$$

5.1.1 Phase error at high frequency

One can employ the same method of approximation. The (normalised) transfer function for high frequency $(|s| >> 1)$ is, to a very good approximation: -

$$T(s) \approx 1 - \frac{1}{s^2} + \frac{a}{s^3}$$

In the more convenient form with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \qquad \text{with} \qquad f_N = \frac{1}{2\pi\sqrt{R_1 C_1 R_2 C_2}}$$

5.2 Equal resistor and capacitor values

If the resistors and capacitors are the same the result is a lower value $(a = 2)$ which proves to be satisfactory in many applications.

5.3 Different values of resistors and capacitors

The method for calculating component values is similar to the actively driven low pass filter and much simplified: -

Define parameters α β γ and δ (all > 0) such that: -

$$R_1 = \alpha R \qquad R_2 = \beta R \qquad C_1 = \gamma C \qquad C_2 = \delta C \qquad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

$$\Rightarrow \qquad a = \gamma(\alpha + \beta) \quad \text{and} \qquad \alpha\beta\gamma\delta = 1$$

The parameter δ is not included in the first expression so that one can choose any value of C_2 and suitable values for R_1, R_2 and C_1 for any value of natural frequency. There is no longer a minimum value for parameter a. One way of simplifying the calculation is to choose widely different and reciprocal values (i.e. pairs of 10/0.1, 2.2/0.47 or 3.3/0.33) for the capacitors and then calculate resistor values according to: -

Example: $\gamma = 0.47 = \frac{1}{\delta} \quad \Rightarrow \quad \delta = 2.2 \quad \Rightarrow \quad \alpha \approx \frac{1}{\beta}$ and $a = 0.47\left(\frac{1}{\beta} + \beta\right)$

This provides for a minimum value $(a = 0.47)$ with $(\alpha = \beta = 1)$ although in most cases $\left(\beta > 1 > \frac{1}{\beta}\right)$ and, to a good approximation: $\beta \approx 2.2a$. I.e. given a value of a it is easy to calculate β and then α.

5.3.1 Example calculation: target $a = 1 \quad \Rightarrow \quad \delta = \beta = 2.2$ and $\gamma = \alpha = 0.47$

For a natural frequency of 1Hz choose, therefore, base values: $C = 1\mu F$ and $R = 159k\Omega$

The nearest preferred values are: $C_1 = 470nF \quad C_2 = 2.2\mu F \quad R_1 = 75k\Omega$ and $R_2 = 330k\Omega$

Result: $f_N = \frac{1}{2\pi\sqrt{R_1 R_2 C_1 C_2}} = 0.995Hz$ and $a = R_2(C_2 + C_1)\omega_N = 0.88$

The value for a is a bit low and so increase R_2 to 390k. The natural frequency is reduced a little but that is less critical: -

$$f_N = 0.915Hz \qquad a = 1.04$$

5.4 Widely different values of resistors and capacitors

If the input is connected to 0V it will be noted that the feedback network consists of a high-pass filter followed by a low-pass filter, with some interaction between the two. The combination is a band-pass filter. If the cut-off frequency of the high-pass filter is lower than the cut-off frequency of the low-pass filter the peak gain approaches unity (i.e. 100% positive feedback) and very low stability margin (c.f. the actively driven two-stage low-pass filter, section 3.4). **As a rule of thumb, therefore:** -

$$R_1C_1 \le R_2C_2$$

An alternative approach is to simplify the calculation by reducing the interaction by choosing a much higher value for C_2 and a lower value for R_2: -

$$a = \gamma(\alpha + \beta) \quad \text{and} \quad \alpha\beta\gamma\delta = 1$$

$$\alpha >> \beta \Rightarrow \quad a \approx \gamma\alpha \quad \text{and} \quad \gamma\alpha = \frac{1}{\beta\delta} \quad \Rightarrow \quad a \approx \sqrt{\frac{R_1C_1}{R_2C_2}} \quad \text{and} \quad R_1C_1 \approx R_2C_2$$

If the resistors and capacitors have the same preferred values other than, say, a factor of 10 the result is a low value for parameter *a*: -

$$C_2 = 10C_1 \quad \text{and} \quad R_1 = 10R_2 \quad \Rightarrow \quad a \approx 1$$

Such pairs are easily obtained for any natural frequency with the slide rule calculator (see appendix 1).

5.5 Input impedance

From the network analysis we obtained the equations: $\mathbf{I} = \mathbf{Z^{-1}V}$ with $\mathbf{Y} = |\mathbf{Z}|\mathbf{Z^{-1}}$ and: -

$$\mathbf{Y} = \begin{pmatrix} R_2 + X_2 & -R_2 \\ -R_2 & R_1 + R_2 + X_1 \end{pmatrix} \quad \text{with} \quad (X_1 = 1/sC_1) \text{ etc: -}$$

$$\text{And} \quad |\mathbf{Z}| = R_2R_1 + R_2X_1 + X_2R_1 + X_2R_2 + X_2X_1$$

The network is connected so that: $V_1 = V_{IN}$ and $V_2 = V_{OUT} = V_{IN}T(s)$

$$\Rightarrow I_1 = \frac{Y_{11}V_1 + Y_{12}V_2}{|\mathbf{Z}|} = \frac{Y_{11} + Y_{12}T(s)}{|\mathbf{Z}|}V_{IN}$$

The input admittance (reciprocal of impedance) is, therefore: $A_{IN} = \frac{I_1}{V_{IN}} = \frac{R_2 + X_2 - R_2T(s)}{|\mathbf{Z}|}$

$$\Rightarrow \quad A_{IN} = \frac{X_2 + R_2(1 - T(s))}{|\mathbf{Z}|}$$

Divide top and bottom by X_1X_2 and note from above: $\frac{|\mathbf{Z}|}{X_1X_2} = s^2C_1C_2R_1R_2 + sR_1C_1 + sR_2C_1 + 1$

One can express this factor in the normalised form $(s = j\omega/\omega_N)$: -

$$\frac{|\mathbf{Z}|}{X_1X_2} = 1 + as + s^2 \quad \text{with} \quad \omega_N = \frac{1}{\sqrt{R_1C_1R_2C_2}}$$

To reveal: $A_{IN} = \left(\frac{1}{1 + as + s^2}\right)\left(\frac{1}{X_1} + \frac{R_2}{X_1X_2}(1 - T(s))\right)$

Also in normalised form, from above: $T(s)=\dfrac{as+s^2}{1+as+s^2}$ so that $1-T(s)=\dfrac{1}{1+as+s^2}$

This simplifies. In normalised form: $\dfrac{R_2}{X_1X_2}=s^2R_2C_1C_2\rightarrow\dfrac{s^2}{R_1}$

I shall retain the capacitive admittance term $1/X_1$ for reasons that will become clear.

Note that, at high frequency, to a good approximation: $(\omega>>\omega_N)\Rightarrow\ \dfrac{1}{1+as+s^2}\approx\dfrac{1}{s^2}\left(1-\dfrac{a}{s}\right)$

$$\Rightarrow\qquad A_{IN}=\frac{1}{s^2}\left(1-\frac{a}{s}\right)\left(\frac{1}{X_1}+\frac{1}{R_1}\left(1-\frac{a}{s}\right)\right)$$

Now $1/X_1$ is imaginary and $1/R_1$ is real and, therefore, the small corrections are insignificant. To a very good approximation, therefore, the significant real and imaginary components are contained within: -

$$A_{IN}=\frac{1}{Z_{IN}}=\frac{1}{s^2}\left(\frac{1}{X_1}+\frac{1}{R_1}\right)$$

This is how impedances add in parallel. The input appears, therefore, as a small negative capacitance (i.e. inductive) in parallel with a large negative resistance. Both impedances are frequency dependent. In more convenient form, with frequency in Hz: -

$$f_N=\frac{1}{2\pi\sqrt{R_1C_1R_2C_2}}\Rightarrow\qquad A_{IN}=-j\left(\frac{f_N}{f}\right)2\pi f_NC_1-\left(\frac{f_N}{f}\right)^2\frac{1}{R_1}$$

The imaginary component of impedance increases in proportion to frequency, relative to the impedance of the capacitor at the natural frequency. The impedance of the real component increases as the frequency squared. Alternatively one can express the components of the parallel impedances due to the capacitor and resistor as follows: -

$$A_{IN}=\frac{1}{Z_{IN}}=\frac{1}{Z_C}+\frac{1}{Z_R}\qquad\text{with}\qquad Z_C=j\left(\frac{f}{f_N}\right)\frac{1}{2\pi f_NC_1}\quad\text{and}\qquad Z_R=-\left(\frac{f}{f_N}\right)^2R_1$$

5.5.1 Example calculation

For the previous example: $C_1=470nF$ and $R_1=75k\Omega$ with $f_N=0.915Hz$

At 0.915Hz the impedance of the capacitor is: $X_1(f_N)=\dfrac{1}{sC_1}=\dfrac{1}{j2\pi f_NC_1}=-j370k\Omega$

At 75Hz this becomes: $Z_C=\dfrac{75}{0.915}\times X_1(f_N)\approx j30M\Omega$

The real component (in parallel) at 75Hz is: $Z_R=-\left(\dfrac{f}{f_N}\right)^2R_1\approx-500M\Omega$

At high frequency the real component is significantly larger than the imaginary component and the effect of the capacitor dominates: the input "looks like" a very large inductor.

6. An LCR low-pass filter with two-stage response

6.1 Introduction

A useful variant, for high cut-off frequency, is a filter based on an inductor, capacitor and resistor. The dynamic response is the same as a two-stage low-pass filter and its output impedance, at low frequency, is sufficiently low (and mainly inductive) to connect directly to the feedback capacitor in a negative capacitor circuit [1].

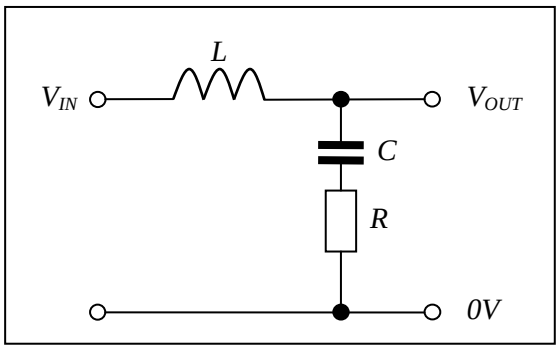


Fig. 6.1.1 An LCR filter with two-stage response

6.2 Circuit analysis

According to Kirchhoff's and Ohm's laws, in the complex representation: -

$$T(s)=\frac{V_{OUT}}{V_{IN}}=\frac{R+\frac{1}{sC}}{R+\frac{1}{sC}+sL}=\frac{1+sRC}{1+sRC+s^2CL}$$

In normalised form $(s=j\omega/\omega_N)$: $T(s)=\frac{1+as}{1+as+s^2}$ with $a=RC\omega_N$ and $\omega_N=\frac{1}{\sqrt{LC}}$

The stability (see section 2.3) and error analysis are the same. In convenient form, with frequency in Hz: -

$$f<<f_N \Rightarrow \quad T(f)\approx 1+\left(\frac{f}{f_N}\right)^2-ja\left(\frac{f}{f_N}\right)^3 \qquad \text{with} \qquad f_N=\frac{1}{2\pi\sqrt{LC}}$$

6.3 Output impedance

The main application is in a circuit for simulating a negative capacitor. It is important that the output is sufficiently low and, preferably, imaginary. When in series with a capacitor the effect is a small reduction in the capacitance but not a phase error. The output impedance is the RC series combination in parallel with the inductance.

$$Z_{OUT}(s)=\frac{\left(R+\frac{1}{sC}\right)sL}{R+\frac{1}{sC}+sL}=\frac{(1+sRC)sL}{1+sRC+s^2CL}$$

At low frequency, the output impedance is, therefore, substantially the inductance: -

$$\left|s^2CL\right|<<1 \quad \Rightarrow \quad Z_{OUT}(s)\approx sL$$

1. Part 6, monograph 2: "A simulated negative capacitor circuit"

Appendix 1: Slide rule calculator for RC combinations

Procedure: -

1. Calculate the required time constant $\tau = RC$ and write it in the form $\alpha \times 10^N$ where α is between 1.0 and 9.9.
2. Record the required value for N.
3. Calculate $X_0 = log_{10}\,\alpha$ (X_0 should be between 0.0 and 1.0).
4. Place the Y_1 marker at X_0 on the X scale and look for the two values of R and C that most closely match.

Note: Apart from factors of 10 $X_0 = \log_{10} RC = \log_{10} R + \log_{10} C = \log_{10} R - \log_{10} \frac{1}{C}$

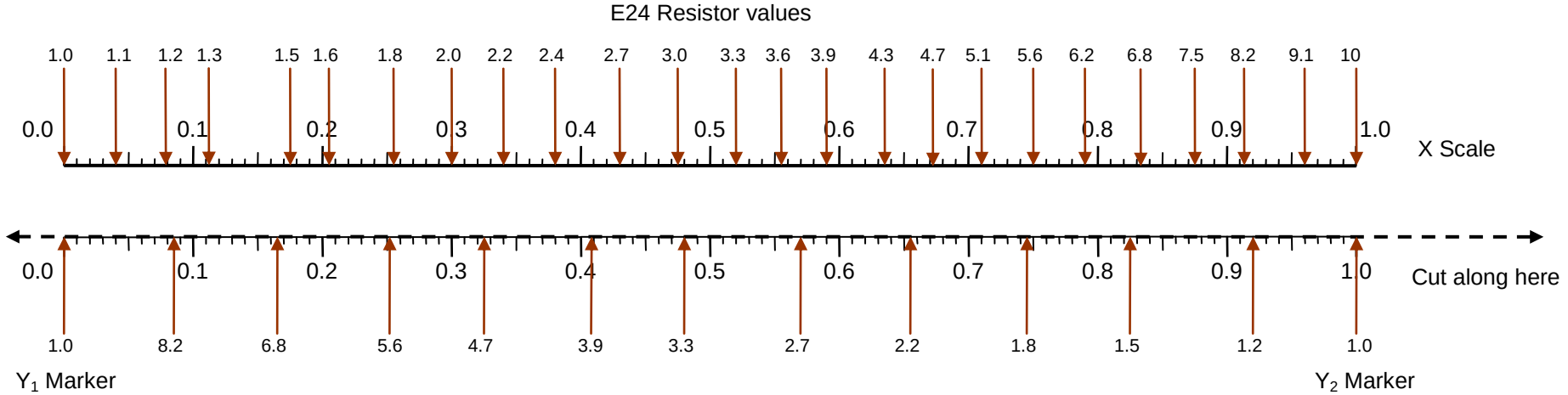
The slide rule performs the addition $X_0 + \log_{10} \frac{1}{C} = \log_{10} R$

5. Try the Y_2 marker at X_0 to see if there is a better match (equivalent to ten times smaller value)
6. Choose suitable values of A and B ($R \times 10^A$ and $C \times 10^B$) so that $RC \times 10^{A+B} = \tau$

Note: Depending on the application A is usually between 2 and 6 (i.e. R is between 100Ω and 1MΩ.) and B can range from minus 12 (pico farads) to minus 2 (large electrolytics).

Example: $\tau = 2.04\mu s = 2.04 \times 10^{-6}$s and so $X_0 = 0.31$
The best fit is R = 3.0 and C = 6.8 so practical values are $R = 3 \times 10^3$ and $C = 6.8 \times 10^{-10}$ i.e. 3kΩ and 680pF.

Note: The same method may be used for calculating inductor/capacitor combinations where $LC = 1/(2\pi f_R)^2$ and f_R is the resonant frequency.



Three-stage filters

1. Introduction

High accuracy in this context refers to not only low phase error (c.f. two-stage filters) but also low magnitude error. With three-stage filters the phase error (imaginary or "quadrature") remains third order, with respect to frequency, but the in-phase (real) error becomes fourth order and negligible at a frequency sufficiently far away from the cut-off frequency.

The generalised networks for low-pass and high-pass filters are: -

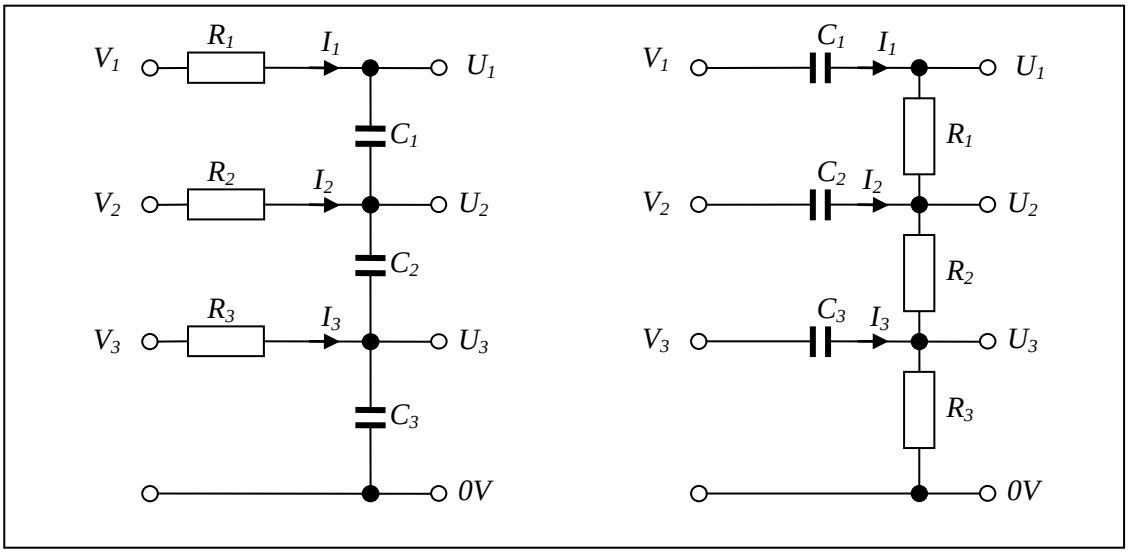


Fig. 1.1 Generalised three-stage filter networks

1.1 Network analysis

In practice the only applications known to the author are for actively driven high-pass filters. If the reader is interested, however, the matrices and equations for the low-pass network and circuits can be easily derived with the substitutions: $R_1 \rightarrow X_1$ and $X_1 \rightarrow R_1$ etc. For the high-pass network, according to Ohm's and Kirchhoff's Laws, the output voltages are: -

$$U_3 = R_3(I_1 + I_2 + I_3) \quad U_2 = R_3(I_1 + I_2 + I_3) + R_2(I_1 + I_2) \qquad U_1 = R_3(I_1 + I_2 + I_3) + R_2(I_1 + I_2) + R_1 I_1$$

The output impedance matrix equation is, therefore: -

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} R_1 + R_2 + R_3 & R_2 + R_3 & R_3 \\ R_2 + R_3 & R_2 + R_3 & R_3 \\ R_3 & R_3 & R_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} \quad \text{or} \quad \mathbf{U} = \mathbf{XI}$$

The input impedance matrix equation is: $\mathbf{V} = \mathbf{ZI}$

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} + \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} \qquad X_1 = \frac{1}{sC_1} \text{ etc}$$

$$\Rightarrow \quad \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} R_1 + R_2 + R_3 + X_1 & R_2 + R_3 & R_3 \\ R_2 + R_3 & R_2 + R_3 + X_2 & R_3 \\ R_3 & R_3 & R_3 + X_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix}$$

Invert this equation to calculate the input currents and impedances.

Define **Y** such that: $\mathbf{Y} = |\mathbf{Z}|\mathbf{Z}^{-1} \Rightarrow \qquad \mathbf{I} = \frac{\mathbf{Y}}{|\mathbf{Z}|}\mathbf{V}$

1.2 Cramer's Rule

For the inverse matrix one can use Cramer's Rule (see appendices 1 to 3): -

Given $V_i = Z_{ij} I_j$ then $I_j = \frac{|\mathbf{Z}[j]|}{|\mathbf{Z}|}$ where $\mathbf{Z}[j]$ is the **Z** matrix with the j^{th} column replaced by V_i.

Each determinant $|\mathbf{Z}[j]|$ is sorted into terms of V_i from which the components of the matrix **Y** can be extracted.

The process is laborious but not difficult. The algebra can be kept to a minimum if terms are retained in the form R^2, RX and X^2 then R^3, R^2X, RX^2 and X^3. These later translate easily into the coefficients in the polynomial in s.

The task is made easier and less error prone if all the equations required are first copied and pasted into a single equation object. The terms can then be copied and pasted and the unnecessary stuff deleted.

The determinant is (see appendix 2): -

$$|\mathbf{Z}| = R_1R_2R_3 + R_2R_3X_1 + R_1(R_2+R_3)X_3 + R_3(R_1+R_2)X_2 + (R_1+R_2+R_3)X_2X_3 + (R_2+R_3)X_1X_3 + R_3X_1X_2 + X_1X_2X_3$$

The components of **Y** for a high-pass network are (see appendix 1): -

First row: -

$$Y_{11} = R_2R_3 + (R_2+R_3)X_3 + R_3X_2 + X_2X_3$$
$$Y_{12} = -R_2R_3 - (R_2+R_3)X_3$$
$$Y_{13} = -R_3X_2$$

Second row: -

$$Y_{21} = -R_2R_3 - (R_2+R_3)X_3$$
$$Y_{22} = (R_1+R_2)R_3 + R_3X_1 + (R_1+R_2+R_3)X_3 + X_1X_3$$
$$Y_{23} = -R_1R_3 - R_3X_1$$

Third row: -

$$Y_{31} = -R_3X_2$$
$$Y_{32} = -R_1R_3 - R_3X_1$$
$$Y_{33} = R_1(R_2+R_3) + (R_2+R_3)X_1 + (R_1+R_2+R_3)X_2 + X_1X_2$$

Combine with the output impedance matrix equation to get the matrix equivalent of the transfer function equation (output versus input): -

$$\mathbf{U} = \mathbf{XI} = \mathbf{XZ}^{-1}\mathbf{V} = \mathbf{TV}$$

Define the matrix $\mathbf{\Omega}$ so that: $\mathbf{\Omega} = |\mathbf{Z}|\mathbf{T} = \mathbf{XY}$ and, therefore: $\mathbf{U} = \frac{1}{|\mathbf{Z}|}\mathbf{\Omega V}$

For the various filter configurations one need only calculate the determinant and those components of $\mathbf{\Omega}$ required.

2. Actively driven three–stage bootstrapped high pass filter (type 1)

2.1 Circuit analysis

The following is a JDY design used in an early version of the F18. The reason for connecting C_3 to the input, rather than being driven by the output (see type 2), remains a mystery.

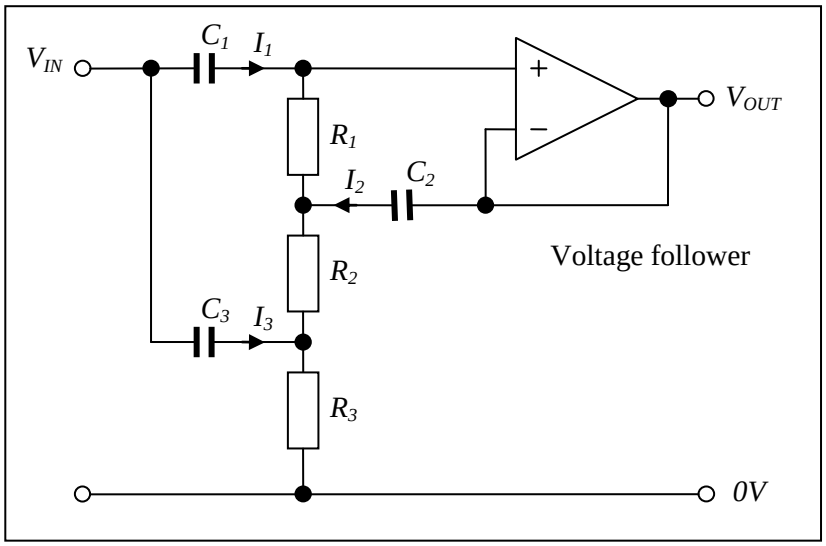


Fig. 2.1.1 A three-stage actively driven high pass filter – the "JDY special"

For the circuit, fig. 2.1.1: $V_{OUT} = U_1 = V_2$ and $V_{IN} = V_1 = V_3$

The output is, therefore: $$V_{OUT} = U_1 = \frac{1}{|\mathbf{Z}|}\left(\Omega_{11}V_1 + \Omega_{12}V_2 + \Omega_{13}V_3\right) = \frac{\left(\Omega_{11} + \Omega_{13}\right)V_{IN} + \Omega_{12}V_{OUT}}{|\mathbf{Z}|}$$

Re-arrange and simplify for the transfer function: $$T(s) = \frac{\Omega_{11} + \Omega_{13}}{|Z| - \Omega_{12}}$$

From appendix 2 the determinant is (with $X_1 = 1/sC_1$) etc: -

$$|\mathbf{Z}| = R_1R_2R_3 + R_2R_3X_1 + R_1(R_2 + R_3)X_3 + R_3(R_1 + R_2)X_2 + (R_1 + R_2 + R_3)X_2X_3 + (R_2 + R_3)X_1X_3 + R_3X_1X_2 + X_1X_2X_3$$

The required matrix elements are (from appendix 3): -

$$\Omega_{11} = R_1R_2R_3 + R_1(R_2 + R_3)X_3 + (R_1 + R_2)R_3X_2 + (R_1 + R_2 + R_3)X_2X_3$$
$$\Omega_{12} = R_2R_3X_1 + (R_2 + R_3)X_1X_3$$
$$\Omega_{13} = R_3X_1X_2$$

$$\Rightarrow T(s) = \frac{R_1R_2R_3 + R_1(R_2 + R_3)X_3 + (R_1 + R_2)R_3X_2 + (R_1 + R_2 + R_3)X_2X_3 + R_3X_1X_2}{R_1R_2R_3 + R_1(R_2 + R_3)X_3 + R_3(R_1 + R_2)X_2 + (R_1 + R_2 + R_3)X_2X_3 + R_3X_1X_2 + X_1X_2X_3}$$

Divide top and bottom by $X_1X_2X_3$ for the more conventional form: -

$$T(s) = \frac{R_1R_2R_3C_1C_2C_3s^3 + \{R_1(R_2 + R_3)C_1C_2 + R_3(R_1 + R_2)C_1C_3\}s^2 + \{(R_1 + R_2 + R_3)C_1 + R_3C_3\}s}{R_1R_2R_3C_1C_2C_3s^3 + \{R_1(R_2 + R_3)C_1C_2 + R_3(R_1 + R_2)C_1C_3\}s^2 + \{(R_1 + R_2 + R_3)C_1 + R_3C_3\}s + 1}$$

2.2 A basic type 1 filter

If the *Rs* and *Cs* have the same values the normalised transfer function $(s = j\omega/\omega_N)$is much simpler and of the form expected. Note that the coefficients are at the upper limit for a practical design:-

$$T(s) = \frac{s^3 + 4s^2 + 4s}{s^3 + 4s^2 + 4s + 1} \qquad \omega_N = \frac{1}{RC}$$

2.3 Different values of resistors and capacitors

The full transfer function is, in normalised form $(s = j\omega/\omega_N)$: $T(s) = \frac{s^3 + bs^2 + as}{s^3 + bs^2 + as + 1}$

With: $a = ((R_1 + R_2 + R_3)C_1 + R_3C_3)\omega_N \qquad b = (R_1(R_2 + R_3)C_1C_2 + R_3(R_1 + R_2)C_1C_3)\omega_N^2$

And $\omega_N = (R_1R_2R_3C_1C_2C_3)^{-\frac{1}{3}}$

Simulation shows that acceptable values for *a* and *b* are between 2 and 4. As with the two-stage filter the lower the value the greater is the resonant peaking and better attenuation at low frequency. Clearly, in this case, one needs to avoid the combination $(a = b = 1)$which results in an infinite peak (the denominator is zero at the natural frequency). The following depicts a range of values of pairs between *a =2 b* = 3 (highest peak), and *a =3 b* = 4 (lowest peak). See section 3.4 for a practical circuit and component values: -

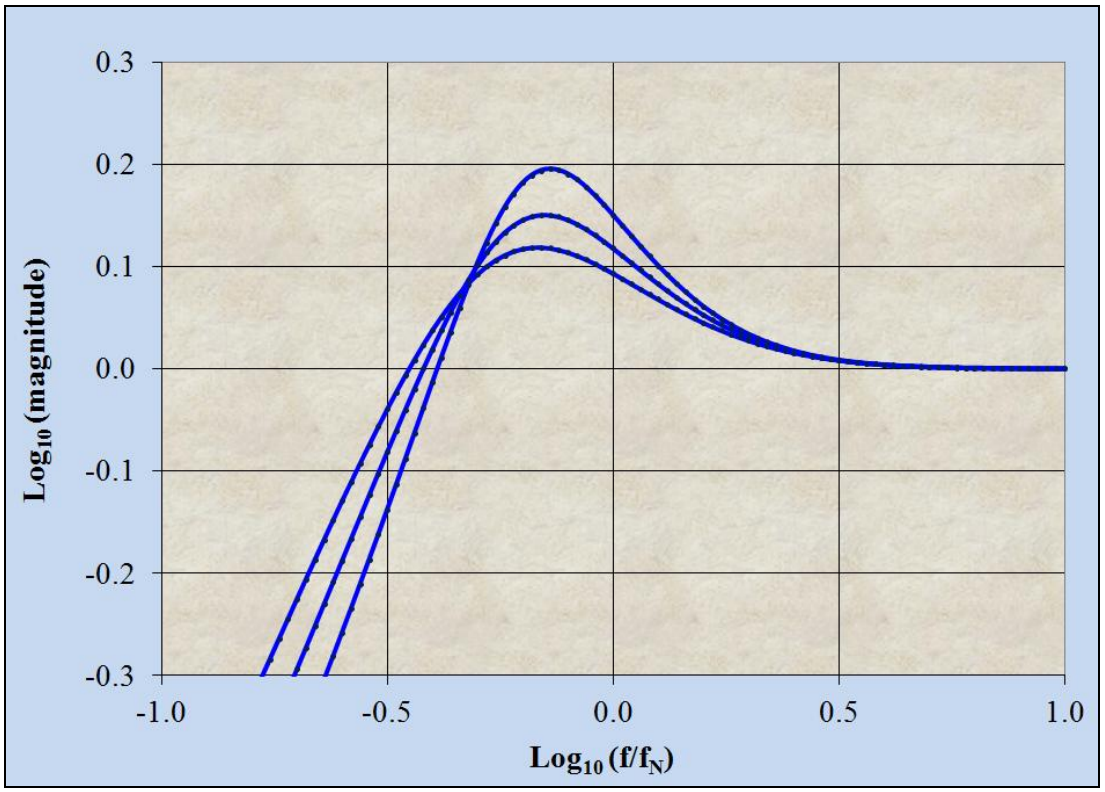


Fig. 2.3.1 Examples of peak response versus attenuation at low frequency (for dB ×20)

Note that the transfer function may be expressed: $T(s) = 1 - \frac{1}{s^3 + bs^2 + as + 1}$

At high frequency the magnitude of the second term is small and inversely proportional to frequency cubed. For a full analysis one needs to extract the significant real (in-phase) and imaginary (quadrature) error components….

2.4 Errors at high frequency $(\omega >> \omega_N)$

In normalised form $(s = j\,\omega/\omega_N)$ being careful to retain terms at least down to order s^{-4}: -

$$T(s) = 1 - \frac{1}{s^3 + bs^2 + as + 1} \approx 1 - \frac{1}{s^3}\left(\frac{1}{1 + \frac{b}{s} + \frac{a}{s^2} + \frac{1}{s^3}}\right) \approx 1 - \frac{1}{s^3}\left(1 - \frac{b}{s}\right)$$

To a very good approximation, therefore: -

$$\omega >> \omega_N \Rightarrow \quad T(s) \approx 1 - \frac{1}{s^3} + \frac{b}{s^4}$$

The real error term is proportional to *b* and of fourth order with respect to frequency. The parameter *a* contributes very little to the imaginary error (fifth order with respect to frequency). In more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 - j\left(\frac{f_N}{f}\right)^3 + b\left(\frac{f_N}{f}\right)^4 \text{ with } \quad f_N = \frac{1}{2\pi(R_1C_1R_2C_2R_3C_3)^{\frac{1}{3}}}$$

Note that the natural frequency is the geometric mean of three characteristic frequencies.

2.4.1 An interesting result

Note also that the phase error is numerically negative (phase lag) but this is only the case for high frequency. For lower frequencies the phase shift is positive (phase lead) as one would expect for a high pass filter: -

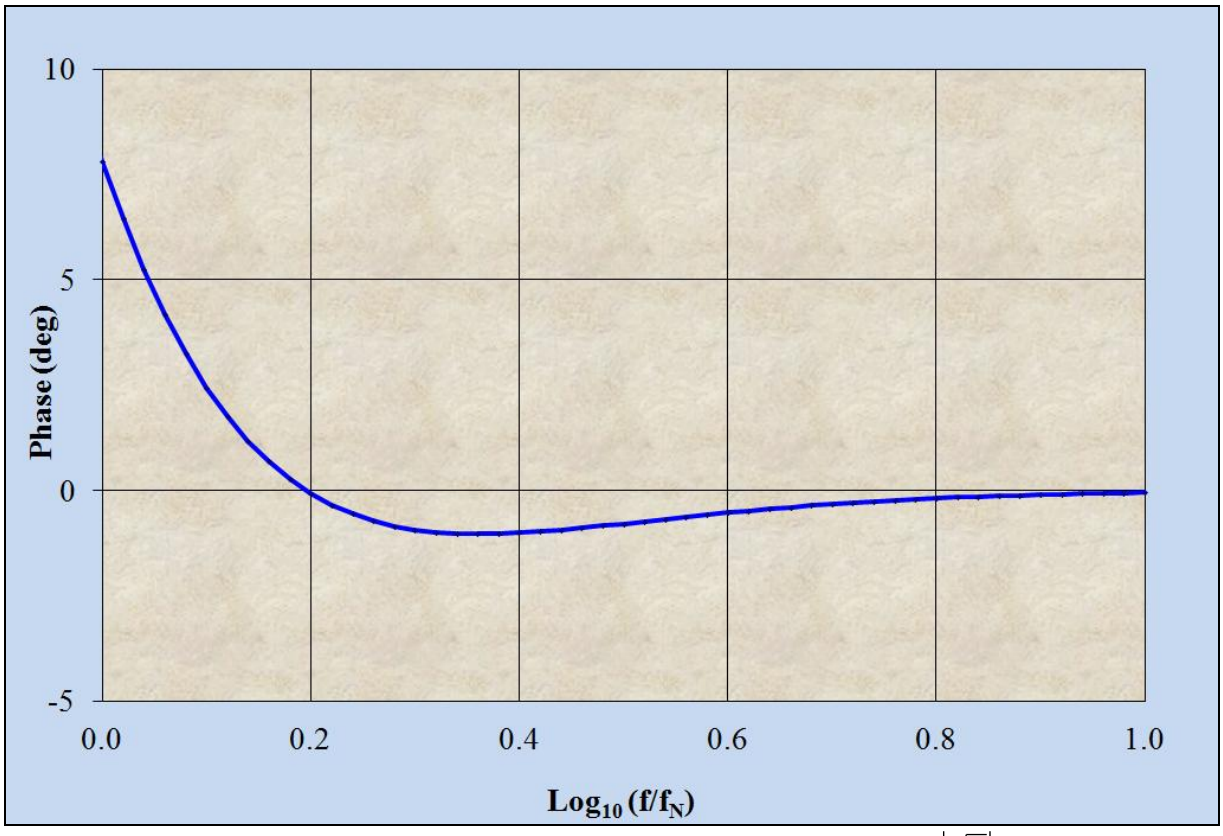


Fig. 2.4.1.1 The phase shift is negative at high frequency (*a* = 4 so $\log\left|\sqrt{a}\right| = 0.301$)

The frequency at which the phase becomes zero is an interesting result. If one assumes that the natural frequency, $\omega_N = 1$ (to simplify the algebra) then the transfer function is: -

$$T(s) = \frac{-b\omega^2 + j(a\omega - \omega^3)}{1 - b\omega^2 + j(a\omega - \omega^3)}$$

Multiply top and bottom by the complex conjugate $1 - b\omega^2 - j(a\omega - \omega^3)$ to make the denominator real: -

$$T(s) = \frac{(-b\omega^2 + j(a\omega - \omega^3))(1 - b\omega^2 - j(a\omega - \omega^3))}{(1 - b\omega^2)^2 + (a\omega - \omega^3)^2}$$

The imaginary part is zero for zero phase shift: -

$$b\omega^2(a\omega - \omega^3) + (1 - b\omega^2)(a\omega - \omega^3) = a\omega - \omega^3 = 0$$

$$a\omega - \omega^3 = 0 \quad \Rightarrow \quad \omega = 0 \qquad \text{or} \qquad \omega = \pm\sqrt{a}$$

Clearly the positive result is valid. More generally, for zero phase shift: $\omega = \sqrt{a}\omega_N$

2.5 Calculating component values

Given the coefficients *a* and *b* one can determine the nearest preferred component values using a similar method as with the two-stage design. This time it is a bit more complicated as there are six parameters to consider: -

Define a baseline capacitance and resistance value so that: $f_N = \frac{1}{2\pi RC}$

If possible choose cardinal values (e.g. $f_N = 1.6Hz; R = 1M\Omega; C = 100nF$) or else use the slide rule calculator (see appendix 4). Define, also, parameters $\alpha \ \beta \ \gamma \ \delta \ \varepsilon \ \theta$ (all > 0) such that: -

$$R_1 = \alpha R \qquad R_2 = \beta R \qquad R_3 = \gamma R \qquad C_1 = \delta C \qquad C_2 = \varepsilon C \qquad C_3 = \theta C$$

$$\text{and} \qquad \alpha\beta\gamma\delta\varepsilon\theta = 1$$

Substitution into the equations for a and b results in some simplification: -

$$a = ((R_1 + R_2 + R_3)C_1 + R_3C_3)\omega_N \qquad b = (R_1(R_2 + R_3)C_1C_2 + R_3(R_1 + R_2)C_1C_3)\omega_N^2$$

$$\Rightarrow \qquad a = \delta(\alpha + \beta + \gamma) + \gamma\theta \qquad b = \delta\{\alpha\varepsilon(\beta + \gamma) + \gamma\theta(\alpha + \beta)\} \qquad \text{and} \qquad \alpha\beta\gamma\delta\varepsilon\theta = 1$$

The equations are still complicated, however, and it is not obvious how to choose the best combination of parameters. Given *a* and *b* there are six "unknowns" but only three equations (constraints). One can afford to introduce more constraints in order to simplify the problem. For example: -

In the equations for a and b the parameters δ and γ seem to play the main role. One could choose δ and γ to be less than 1 in order to achieve near the minimum values – usually the case. One or more of the other parameters must compensate by being greater than 1. Looking at how they combine the most likely candidates are ε and θ. Consider the further constraints: -

$$\delta = \frac{1}{\varepsilon} \qquad \text{and} \qquad \gamma = \frac{1}{\theta} \qquad \Rightarrow \qquad \alpha = \frac{1}{\beta}$$

The equations are now much simpler with the lower limits $a = b = 1$: -

$$a = 1 + \delta\gamma + \delta\left(\alpha + \frac{1}{\alpha}\right) \quad \text{and} \quad b = 1 + \alpha\gamma + \delta\left(\alpha + \frac{1}{\alpha}\right) \quad \Rightarrow \quad b - a = (\alpha - \delta)\gamma$$

There are now three unknowns and only two equations. It is possible, in principle, therefore to select a value for one of the unknowns and then calculate the others. The algebra turns out to be a bit complicated but, fortunately, unnecessary. Simulation shows that *a* = *b* provides a useful range of choices. Consider one more constraint: -

$$\alpha = \delta \Rightarrow \qquad a = b = 2 + \delta\gamma + \delta^2$$

The minimum value for both *a* and *b* is now 2, but that is also OK. One can choose one of the limited practical values for δ and then calculate the value of γ required to achieve the target *a* and *b*.

Consider one last constraint to make the calculation even simpler: -

$$\gamma = \delta \Rightarrow \qquad a = b = 2 + 2\delta^2$$

The range of choice is now quite restricted. Fortunately the range of standard capacitor values is sufficient for the range of filter characteristics usually required $(2 < a, b < 4)$. A useful example is a design for operation at 75Hz (1.5 × supply frequency). A natural frequency of 1.6Hz provides sufficiently low in-phase and quadrature errors:-

$$f_N = 1.6Hz \quad \text{and} \quad f = 75Hz \Rightarrow \; T(f) \approx 1 - j\left(\frac{f_N}{f}\right)^3 + b\left(\frac{f_N}{f}\right)^4 \approx 1 - j10^{-5} + b \times 2 \times 10^{-7}$$

It is no coincidence that a natural frequency of 1.6Hz allows for cardinal values for base values *R* and *C* and so the reciprocal relations between the most common preferred values makes selection easy.

$0.1 \approx \frac{1}{10}$ $\quad 0.22 \approx \frac{1}{4.7}$ $\quad 0.33 \approx \frac{1}{3.3}$ $\quad 0.68 \approx \frac{1}{1.5}$ and vice versa, with a worst case error of 10%

$\alpha = \delta = \gamma =$	0.1	0.22	0.33	0.47	0.68	1.0
$\beta = \varepsilon = \theta =$	10	4.7	3.3	2.2	1.5	1.0
$a = b =$	2.02	2.10	2.22	2.44	2.92	4

According to the strategy above the values of R_1, C_1 and R_3 are kept low while R_2, C_2 and C_3 are correspondingly high relative to the base values: -

$$\alpha = \delta = \gamma = \frac{1}{\beta} = \frac{1}{\varepsilon} = \frac{1}{\theta} \qquad \Rightarrow \qquad a = b = 2 + 2\delta^2$$

This is equivalent to having a relatively low time constant R_1C_1, a high value for R_2C_2 and a medium value for R_3C_3 so that the geometric mean of the corresponding frequencies is the natural frequency.

$$f_N = \frac{1}{2\pi\left(R_1C_1R_2C_2R_3C_3\right)^{\frac{1}{3}}}$$

As a rule of thumb, therefore, a low value for parameters *a* and *b* is obtained by a larger difference between the low time constant and the high time constant (equivalent to a low value for δ). It should be noted, however, that the parameters γ and θ do not appear in the equations in a symmetrical way and, therefore, the medium time constant must be obtained by a low value for R_3 and a correspondingly high value for C_3.

2.5.1 Example calculation

For part of an early F18 design JDY found a nice combination by a method which remains a mystery.

$$R_1 = R_2 = 160k\Omega \qquad R_3 = 91k\Omega \qquad C_1 = 680nF \qquad C_2 = 1\mu F \qquad C_3 = 2.2\mu F$$

$$f_N = \frac{1}{2\pi\left(R_1C_1R_2C_2R_3C_3\right)^{\frac{1}{3}}} = 1.05Hz$$

Clearly a natural frequency of 1Hz was the target. The most convenient way to calculate *a* and *b* is now to choose base values for 1.05Hz: -

$$C = 1\mu F \qquad \text{and} \qquad R = 152k\Omega$$

$$a = (\alpha + \beta + \gamma)\delta + \gamma\theta = \frac{(160 + 160 + 91) \times 0.68 + 91 \times 2.2}{152} = 3.15$$

$$b = \delta\{\alpha\varepsilon(\beta + \gamma) + \gamma\theta(\alpha + \beta)\} = 0.68 \times \left\{\frac{160}{152}\frac{(160 + 91)}{152} + \frac{91}{152} \times 2.2 \times \frac{160 + 160}{152}\right\} = 3.07$$

The lowest operating frequency is 25Hz and the application was for AC coupling the last three digits (decades 5, 6 and 7) from a multiplying R-2R DAC to a pair of two-stage transformers (each of 100:1 step down) with a low DC error follower (AD547 op-amp). At 25Hz: -

$$f_N = 1.0Hz \Rightarrow \qquad T(f) \approx 1 - j\left(\frac{f_N}{f}\right)^3 + b\left(\frac{f_N}{f}\right)^4 \approx 1 - j6 \times 10^{-5} + 7 \times 10^{-6}$$

With 7ppm in-phase error the performance is well within the error budget required (i.e. 100ppm in-phase to be safe) and the quadrature error is negligible (also reduced by a factor of 10,000). A natural frequency of 1.6Hz would have been perfectly satisfactory. Using the method above with a slightly reduced target for $a = b = 2.44$: -

Choose base values: $C = 1\mu F$ and $R = 100k\Omega$ $f_N = 1.6Hz$

$$\alpha = \delta = \gamma = 0.47 \qquad \beta = \varepsilon = \theta = 2.2$$

$$a = b = 2 + 2\delta^2 = 2.44$$

$$R_1 = 47k\Omega \qquad R_2 = 220k\Omega \qquad R_3 = 47k\Omega \qquad C_1 = 470nF \qquad C_2 = 2.2\mu F \qquad C_3 = 2.2\mu F$$

Alternatively one could increase the resistors and decrease the capacitors by a factor of ten. At 25Hz: -

$$f_N = 1.6Hz \Rightarrow \qquad T(f) \approx 1 - j\left(\frac{f_N}{f}\right)^3 + b\left(\frac{f_N}{f}\right)^4 \approx 1 - j2.6 \times 10^{-4} + 4.1 \times 10^{-5}$$

The in-phase error of 41ppm, reduced by a factor of 10,000, contributes an error of 4.1ppb!

The reader may ask "what about the voltage follower error?" At low frequency this is not usually an issue and a reasonably good single op-amp should suffice [1]. There is always the option, however, to use a high accuracy voltage follower [2], based on a two or three-stage high gain block [1].

1. Part 4, monograph 1: "High gain blocks"
2. Part 4, monograph 2: "High accuracy voltage followers"

3. Actively driven three–stage high-pass filter (type 2)

3.1 Circuit analysis

An interesting variant is to drive both lower stages of the network from the output. The filter characteristic is not hugely different compared to type 1 but it does have much higher input impedance. It also opens up the possibility of extremely high input impedance (very small C_1 and very large R_1). See the monograph "An ultra-high input impedance high-pass filter" by the same author [1].

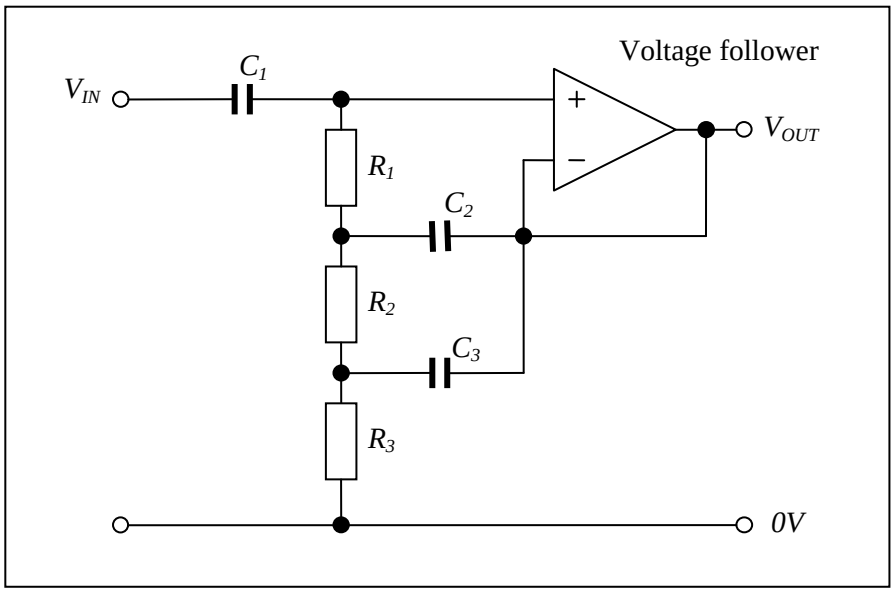


Fig. 3.1.1 A three-stage actively driven high pass filter

Fortunately most of the analysis required has been done for the network (see section 1.1). For the circuit fig. 3.1, assuming an ideal follower: -

$$V_{OUT} = U_1 = V_2 = V_3 \quad \text{and} \quad V_{IN} = V_1$$

The output is, therefore: -

$$V_{OUT} = U_1 = \frac{\Omega_{11}V_1 + \Omega_{12}V_2 + \Omega_{13}V_3}{|\mathbf{Z}|} = \frac{\Omega_{11}V_{IN} + (\Omega_{12} + \Omega_{13})V_{OUT}}{|\mathbf{Z}|}$$

Re-arrange and simplify for the transfer function: $$T(s) = \frac{\Omega_{11}}{|\mathbf{Z}| - \Omega_{12} - \Omega_{13}}$$

From appendix 2 the determinant is (with $X_1 = 1/sC_1$) etc: -

$$|\mathbf{Z}| = R_1R_2R_3 + R_2R_3X_1 + R_1(R_2 + R_3)X_3 + R_3(R_1 + R_2)X_2 + (R_1 + R_2 + R_3)X_2X_3 + (R_2 + R_3)X_1X_3 + R_3X_1X_2 + X_1X_2X_3$$

The required matrix elements are (from appendix 3): -

$$\Omega_{11} = R_1R_2R_3 + R_1(R_2 + R_3)X_3 + (R_1 + R_2)R_3X_2 + (R_1 + R_2 + R_3)X_2X_3$$
$$\Omega_{12} = R_2R_3X_1 + (R_2 + R_3)X_1X_3$$
$$\Omega_{13} = R_3X_1X_2$$

1. Part 2, monograph 3: "An ultra-high input impedance high-pass filter"

With the result: -

$$T(s)=\frac{R_1R_2R_3+R_1(R_2+R_3)X_3+(R_1+R_2)R_3X_2+(R_1+R_2+R_3)X_2X_3}{R_1R_2R_3+R_1(R_2+R_3)X_3+R_3(R_1+R_2)X_2+(R_1+R_2+R_3)X_2X_3+X_1X_2X_3}$$

Divide top and bottom by $X_1X_2X_3$: -

$$T(s)=\frac{R_1R_2R_3C_1C_2C_3s^3+\{R_1(R_2+R_3)C_1C_2+R_3(R_1+R_2)C_1C_3\}s^2+(R_1+R_2+R_3)C_1s}{R_1R_2R_3C_1C_2C_3s^3+\{R_1(R_2+R_3)C_1C_2+R_3(R_1+R_2)C_1C_3\}s^2+(R_1+R_2+R_3)C_1s+1}$$

3.2 A basic type 2 filter

If the *Rs* and *Cs* are the same the normalised transfer function is $(s=j\omega/\omega_N)$: -

$$T(s)=\frac{s^3+4s^2+3s}{s^3+4s^2+3s+1}\qquad \omega_N=(R_1R_2R_3C_1C_2C_3)^{-\frac{1}{3}}$$

The result is a slight increase in peaking compared to the basic type 1 three-stage filter. The quadrature and in-phase errors are the same: -

$$f>>f_N\Rightarrow\quad T(f)\approx 1+j\left(\frac{f_N}{f}\right)^3+4\left(\frac{f_N}{f}\right)^4\qquad f_N=\frac{1}{2\pi RC}$$

3.3 Different values of resistors and capacitors

The full transfer function is, in normalised form $(s=j\omega/\omega_N)$: $\quad T(s)=\frac{s^3+bs^2+as}{s^3+bs^2+as+1}$

With $\quad a=(R_1+R_2+R_3)C_1\omega_N\qquad b=(R_1(R_2+R_3)C_1C_2+R_3(R_1+R_2)C_1C_3)\omega_N^2$

And $\quad \omega_N=(R_1R_2R_3C_1C_2C_3)^{-\frac{1}{3}}$

Note that the equation for *b* is the same as for a type 1 high pass filter but the equation for *a* has lost a term.

Note, also, that the transfer function may again be expressed as: -

$$T(s)=1-\frac{1}{s^3+bs^2+as+1}$$

To a very good approximation, therefore, also in normalised form $(s=j\omega/\omega_N)$ the error analysis is the same as for a type 1 filter (see section 2.4): -

$$\omega>>\omega_N\qquad\Rightarrow\qquad T(s)\approx 1-\frac{1}{s^3}+\frac{b}{s^4}$$

In more convenient form, with frequency in Hz: -

$$f>>f_N\Rightarrow\quad T(f)\approx 1-j\left(\frac{f_N}{f}\right)^3+b\left(\frac{f_N}{f}\right)^4\ \text{with}\qquad f_N=\frac{1}{2\pi(R_1C_1R_2C_2R_3C_3)^{\frac{1}{3}}}$$

3.4 Calculating component values

As before choose baseline capacitor and resistor values so that: $\omega_N = \frac{1}{RC}$

Define parameters α β γ δ ε θ (all > 0) such that: -

$$R_1 = \alpha R \qquad R_2 = \beta R \qquad R_3 = \gamma R \qquad C_1 = \delta C \qquad C_2 = \varepsilon C \qquad C_3 = \theta C$$

and $\alpha\beta\gamma\delta\varepsilon\theta = 1$

Substitution into the equation for a and b results in some simplification: -

$$a = \delta(\alpha + \beta + \gamma) \qquad b = \delta\{\alpha\varepsilon(\beta + \gamma) + \gamma\theta(\alpha + \beta)\}$$

N.B. The equation for *a* has lost the term $\gamma\theta$. To simplify further apply the same constraints as before: -

$$\varepsilon = \frac{1}{\delta} \quad \theta = \frac{1}{\gamma} \quad \Rightarrow \alpha = \frac{1}{\beta}$$

$$a = \delta\gamma + \delta\left(\alpha + \frac{1}{\alpha}\right) \qquad b = 1 + \alpha\gamma + \delta\left(\alpha + \frac{1}{\alpha}\right)$$

The difference is: $b - a = 1 + (\alpha - \delta)\gamma$

One could make *a* and *b* the same but further analysis shows that both would have to be large. Fortunately it is OK to have $b = a + 1$ with the further simplification: -

$$\alpha = \delta \Rightarrow \qquad a = 1 + \delta\gamma + \delta^2 \qquad \text{and} \qquad b = 2 + \delta\gamma + \delta^2$$

Finally, consider the constraint: $\gamma = \frac{1}{\delta} \Rightarrow$ $\qquad a = 2 + \delta^2 \qquad$ and $\qquad b = 3 + \delta^2$

The range of choice is again restricted but the range of standard values is sufficient for most applications: -

$$\alpha = \delta = \theta = \frac{1}{\beta} = \frac{1}{\gamma} = \frac{1}{\varepsilon}$$

$\alpha = \delta = \theta =$	0.1	0.22	0.33	0.47	0.68	1.0
$\beta = \gamma = \varepsilon =$	10	4.55(4.7)	3.03(3.3)	2.13(2.2)	1.47(1.5)	1.0
$a =$	2.01	2.05	2.11	2.22	2.46	3
$b =$	3.01	3.05	3.11	3.22	3.46	4

N.B. In this case a good compromise is the simplest: the component values being the same (*a* = 3, *b* = 4) and another advantage of the type 2 filter. The following includes, for the highest peak (a = 2, b = 3), middle (a = 2.46, b = 3.46) and the lowest (a = 3, b = 4): -

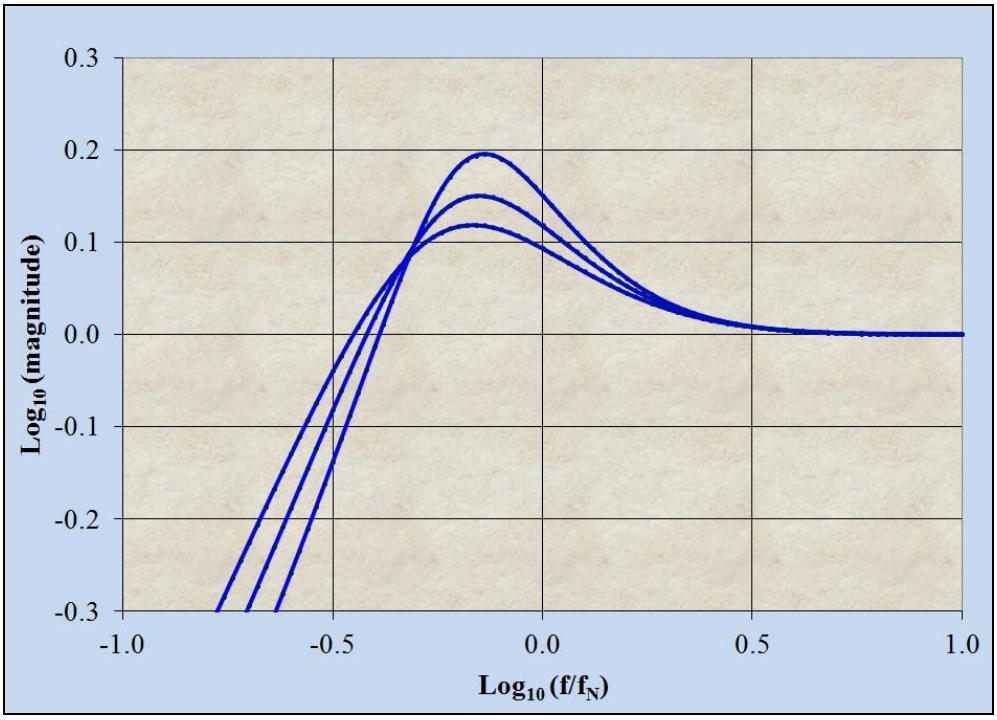


Fig.3.4.1 Peaking versus attenuation (for dB ×20)

3.5 Input impedance

One of the main advantages of a type 2 high pass filter is the extra boost to the input impedance, compared to a two-stage filter. The input current is: -

$$I_1 = \frac{Y_{11}V_1 + Y_{12}V_2 + Y_{13}V_3}{|\mathbf{Z}|} \qquad \text{with} \qquad V_1 = V_{IN} \qquad \text{and} \qquad V_2 = V_3 = V_{OUT} = T(s)V_{IN}$$

From appendix 2 the determinant is (with $X_1 = 1/sC_1$) etc: -

$$|\mathbf{Z}| = R_1R_2R_3 + R_2R_3X_1 + R_1(R_2 + R_3)X_3 + R_3(R_1 + R_2)X_2 + (R_1 + R_2 + R_3)X_2X_3 + (R_2 + R_3)X_1X_3 + R_3X_1X_2 + X_1X_2X_3$$

The first row of the matrix consists of the following elements (see appendix 1): -

$$Y_{11} = R_2R_3 + (R_2 + R_3)X_3 + R_3X_2 + X_2X_3 \qquad Y_{12} = -R_2R_3 - (R_2 + R_3)X_3 \qquad \text{and} \qquad Y_{13} = -R_3X_2$$

$$I_1 = \frac{Y_{11} + (Y_{12} + Y_{13})T(s)}{|\mathbf{Z}|}V_{IN}$$

$$\Rightarrow \qquad I_1 = \frac{R_2R_3 + (R_2 + R_3)X_3 + R_3X_2 + X_2X_3 - (R_2R_3 + (R_2 + R_3)X_3 + R_3X_2)T(s)}{|\mathbf{Z}|}V_{IN}$$

$$\Rightarrow \qquad I_1 = \frac{X_2X_3 + (R_2R_3 + (R_2 + R_3)X_3 + R_3X_2)(1 - T(s))}{|\mathbf{Z}|}V_{IN}$$

At high frequency the impedance of the capacitors is much smaller than the resistors and so, to a good approximation, the input admittance is: -

$$A_{IN} = \frac{I_1}{V_{IN}} \approx \frac{X_2X_3 + R_2R_3(1-T(s))}{|\mathbf{Z}|}$$

Divide top and bottom by $X_1X_2X_3$ (with $X_1 = 1/sC_1$) etc but retain the term in $1/X_1$ for reasons that will become clear. Retain, also, the second largest term in the denominator until it becomes clear when it is insignificant: -

$$A_{IN} = \frac{1/X_1 + s^3C_1R_2C_2R_3C_3(1-T(s))}{s^3R_1C_1R_2C_2R_3C_3 + s^2(R_2C_2R_3C_3 + R_1(R_2+R_3)C_1C_2 + R_3(R_1+R_2)C_1C_3) + \ldots}$$

The factor $(1-T(s))$ is small at high frequency and, in normalised form $(s = j\omega/\omega_N)$: -

From section 3.3: $\omega >> \omega_N \Rightarrow \quad T(s) \approx 1 - \frac{1}{s^3} + \frac{b}{s^4}$ with $\omega_N = (R_1R_2R_3C_1C_2C_3)^{-\frac{1}{3}}$

$$\Rightarrow \qquad 1 - T(s) \approx \frac{1}{s^3}\left(1 - \frac{b}{s}\right)$$

Also in normalised form: $s^3C_1R_2C_2R_3C_3 \rightarrow \frac{s^3}{R_1}$

And: $s^3R_1C_1R_2C_2R_3C_3 + s^2(R_2C_2R_3C_3 + R_1(R_2+R_3)C_1C_2 + R_3(R_1+R_2)C_1C_3) \rightarrow s^3\left(1 + \frac{\alpha}{s}\right)$

With the dimensionless parameter: -

$$\alpha = (R_2C_2R_3C_3 + R_1(R_2+R_3)C_1C_2 + R_3(R_1+R_2)C_1C_3)\omega_N^2$$

To a very good approximation, therefore, in normalised form $(s = j\omega/\omega_N)$: -

$$A_{IN} \approx \left(\frac{1}{X_1} + \frac{1}{R_1}\left(1 - \frac{b}{s}\right)\right)\frac{1}{s^3}\left(1 - \frac{\alpha}{s}\right)$$

Now $1/X_1$ is imaginary and $1/R_1$ is real and the small corrections due to α and b are insignificant. To a very good approximation, therefore, the significant real and imaginary components are contained within: -

$$A_{IN} \approx \frac{1}{s^3}\left(\frac{1}{X_1} + \frac{1}{R_1}\right)$$

The input admittance is the capacitance, C_1, in parallel with resistance, R_1 reduced in proportion to frequency cubed and also subject to a phase shift. In more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \quad A_{IN} \approx j\left(\frac{f_N}{f}\right)^3\left(\frac{1}{X_1} + \frac{1}{R_1}\right)$$

Alternatively one can express the components of the parallel impedances due to the capacitor and resistor as follows: -

$$A_{IN} = \frac{1}{Z_{IN}} = \frac{1}{Z_C} + \frac{1}{Z_R} \qquad \text{with} \qquad Z_C = -\left(\frac{f}{f_N}\right)^2 \frac{1}{2\pi f_N C_1} \text{ and} \qquad Z_R = -j\left(\frac{f}{f_N}\right)^3 R_1$$

The component due to the capacitor becomes negative real and that due to the resistor is negative imaginary (i.e. a very large negative resistance in parallel with a very small capacitance). Both are frequency dependent. In practice the impedance of the capacitor is smaller than that of the resistor (i.e. its admittance is much larger) and the input impedance is, therefore, substantially a large negative resistor.

3.5.1 Example calculation

It is entirely possible for a high input impedance (e.g. JFET input) follower to permit a value of R_1 as high as 5GΩ. Such resistors are readily available and quite accurate and stable. The corresponding capacitance value for a stable filter with a natural frequency of 1Hz thus results in: -

$$R_1 = 5G\Omega \qquad \text{and} \qquad f_N \approx 1Hz \qquad \Rightarrow \qquad C_1 \approx 33pF$$

The impedance of C_1 at the natural frequency is: $X_1(f_N) = \frac{1}{sC_1} = \frac{1}{j2\pi f_N C_1} = -j4.8G\Omega$

At 75Hz this becomes: $Z_C = -\left(\frac{75}{1}\right)^2 \times \frac{1}{2\pi f_N C_1} \approx -27T\Omega$

The imaginary component (due to the resistor) at 75Hz is: $Z_R = -j\left(\frac{f}{f_N}\right)^3 R_1 \approx -2\,jP\Omega$ (Peta = 10^{15})

At high frequency the imaginary component is significantly larger than the real component and the effect of the capacitor dominates. Clearly this is so large that other factors must be considered. It is possible, however, to employ very good insulators (e.g. PTFE, ceramic or sapphire) an active guard and the "inside-out" configuration to eliminate the effects of stray capacitance. See the monograph "An ultra-high input impedance high-pass filter" by the same author [1].

1. Part 2, monograph 3: "An ultra-high input impedance high-pass filter"

Appendix 1: Cramer's method for matrix inversion (Three-stage high-pass network)

Given $V_i = Z_{ij} I_j$ then $I_j = \dfrac{|\mathbf{Z}[j]|}{|\mathbf{Z}|}$ where $\mathbf{Z}[j]$ = the **Z** matrix with the j^{th} column replaced by V_i.

The determinant is moved to the left hand side for simplicity: -

For $\boldsymbol{I_1}$ the $\boldsymbol{V_i}$ go in the first column $(j=1)$: -

$$|\mathbf{Z}|I_1 = \begin{vmatrix} V_1 & Z_{12} & Z_{13} \\ V_2 & Z_{22} & Z_{23} \\ V_3 & Z_{32} & Z_{33} \end{vmatrix} = V_1(Z_{22}Z_{33} - Z_{23}Z_{32}) + Z_{12}(Z_{23}V_3 - V_2Z_{33}) + Z_{13}(V_2Z_{32} - Z_{22}V_3)$$

$$= (Z_{22}Z_{33} - Z_{23}Z_{32})V_1 + (Z_{13}Z_{32} - Z_{12}Z_{33})V_2 + (Z_{12}Z_{23} - Z_{13}Z_{22})V_3 = Y_{11}V_1 + Y_{12}V_2 + Y_{13}V_3$$

Second column $(j=2)$: -

$$|\mathbf{Z}|I_2 = \begin{vmatrix} Z_{11} & V_1 & Z_{13} \\ Z_{21} & V_2 & Z_{23} \\ Z_{31} & V_3 & Z_{33} \end{vmatrix} = Z_{11}(V_2Z_{33} - Z_{23}V_3) + V_1(Z_{23}Z_{31} - Z_{21}Z_{33}) + Z_{13}(Z_{21}V_3 - V_2Z_{31})$$

$$= (Z_{23}Z_{31} - Z_{21}Z_{33})V_1 + (Z_{11}Z_{33} - Z_{13}Z_{31})V_2 + (Z_{13}Z_{21} - Z_{11}Z_{23})V_3 = Y_{21}V_1 + Y_{22}V_2 + Y_{23}V_3$$

Third column $(j=3)$: -

$$|\mathbf{Z}|I_3 = \begin{vmatrix} Z_{11} & Z_{12} & V_1 \\ Z_{21} & Z_{22} & V_2 \\ Z_{31} & Z_{32} & V_3 \end{vmatrix} = Z_{11}(Z_{22}V_3 - V_2Z_{32}) + Z_{12}(V_2Z_{31} - Z_{21}V_3) + V_1(Z_{21}Z_{32} - Z_{22}Z_{31})$$

$$= (Z_{21}Z_{32} - Z_{22}Z_{31})V_1 + (Z_{12}Z_{31} - Z_{11}Z_{32})V_2 + (Z_{11}Z_{22} - Z_{12}Z_{21})V_3 = Y_{31}V_1 + Y_{32}V_2 + Y_{33}V_3$$

The matrix for **Z** is restated for convenience: $\mathbf{Z} = \begin{pmatrix} R_1 + R_2 + R_3 + X_1 & R_2 + R_3 & R_3 \\ R_2 + R_3 & R_2 + R_3 + X_2 & R_3 \\ R_3 & R_3 & R_3 + X_3 \end{pmatrix}$

From which the first row of **Y** can be calculated: -

$$Y_{11} = (Z_{22}Z_{33} - Z_{23}Z_{32}) = (R_2 + R_3 + X_2)(R_3 + X_3) - R_3^2 = R_2R_3 + (R_2 + R_3)X_3 + R_3X_2 + X_2X_3$$

$$Y_{12} = (Z_{13}Z_{32} - Z_{12}Z_{33}) = R_3^2 - (R_2 + R_3)(R_3 + X_3) = -R_2R_3 - (R_2 + R_3)X_3$$

$$Y_{13} = (Z_{12}Z_{23} - Z_{13}Z_{22}) = (R_2 + R_3)R_3 - R_3(R_2 + R_3 + X_2) = -R_3X_2$$

Second row: -

$$Y_{21} = (Z_{23}Z_{13} - Z_{21}Z_{33}) = R_3^2 - (R_2 + R_3)(R_3 + X_3) = -R_2R_3 - (R_2 + R_3)X_3$$

$$Y_{22} = (Z_{11}Z_{33} - Z_{13}Z_{31}) = (R_1 + R_2 + R_3 + X_1)(R_3 + X_3) - R_3^2 = (R_1 + R_2)R_3 + R_3X_1 + (R_1 + R_2 + R_3)X_3 + X_1X_3$$

$$Y_{23} = (Z_{13}Z_{21} - Z_{11}Z_{23}) = R_3(R_2 + R_3) - (R_1 + R_2 + R_3 + X_1)R_3 = -R_1R_3 - R_3X_1$$

Third row: -

$$Y_{31} = (Z_{21}Z_{32} - Z_{22}Z_{31}) = (R_2 + R_3)R_3 - (R_2 + R_3 + X_2)R_3 = -R_3X_2$$

$$Y_{32} = (Z_{12}Z_{31} - Z_{11}Z_{32}) = (R_2 + R_3)R_3 - (R_1 + R_2 + R_3 + X_1)R_3 = -R_1R_3 - R_3X_1$$

$$Y_{33} = (Z_{11}Z_{22} - Z_{12}Z_{21}) = (R_1 + R_2 + R_3 + X_1)(R_2 + R_3 + X_2) - (R_2 + R_3)(R_2 + R_3)$$

$$= R_1(R_2 + R_3) + (R_2 + R_3)X_1 + (R_1 + R_2 + R_3)X_2 + X_1X_2$$

Appendix 2: Calculating the determinant (Three-stage high-pass network)

$$|\mathbf{Z}| = \begin{vmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{vmatrix} = Z_{11}\left(Z_{22}Z_{33} - Z_{21}Z_{32}\right) + Z_{12}\left(Z_{23}Z_{31} - Z_{21}Z_{33}\right) + Z_{13}\left(Z_{21}Z_{32} - Z_{22}Z_{31}\right)$$

Take each element in turn: -

$$\begin{aligned} &Z_{11}\left(Z_{22}Z_{33} - Z_{21}Z_{32}\right) \\ &= \left(R_1 + R_2 + R_3 + X_1\right)\left\{\left(R_2 + R_3 + X_2\right)\left(R_3 + X_3\right) - R_3^2\right\} = \left(R_1 + R_2 + R_3 + X_1\right)\left\{R_2R_3 + \left(R_2 + R_3\right)X_3 + R_3X_2 + X_2X_3\right\} \\ &= \left(R_1 + R_2 + R_3\right)R_2R_3 + R_2R_3X_1 + \left(R_1 + R_2 + R_3\right)\left(R_2 + R_3\right)X_3 + \left(R_1 + R_2 + R_3\right)R_3X_2 + \left(R_1 + R_2 + R_3\right)X_2X_3 + \left(R_2 + R_3\right)X_1X_3 + R_3X_1X_2 + X_1X_2X_3 \end{aligned}$$

$$\begin{aligned} &Z_{12}\left(Z_{23}Z_{31} - Z_{21}Z_{33}\right) \\ &= \left(R_2 + R_3\right)\left\{R_3^2 - \left(R_2 + R_3\right)\left(R_3 + X_3\right)\right\} = \left(R_2 + R_3\right)\left\{-R_2R_3 - \left(R_2 + R_3\right)X_3\right\} \\ &= -\left(R_2 + R_3\right)R_2R_3 - \left(R_2 + R_3\right)^2 X_3 \end{aligned}$$

$$\begin{aligned} &Z_{13}\left(Z_{21}Z_{32} - Z_{22}Z_{31}\right) \\ &= R_3\left\{\left(R_2 + R_3\right)R_3 - \left(R_2 + R_3 + X_2\right)R_3\right\} \\ &= -R_3^2X_2 \end{aligned}$$

Add together, cancelling terms (looking for matches of each type: R^3 , R^2X , RX^2 and X^3): -

$$|\mathbf{Z}| = R_1R_2R_3 + R_2R_3X_1 + R_1\left(R_2 + R_3\right)X_3 + R_3\left(R_1 + R_2\right)X_2 + \left(R_1 + R_2 + R_3\right)X_2X_3 + \left(R_2 + R_3\right)X_1X_3 + R_3X_1X_2 + X_1X_2X_3$$

Appendix 3: Calculating transfer functions

The following is repeated for convenience: $\mathbf{X} = \begin{pmatrix} R_1+R_2+R_3 & R_2+R_3 & R_3 \\ R_2+R_3 & R_2+R_3 & R_3 \\ R_3 & R_3 & R_3 \end{pmatrix}$ with $\mathbf{\Omega} = \mathbf{XY}$

The matrix elements of $\mathbf{\Omega}$ and $\mathbf{Y}$ required for transfer functions are: -

$$\Omega_{11} = R_{11}Y_{11} + R_{12}Y_{21} + R_{13}Y_{31} \qquad \Omega_{12} = R_{11}Y_{12} + R_{12}Y_{22} + R_{13}Y_{32} \qquad \Omega_{13} = R_{11}Y_{13} + R_{12}Y_{23} + R_{13}Y_{33}$$

From appendix 1: -

$Y_{11} = R_2R_3 + (R_2+R_3)X_3 + R_3X_2 + X_2X_3$ $\quad Y_{12} = -R_2R_3 - (R_2+R_3)X_3$ $\quad Y_{13} = -R_3X_2$

$Y_{21} = -R_2R_3 - (R_2+R_3)X_3$ $\quad Y_{22} = (R_1+R_2)R_3 + R_3X_1 + (R_1+R_2+R_3)X_3$ $\quad Y_{23} = -R_1R_3 - R_3X_1$

$Y_{31} = -R_3X_2$ $\quad Y_{32} = -R_1R_3 - R_3X_1$ $\quad Y_{33} = R_1(R_2+R_3) + (R_2+R_3)X_1 + (R_1+R_2+R_3)X_2 + X_1X_2$

With results: -

$$\begin{aligned}\Omega_{11} &= (R_1+R_2+R_3)(R_2R_3 + (R_2+R_3)X_3 + R_3X_2 + X_2X_3) + (R_2+R_3)(-R_2R_3 - (R_2+R_3)X_3) - R_3^2X_2 \\ &= (R_1+R_2+R_3)R_2R_3 + (R_1+R_2+R_3)(R_2+R_3)X_3 + (R_1+R_2+R_3)R_3X_2 + (R_1+R_2+R_3)X_2X_3 - (R_2+R_3)R_2R_3 - (R_2+R_3)^2X_3 - R_3^2X_2 \\ &= R_1R_2R_3 + R_1(R_2+R_3)X_3 + (R_1+R_2)R_3X_2 + (R_1+R_2+R_3)X_2X_3\end{aligned}$$

$$\begin{aligned}\Omega_{12} &= (R_1+R_2+R_3)(-R_2R_3 - (R_2+R_3)X_3) + (R_2+R_3)((R_1+R_2)R_3 + R_3X_1 + (R_1+R_2+R_3)X_3 + X_1X_3) + R_3(-R_1R_3 - R_3X_1) \\ &= -(R_1+R_2+R_3)R_2R_3 - (R_1+R_2+R_3)(R_2+R_3)X_3 + (R_2+R_3)(R_1+R_2)R_3 + (R_2+R_3)R_3X_1 + (R_2+R_3)(R_1+R_2+R_3)X_3 - R_1R_3^2 - R_3^2X_1 \\ &= R_2R_3X_1 + (R_2+R_3)X_1X_3\end{aligned}$$

$$\begin{aligned}\Omega_{13} &= (R_1+R_2+R_3)(-R_3X_2) + (R_2+R_3)(-R_1R_3 - R_3X_1) + R_3(R_1(R_2+R_3) + (R_2+R_3)X_1 + (R_1+R_2+R_3)X_2 + X_1X_2) \\ &= R_3X_1X_2\end{aligned}$$

Lovely!

Appendix 4: Slide rule calculator for RC combinations

Procedure: -

1. Calculate the required time constant $\tau = RC$ and write it in the form $\alpha \times 10^N$ where α is between 1.0 and 9.9.
2. Record the required value for N.
3. Calculate $X_0 = log_{10}\,\alpha$ (X_0 should be between 0.0 and 1.0).
4. Place the Y_1 marker at X_0 on the X scale and look for the two values of R and C that most closely match.

Note: Apart from factors of 10 $X_0 = \log_{10} RC = \log_{10} R + \log_{10} C = \log_{10} R - \log_{10} \frac{1}{C}$

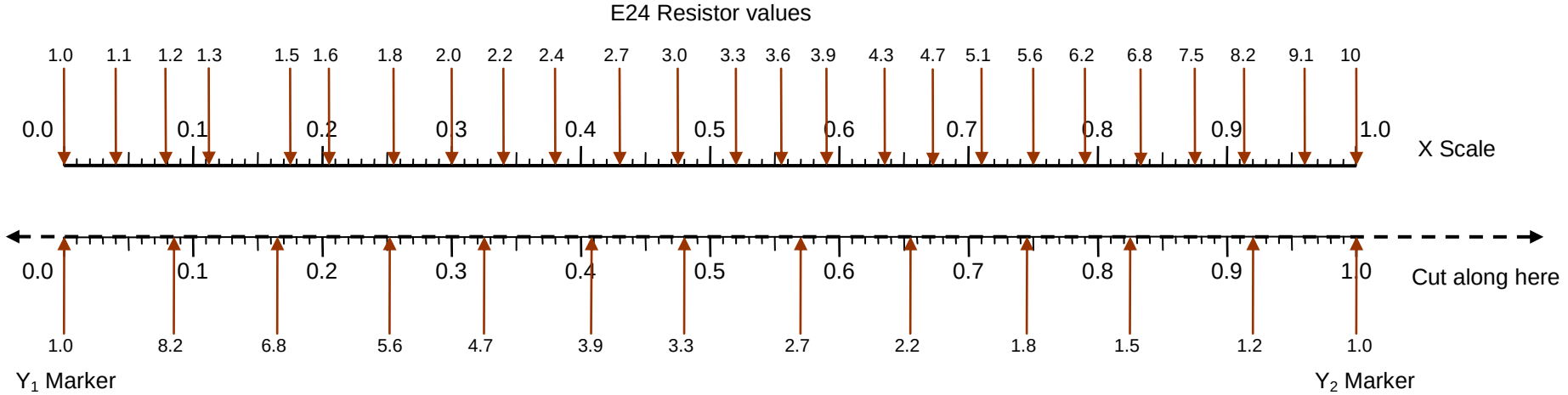
The slide rule performs the addition $X_0 + \log_{10} \frac{1}{C} = \log_{10} R$

5. Try the Y_2 marker at X_0 to see if there is a better match (equivalent to ten times smaller value)
6. Choose suitable values of A and B ($R \times 10^A$ and $C \times 10^B$) so that $RC \times 10^{A+B} = \tau$

Note: Depending on the application A is usually between 2 and 6 (i.e. R is between 100Ω and 1MΩ.) and B can range from minus 12 (pico farads) to minus 2 (large electrolytics).

Example: $\tau = 2.04\mu s = 2.04 \times 10^{-6}$s and so $X_0 = 0.31$
The best fit is R = 3.0 and C = 6.8 so practical values are $R = 3 \times 10^3$ and $C = 6.8 \times 10^{-10}$ i.e. 3kΩ and 680pF.

Note: The same method may be used for calculating inductor/capacitor combinations where $LC = 1/(2\pi f_R)^2$ and f_R is the resonant frequency.



An ultra-high input impedance high-pass filter

1. Introduction

It is shown elsewhere [1] that a three-stage high pass filter (type 2) has a high input impedance by virtue of the "bootstrap" effect – a form of positive feedback. This property is not normally required but it does raise the possibility of novel applications which may be of interest to the reader.

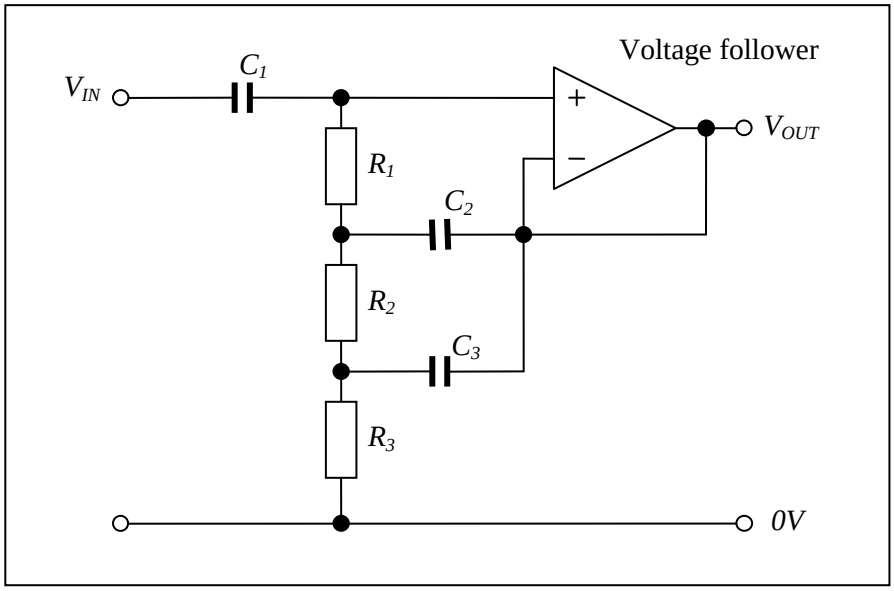


Fig. 1.1 A three-stage high-pass filter

Analysis shows that the input impedance is boosted by a factor proportional to the frequency cubed. In normalised form $\left(s = j\,\omega/\omega_N\right)$, at a frequency well above the natural frequency, the input admittance is [1]: -

$$\omega >> \omega_N \;\Rightarrow\; A_{IN} \approx \frac{1}{s^3}\left(\frac{1}{X_1}+\frac{1}{R_1}\right) \quad \text{with} \quad X_1 = 1/sC_1$$

This can be expressed as a parallel combination of a large negative resistor and small capacitor. In more convenient form, with frequency in Hz: -

$$A_{IN} = \frac{1}{Z_{IN}} = \frac{1}{Z_C}+\frac{1}{Z_R} \qquad \text{With} \quad Z_C = -\left(\frac{f}{f_N}\right)^2 \frac{1}{2\pi f_N C_1} \text{ and } \qquad Z_R = -j\left(\frac{f}{f_N}\right)^3 R_1$$

$$\text{and} \qquad f_N = \frac{1}{2\pi\left(R_1C_1R_2C_2R_3C_3\right)^{\frac{1}{3}}}$$

With readily available components it is possible, in theory, to achieve an input impedance which is much higher than conventional JFET or MOSFET devices. The input impedance of the follower then sets the upper limit. If, for example [1] $C_1 = 33pF$, $R_1 = 5G\Omega$ and $f_N \approx 1Hz$.

At 75Hz: $\quad Z_C \approx -27\times10^{12}\Omega$ and $Z_R \approx -2\,j\times10^{15}\Omega$

1. Part 2, monograph 2: "Three-stage filters". See section 3.

2. Outline design proposal

One possibility is to employ an "inside-out" voltage follower [1], based on a two or three-stage high gain block (HGB) [2], with a very low leakage input stage (MOSFET). Inside-out operation not only ensures very high accuracy (no common mode at the inputs) but also bootstraps input capacitance and resistance - the floating power supply and, therefore, the input stage of the HGB is driven to follow the AC input voltage.

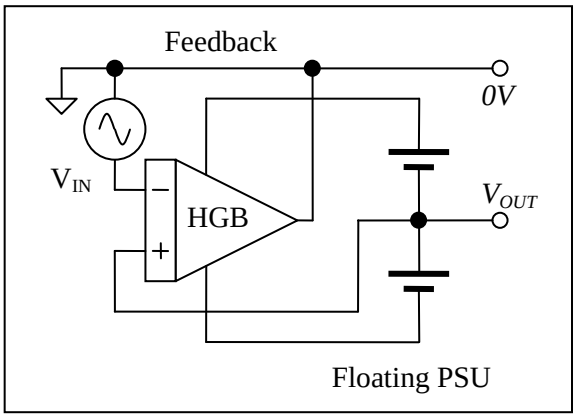


Fig. 2.1 The inside-out voltage follower

The output of the follower is fed back, via suitable filters, to bootstrap a large resistor and stray/screen capacitance. The upper limit of source resistance is determined by the need for stability: -

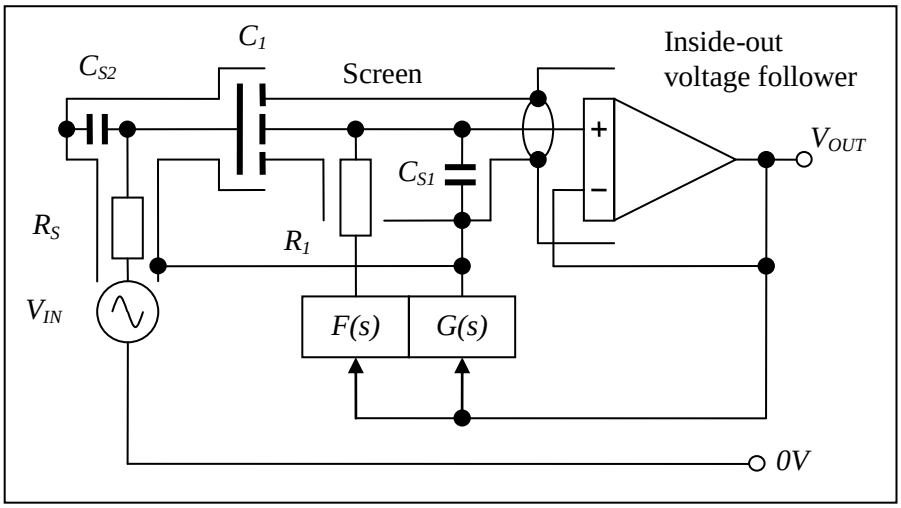


Fig. 2.2 Outline schematic

The voltage follower is drawn in the conventional way to aid interpretation of the diagram. The screen (active guard) must completely surround the high impedance nodes (usually both sides of C_1, most of R_1 and the first stage of the voltage follower).

The high-pass filter, *F(s)*, and low-pass filter, *G(s)*, ensure stability. Each feedback route has a band-pass characteristic so that the feedback factor is lower than 100% when the phase shift is zero.

The input capacitor could be a small parallel plate capacitor (air dielectric and typically ≈ 10pF) [3], constructed with high resistivity insulators (PTFE or, better still, sapphire) [4].

1. Part 4, monograph 2: "High accuracy voltage followers".
2. Part 4, monograph 1: "High gain blocks".
3. Part 1, monograph 5: "Variable gap capacitive displacement transducers".
4. Yeager, John. & Hrusch-Tupta, Mary Anne et al: "Low Level Measurement", 5th edition, available (free) from Keithley Instruments Inc. For more practical tips on very high resistance and very low current measurements see sections 2 and 4.

The very high input impedance $(>1T\Omega)$ and low leakage $(\approx 10\,fA)$ input stage permits the use of a higher value resistor $(R_1 \approx 100G\Omega)$ reducing the natural frequency (typically ≈ 0.16Hz or $\omega_N \approx 1$).

The large resistor provides a route for the DC leakage current but, at the operating frequency (typically 25Hz or 75Hz), the potential difference across it is very small and the current through it is substantially reduced. The stray capacitance, from the high impedance node to the surrounding screen, is similarly "bootstrapped".

For example: at the operating frequency the follower and filters can be accurate to better than 1ppm (e.g. with a three-stage HGB and high accuracy filters). The potential difference across the resistance and stray/screen capacitance is < 10^{-6} of the input voltage and their impedances are boosted by a factor of > 10^6. The impedance of 6pF stray capacitance, for example, has an impedance of 1GΩ at 25Hz. This would be boosted to over 1PΩ ($10^{15}\Omega$)!

The filters, *F(s)* and *G(s)*, are probably best made adjustable so that loop stability can be optimised for the particular application (depending on the source resistance). Both can be two-stage filters for low phase error at the operating frequency [1] or three-stage for low phase and low magnitude error [2].

The main problem is likely to be the gate-drain and gate-source capacitance of the HGB input stage. The positive feedback through that route probably sets the upper limit on the impedance of source resistance, R_S, and input capacitor, C_1. The equivalent feedback network is quite complicated: -

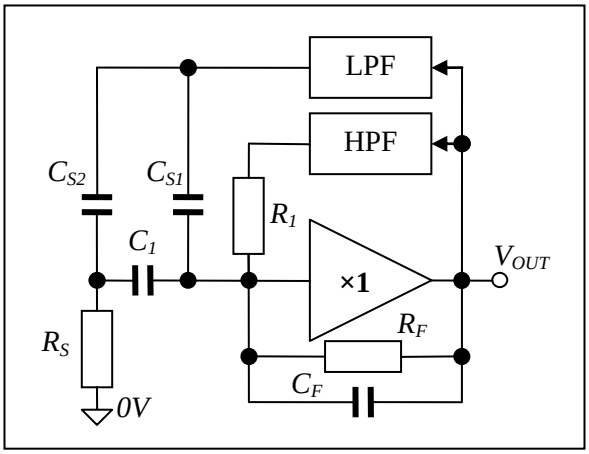


Fig. 2.3 The feedback model at the operating frequency

The solution could be a separate power supply for the first stage, bootstrapped via a suitable filter. The output of the first stage and the input of the second stage of the HGB would need to be differential: -

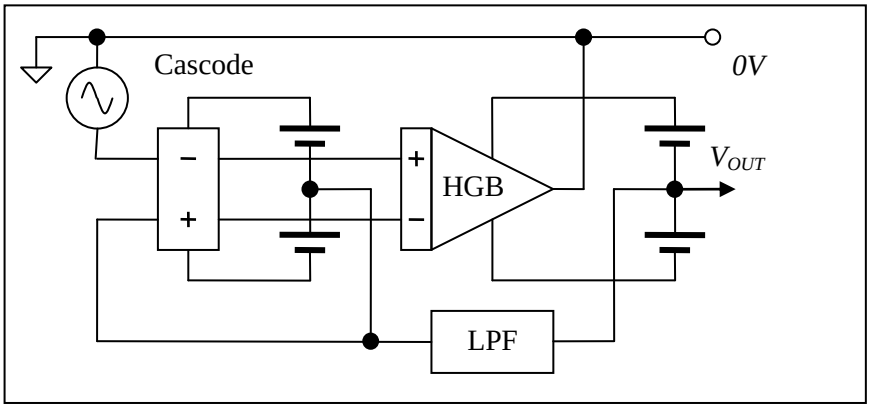


Fig. 2.4 A double bootstrapped inside-out follower

1. Part 2, monograph 1: "Two-stage filters"
2. Part 2, monograph 2: "Three-stage filters"

3. A low cost version

The high value resistor is a costly item. Also, MOSFETs tend to be noisy, compared to their JFET cousins. A possible alternative is an ultra-low leakage pre-amplifier based on a pair of low cost JFETs operated in voltage controlled resistor (VCR) mode [1]. The following ideas, dear reader, are untested – caveat emptor: -

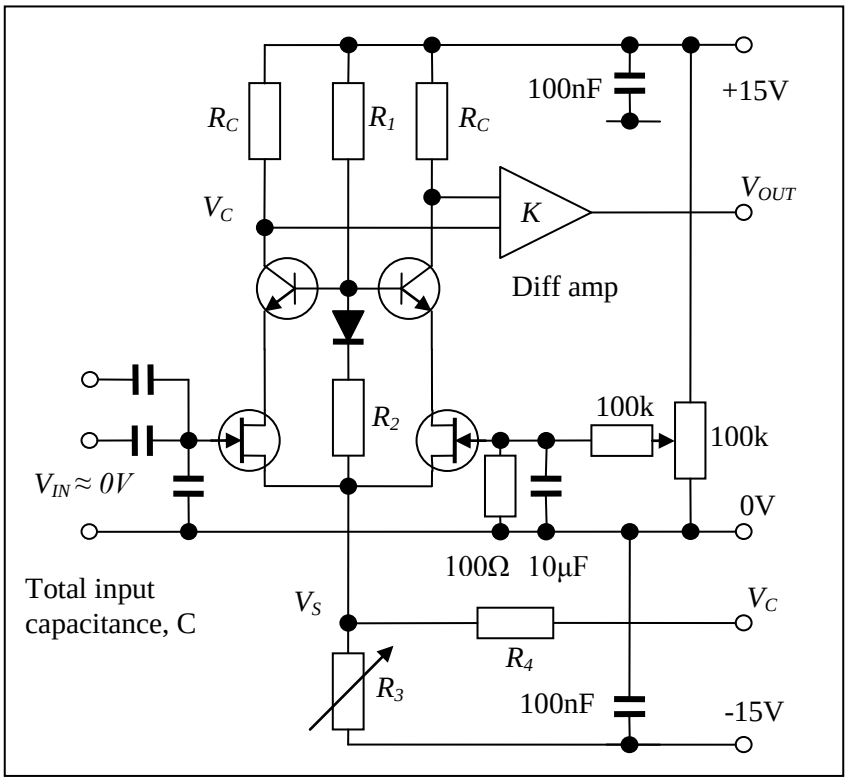


Fig.3.1 An ultra-low leakage pre-amplifier

The basic idea is that the gate-drain and gate-source potential differences are small and (approximately) equal and opposite, relative to the gate (at 0V). Small adjustments to the drain and source DC levels are made via the control signal, V_C. A slowly acting feedback loop operates to maintain the input gate at 0V. This is probably best implemented with a microcontroller and adaptive algorithm [1 see section 4]. The average gate-drain and gate-source leakage currents are not only equal and opposite but also inherently very small (typically $I_L \approx 50\,fA$ at 20ºC). The noise current is, therefore, also very small [2]: -

$$I_N = \sqrt{2eI_L B} \quad \text{and} \quad I_L \approx 50\,fA \;\Rightarrow\; I_N \approx 0.13\,fA/\sqrt{Hz}$$

With discrete JFETs it would also be possible to incorporate the input JFET in the capacitor: -

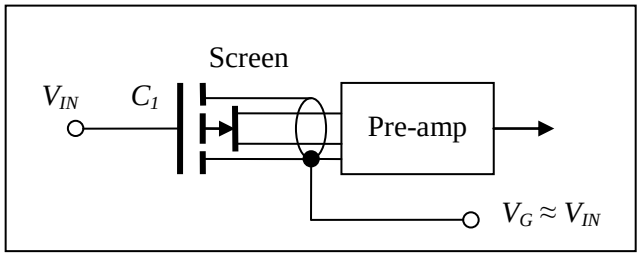


Fig. 3.2 A JFET in the capacitor

1. Part 5, monograph 3: "Low noise JFET pre-amps".
2. Some proprietary low leakage devices employ extra circuitry to match the leakage current, including its temperature dependent characteristic. The net average leakage current is then very small but the noise current remains high.

The leakage control loop results in a second order high-pass characteristic, which is not ideal for operation within an overall feedback loop, such as the inside-out voltage follower. It should be possible, however, to disable the control loop for a few seconds while a measurement is made. The DC level at the pre-amp output will change slowly but not sufficiently to affect the signal at the operating frequency.

During auto-zero mode the power supply centres are connected to local 0V. With a notch filter in the leakage control loop it may not be necessary to disconnect the input signal.

The high-pass filter is no longer required but the second floating power supply and a band-pass filter for the transistor leakage resistance and stray capacitance is probably necessary: -

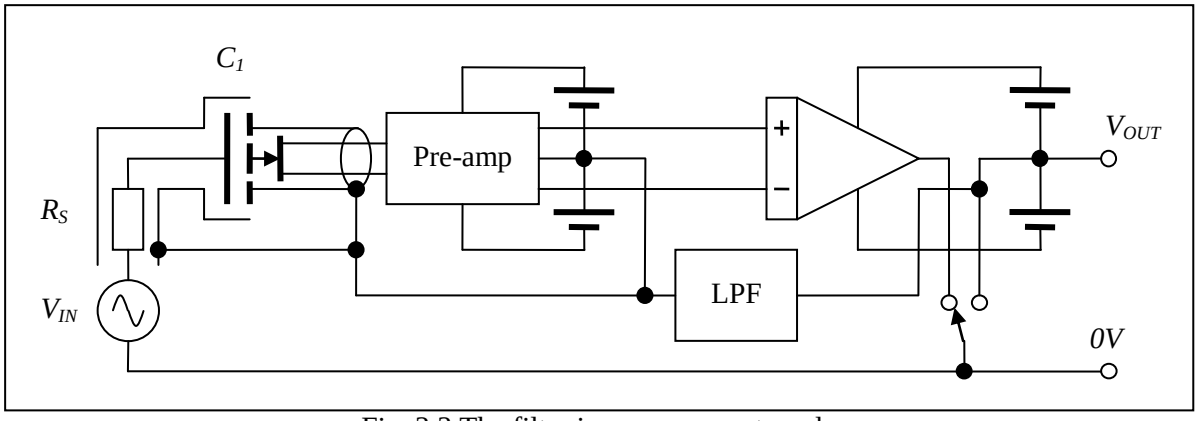


Fig. 3.3 The filter in measurement mode

The circuit may be used as a low frequency AC high pass filter with sub-ppm accuracy even if the source resistance is quite high source resistance (typically up to 100MΩ). It could also be used as the low noise front end of a null detector in a bridge for comparing very high value resistors.

Another possible application is multi-phase electromagnetic flow tomography – measuring small AC signals through a dielectric medium with an array of capacitive sensors.

If you are interested in building and testing a prototype please contact the author.

Inductive voltage dividers and ratio transformers – the basics

1. Introduction

Inductive voltage dividers (IVDs) and Ratio transformers (RTs) are very accurate electronic multipliers. With a voltage transformer, for example, the output voltage is the input voltage, usually a sine wave, multiplied by a number, set by the turns ratio: –

$$V_{OUT} = V_{IN} \times n \qquad \text{Where } n \text{ is a number.}$$

A number of stages can be combined to provide 6 or more decades so that the ratio may be set, for example, from n = 0.0000000 to n = 0.9999999. Clearly, the accuracy of the first decade is most critical.

When used in a null balance "bridge" configuration with a very sensitive (low noise) null detector, the ratio of two AC voltages or currents can be measured with an accuracy better than one part per million (1ppm or 0.0001%). With care an accuracy approaching one part per billion (1ppb) is possible. The main disadvantages, compared to resistive dividers are size and cost.

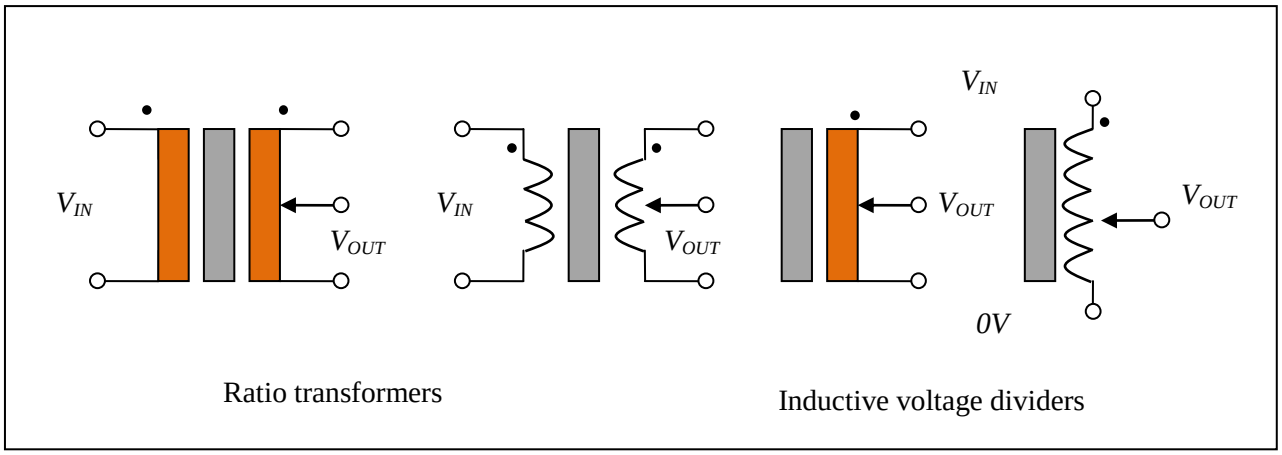


Fig. 1.1 Frequently used circuit symbols (the dot indicates polarity)

High accuracy is made possible by the ready availability of soft magnetic materials (principally metallic glasses with trade names such as "permalloy", "mumetal" and "super-mumetal") with extraordinarily high and sustainable relative permeability (50,000 to 500,000). High permeability means that magnetic flux "flows" easily through the material, compared to air, plastic or other non-magnetic materials. A number of companies supply toroidal cores based on these alloys, in a variety of sizes and grades, which can be used to construct IVDs and RTs. The material is manufactured in thin sheets and then wound, like a tight clock spring, into the shape of a toroid, with rectangular cross-section, and encased in plastic.

Fig. 1.2 Toroidal cores are available in a variety of sizes

1.1 Applications

The main applications are in the areas of electrical primary and transfer standards, for measuring the ratio of resistors and capacitors, platinum resistance thermometry and a wide variety of research activities, where pushing the boundaries of accuracy or resolution is a key objective.

Fig. 1.1.1 The F900 resistance thermometry bridge (picture courtesy ASL Ltd)

Most commercially available IVDs are manually operated and are used mainly to construct ad-hoc bridge configurations for research applications. At least two IVD sets are usually required, taking up a lot of bench space. Whereas the first automatically balancing bridge was based on a pair of IVDs (ASL's model A7 double Kelvin thermometry bridge [1]) this has been superseded by RT based designs such as ASL's F900.

Fig. 1.1.2 A 7-decade IVD (picture courtesy Tinsley)

A potential area of application is in ultra-precision engineering and nano-technology. Variable capacitance displacement transducers, for example, have remarkable accuracy and resolution, which can be realised only with a ratio transformer based capacitance bridge.

1. Hill J. J. and Miller A.P.: "An AC double Bridge with inductively coupled ratio arms for precision platinum resistance thermometry." Proc. I.E.E. 1963 110 No. 2 pp 453 - 458

1.2 Leakage flux

The high permeability and toroidal form ensure that a very high proportion of the magnetic flux is retained within the body of the toroid. Some of the flux "leaks out" of the toroid, however, depending on the distribution of currents flowing around the core. Some flux also leaks out at the ends of the high permeability metal strip (at one point each on the OD and ID). The effect of the leakage flux on accuracy can be minimised by using a "rope" and a symmetrical, evenly distributed, winding scheme. In practice an accuracy of 1ppm is easily achieved. Magnetic screening is possible but there are more practicable ways of reducing errors to the ppb range.

1.3 Winding schemes

1.3.1 The "no-net-loop" (NNL) method

The method most widely used, for all types of winding, results in no net loop around the core. The winding is started at point A and progresses around the core to complete the circle. The direction is then reversed - back to A (see fig. 1.3.1). The art of IVD and RT design and construction is to choose the best combination of core size and wire thickness that produces a uniform distribution. A good choice, for example, just fills the outer circumference with a single layer and the inner circumference with a double layer with the required number of turns.

The NNL method ensures that any current carrying winding does not produce an external magnetic field (passing through the centre of the toroid). This is a common problem with machine wound toroidal power transformers.

Similarly, any flux passing through the centre of the toroid does not result in an induced voltage in the winding.

Also, one or more net loops adds undesirable series inductance.

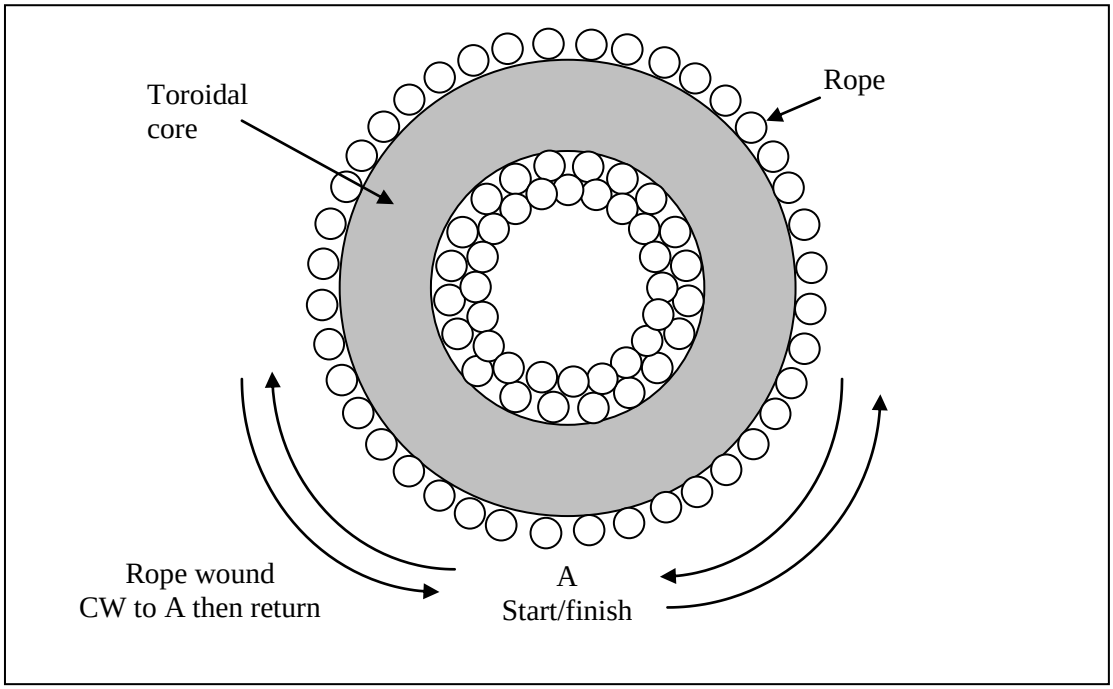


Fig. 1.3.1 The "no-net loop" winding technique

The other key advantage of the NNL method is that the effects of interwinding capacitance can be minimised. An energising winding, for example, can be driven by a balanced voltage source. Each part of the winding is in close proximity to another part which is at equal but opposite voltage. Capacitive currents to a neighbouring winding substantially cancel.

1.3.2 The "balanced no-net-loop" (BNNL) method

A practicable method for winding a large number of turns is to first wind the rope or wire on a shuttle: a wooden batten with slots at each end. The batten and rope needs to be sufficiently thin to pass through the centre of the core. This is not always possible for the whole winding and so a balanced no-net-loop (BNNL) method is employed. Two shuttles are then required with half the rope on each.

Shuttle 1: half is wound in a clockwise direction - from A to B and back again. The other half of the rope (shuttle 2) is then wound in a counter-clockwise (CCW) direction from A to B and back again (see fig. 1.3.2).

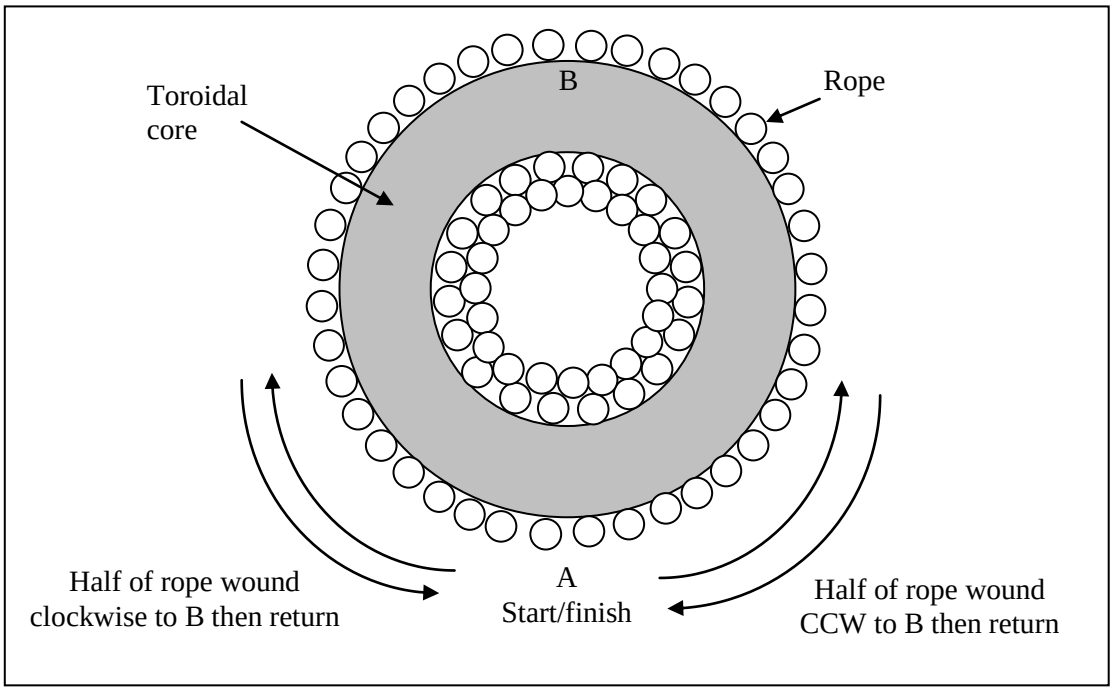


Fig. 1.3.2 The balanced zero loop winding technique

1.3.3 The NNL and BNNL methods applied to low number of turns of energising windings.

The same principles can be applied to energising windings with a low number of net turns. A number of windings are applied, each covering a proportion of the circumference so that the whole core is covered uniformly. The windings are then connected in parallel.

A practicable example is six windings of 40 turns each, each occupying one sixth of the core circumference with fairly thin wire. The result approximates to a uniform sheet of energising current around the toroid and very low leakage flux. A typical application would be a step-up transformer for a capacitance bridge. The high operating frequency means that interwinding capacitance can be an issue and so the NNL method is employed.

2. The basic inductive voltage divider

The most basic design is an IVD with a single rope winding on a single toroidal core. A typical example has a 10 strand rope of 40 turns (a total of 400 turns) wound as in fig. 2.1.

The basic principle is that the self inductance and resistance of each strand of the rope is very nearly equal. The former is achieved by selecting a suitably high permeability core and the latter by using wire with a uniformly constant diameter, usually taken from the same reel. At a sufficiently high frequency the input impedance is high, the current low and the resistive voltage drop small, compared to the induced voltage. The resistive voltage drop errors are in quadrature (at 90 degrees relative phase) and would be largely rejected by a suitable phase sensitive null detector and/or separate quadrature balance.

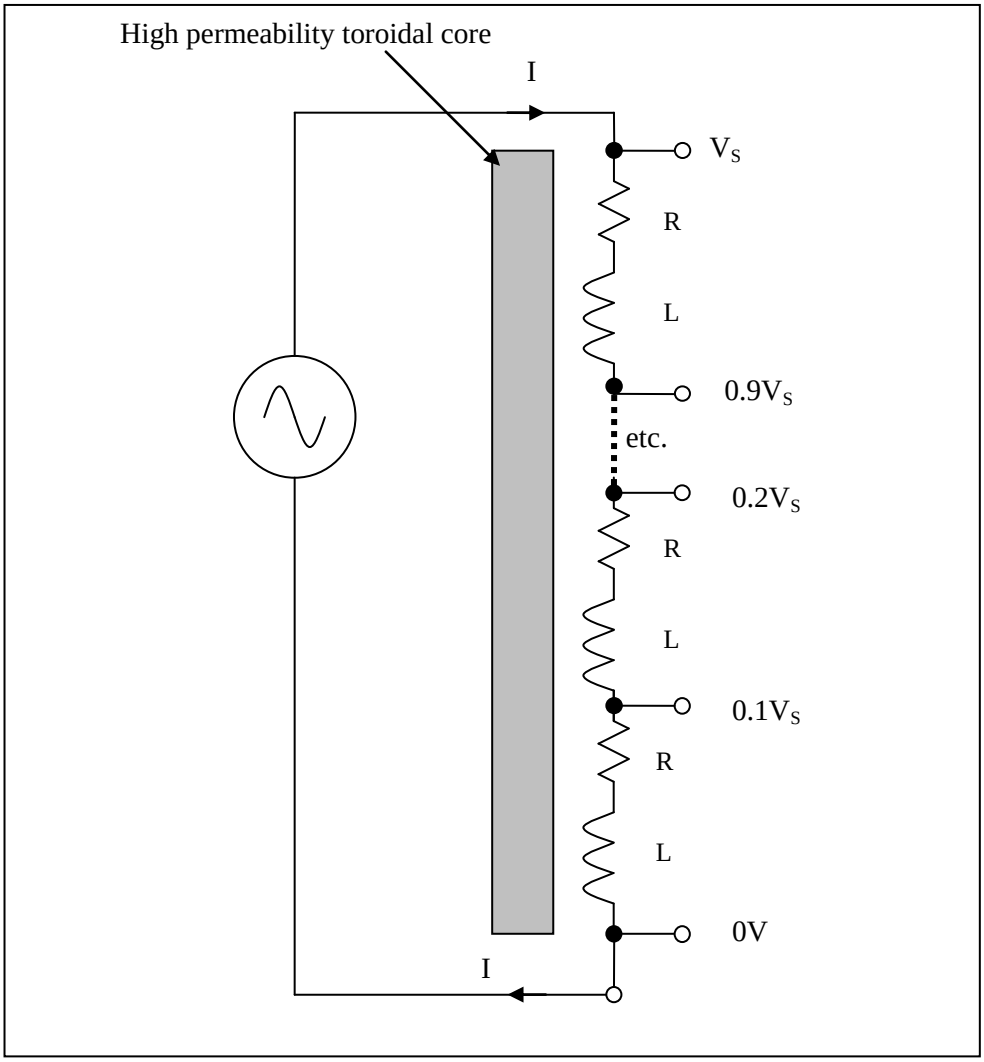


Fig. 2.1 The basic single-stage IVD

The result is a very simple device, easily manufactured, with a ratio accuracy of better than 1ppm over a frequency range of 50Hz to 10 kHz. At low frequency the main problem is low input impedance and at higher frequency the main problem is the interwinding capacitance.

The characteristics of IVDs have been thoroughly investigated with very good correlation between actual performance and theoretical modelling. See, for example [1 and 2].

1. Hill, J. J. and Deacon, T.A. "Theory, design and measurement of inductive voltage dividers". Proceedings of the IEE, vol. 115, pp. 727-735 (May 1968).
2. Hill, J. J. and Deacon, T.A. "Voltage ratio measurements with a precision of parts in 10^9 and performance of inductive voltage dividers". IEEE Trans. Instrum. Meas., vol. IM-17, pp269-278, Dec 1968.

2.1 Multi-decade IVDs

Commercially manufactured IVDs are available with five or more decades of switching so that ratio settings ranging from, for example, 0.000000 to 0.999999 are possible.

The basic idea is that the second decade IVD is energised by tapping one tenth of the voltage from the first decade and so on for subsequent decades [1]. This is achieved with a two-pole ten-way switching mechanism, using good quality, low contact resistance, manual rotary switches or electro-mechanical relays. The loading effect of the second and subsequent decades depends on the source impedance and only those bridge configurations which present low source impedance to the IVD input are practicable. This is one of the main limitations of the IVD. In one of the main applications (measurement of resistance ratio) the result is a large and expensive bridge [2] with two complete sets of IVDs. The relatively low input impedance of each IVD stage also requires quite low switch or relay contact resistance - big, chunky switch contacts or mercury whetted reed relays and thick wire interconnections are necessary.

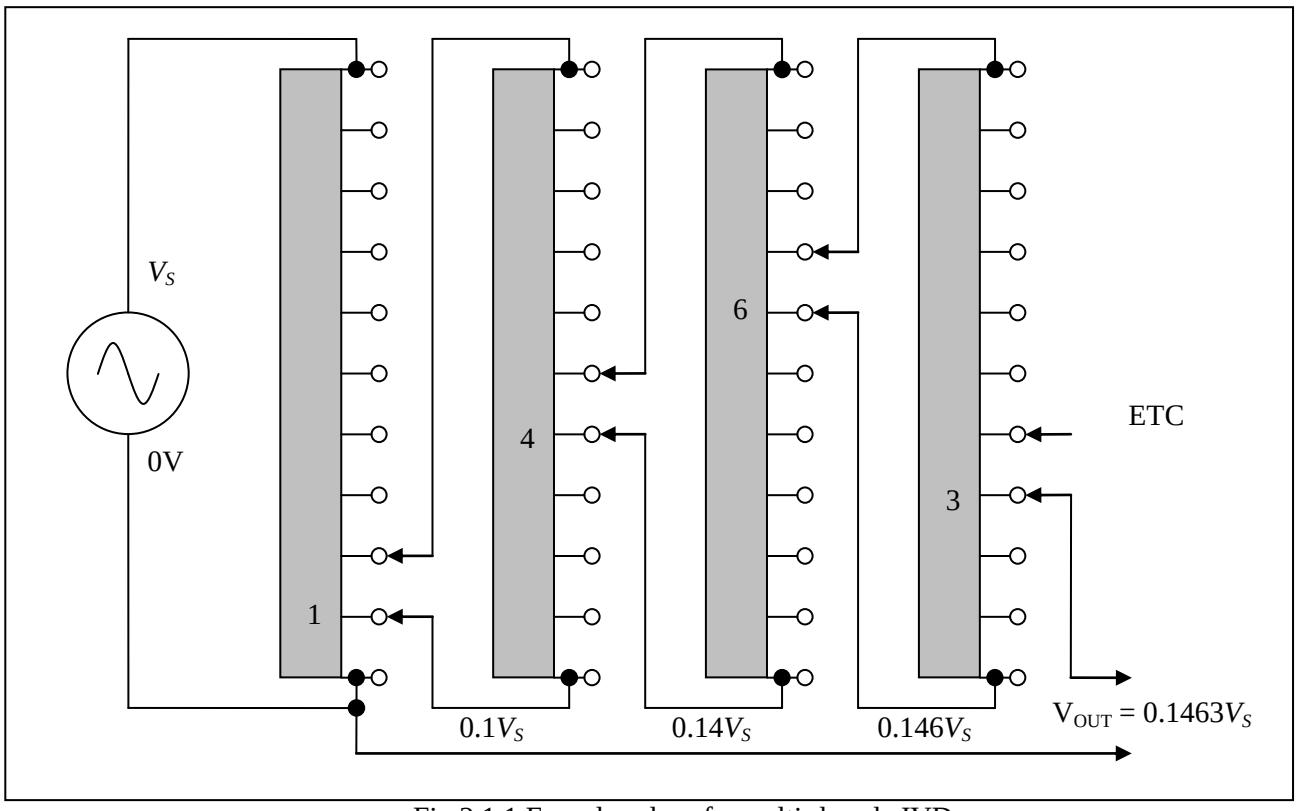


Fig.2.1.1 Four decades of a multi-decade IVD

Despite the practical problems the basic IVD can achieve ppb accuracy and commercial multi-decade IVD are still used for ad-hoc bridge configurations to this day.

1. Often referred to in the literature as the “Kelvin–Varley method” by analogy with the original kelvin-Varley resistive divider (ref. required: see Wikipedia).
2. The market leader for many years was the model A7 from Automatic Systems Laboratories Ltd, based on the original design by Hill and Miller.

Fig. 2.1.2 The A7 employing two seven decade dividers (Picture courtesy ASL Ltd). (N.B. weight about 70kg!)

2.2 Two-stage IVDs

The main disadvantage with the basic single stage IVD is the low input impedance at low frequency. The ratio winding also carries the current to provide the flux in the core – acting as both energising and ratio winding. To achieve high accuracy it is often necessary to use a large, expensive, high grade core (very high permeability e.g. super-mumetal SM200), at least in the first decade, to achieve sufficiently high inductance. There is, however, a much simpler solution - a two-stage design, where two toroidal cores and a separate energising winding are used. The result is a much reduced current flowing in the ratio winding and connecting leads.

Historical note: This technique was first used by Brook and Holtz [1], early in the 20th century, to improve the accuracy of current measuring transformers.

The two-stage IVD is best depicted, diagrammatically, by using the following convention: -
1. Cores and windings are depicted as vertical rectangles. It is fairly clear, even in monochrome, which is which – the windings have connections. I shall adopt the convention of orange shading for windings and grey for cores.
2. A winding that is horizontally adjacent to one core is wound only around that core.
3. A winding that is horizontally adjacent to two or more cores is wound around all cores.

It will be seen later that this convention can be extended to much more complicated arrangements of windings and cores.

In fig 2.2.1, for example, the energising winding is wound around one core only – the "bottom" core. The second core is then placed on top and the ratio winding is wound around both cores. Both windings comply with the BNNL winding scheme with exactly the same number of turns. If, for example, the ratio winding has ten strands and 40 turns, the energising winding must have exactly 400 turns.

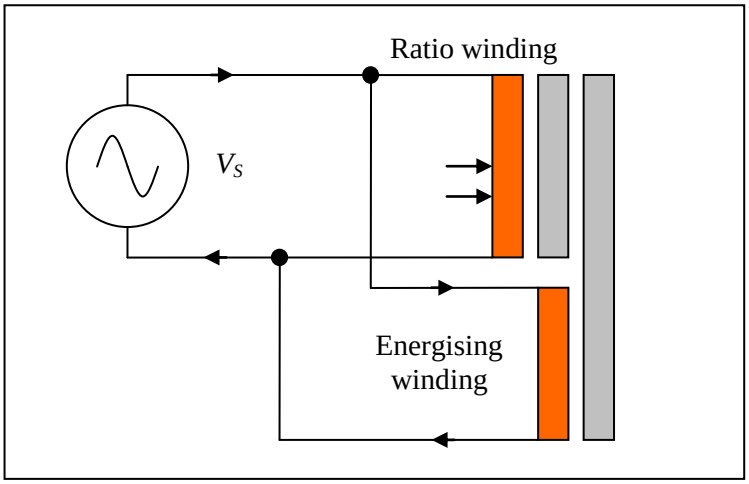


Fig. 2.2.1 The "two-stage" technique with separate energising winding

The reduction in current flowing in the ratio winding makes it possible to use much thinner wire for the rope – making construction much simpler. Thinner wire can also be used for interconnections and the resistance of connectors, switches and relays is far less of a problem. The benefits to the designer are many. As a rule of thumb the current in the ratio winding can be reduced by at least a factor of a hundred and up to one thousand over the low audio frequency range (20 – 100Hz).

1. Brooks, H. B., and Holtz F. C., "The two stage current transformer". AIEE Trans. Vol. 41, 1922, p382

The principle of operation is very simple – the current flowing through the energising winding sets up very nearly the flux corresponding to the applied voltage. The difference is due to the resistance of the energising winding. The ratio winding shares this flux and so a voltage is induced which is very nearly equal to the applied voltage. The difference gives rise to very small current flowing in the ratio winding, setting up the remainder of the flux in the top core. The ratio winding is thus "bootstrapped", presenting much higher input impedance to the voltage source. See the monograph "Two-stage IVDs and RTs" by the same author for more detailed analysis and example calculations [1].

2.3 The two-stage IVD with high accuracy voltage followers

In the basic two-stage design the energising winding still presents low impedance to the voltage source. Any current flowing through the connecting wires may cause a significant voltage drop, complicating the design of bridge configurations. The effect can be eliminated completely by employing low noise, high accuracy voltage followers with very high input impedance. The very high loop gain also ensures that the voltage followers have very low output impedance - at the point where the feedback voltage is sensed. For more detail see the monograph "High Accuracy Voltage Followers" by the same author [2]. The main disadvantage is the extra noise due to the followers, which is significantly greater than the Johnson noise generated in the winding resistance. It is, however, of the same order as the noise generated elsewhere (e.g. 100Ω source resistance: $\approx 1-2\,nV/\sqrt{Hz}$) and sub-ppm accuracy and resolution is still practicable in a modest bandwidth $(\approx 0.1-1Hz)$.

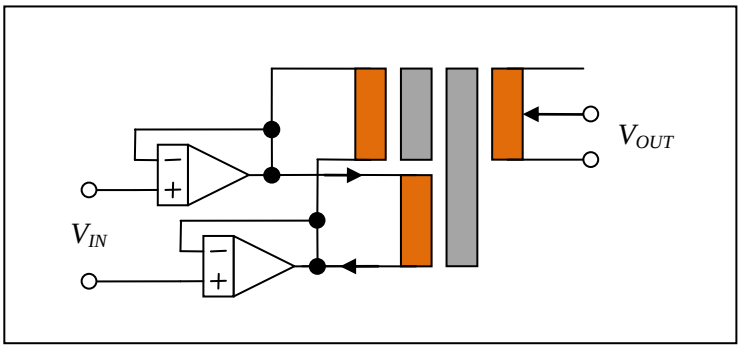


Fig. 2.3.1The two-stage technique with voltage followers

Note the order of the negative feedback connection – the larger current flowing through the energising winding should not flow through any part of the connection to the ratio winding.

1. Part 3, monograph 3: "Two-stage inductive voltage dividers and ratio transformers".
2. Part 4, monograph 2: "High accuracy voltage followers".

3. The basic ratio transformer

The most basic ratio transformer has a primary winding plus a fixed and a variable secondary winding. The fixed and variable windings are part of the same rope and so high accuracy of ratio is achieved. The primary winding is used only to energise the core and does not usually require a high degree of ratio accuracy relative to the ratio windings. The primary is usually, therefore, a single strand, in a single layer, with the number of turns to provide the required secondary voltage. A high step up ratio can be achieved by having multiple primary windings, each with a low number of turns, connected in parallel (see section 1.3.3). The primary winding must, however, be uniformly distributed around the core, each complying with the NNL or BNNL scheme, to provide, as near as possible, a uniform sheet of current around the core and minimum capacitive coupling to the secondary. The ratio windings are then wound over the top. Capacitance between primary and secondary windings can be further reduced with a thick layer of insulating tape (e.g. PTFE plumber's tape or polyester ribbon). With a single core it is easily possible to provide 2 or even 3 decades of a multi-decade bridge.

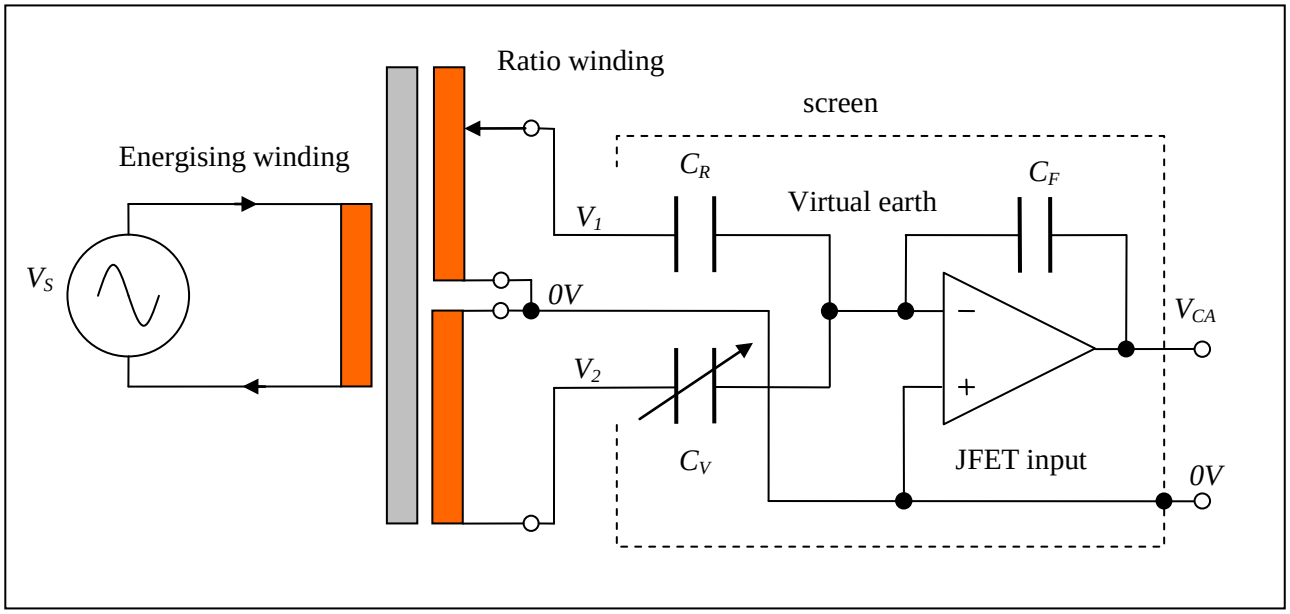


Fig. 3.1 The most basic ratio transformer in a capacitance bridge

$$n = -\frac{V_1}{V_2}$$ With n variable (e.g. 0.000000 to 0.999999)

At null balance: $$V_{CA} = \frac{V_1 C_R + V_2 C_V}{C_F} = 0 \qquad \Rightarrow \frac{C_V}{C_R} = -\frac{V_1}{V_2} = n$$

The basic ratio transformer is useful for measuring the ratio of high impedances (e.g. high value resistors or low value capacitors). Capacitive loading of the ratio windings (e.g. due to long cables) is not usually a problem but resistive loading can be, reducing accuracy.

Fig. 3.2 A commercial capacitance bridge (picture courtesy Andeen Hagerling)

A multi-decade transformer can be implemented by using an extra strand of a rope to energise subsequent cores. A six decade capacitance bridge operating at moderately high frequency (1.6kHz) would, for example, look something like the following: -

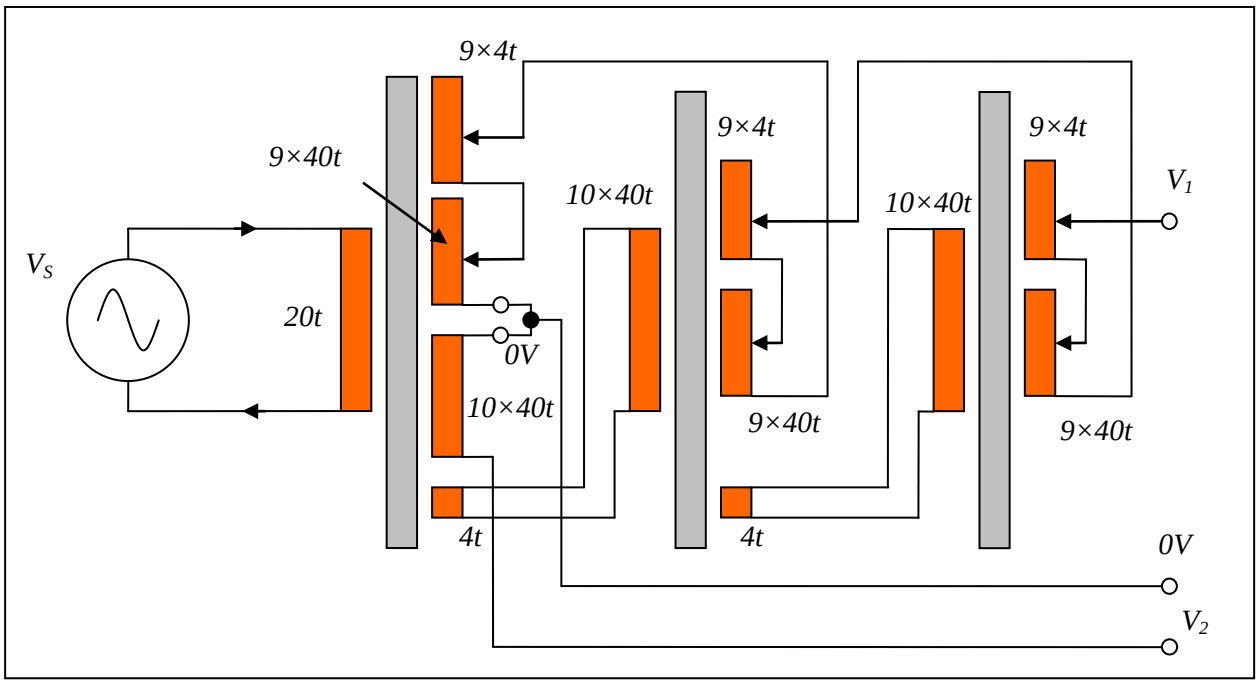


Fig. 3.3 A six decade bridge (0.000000 to 0.999999)

Surprisingly accurate results (sub-ppm) can be obtained with a simple bridge (e.g. measuring the ratio of three terminal capacitors) as long as the currents flowing in the ratio windings are kept low. If long connecting leads are required the capacitance of the cables becomes an issue. Here again, however, the effect is mainly quadrature and can be reduced significantly with a separate quadrature balance. For the highest possible accuracy tri-axial leads can be used, with the inner screen driven to be at the same voltage as the inner conductor.

A useful variation is when used with a transducer which is inherently differential. In this case the selected winding is connected to ground: -

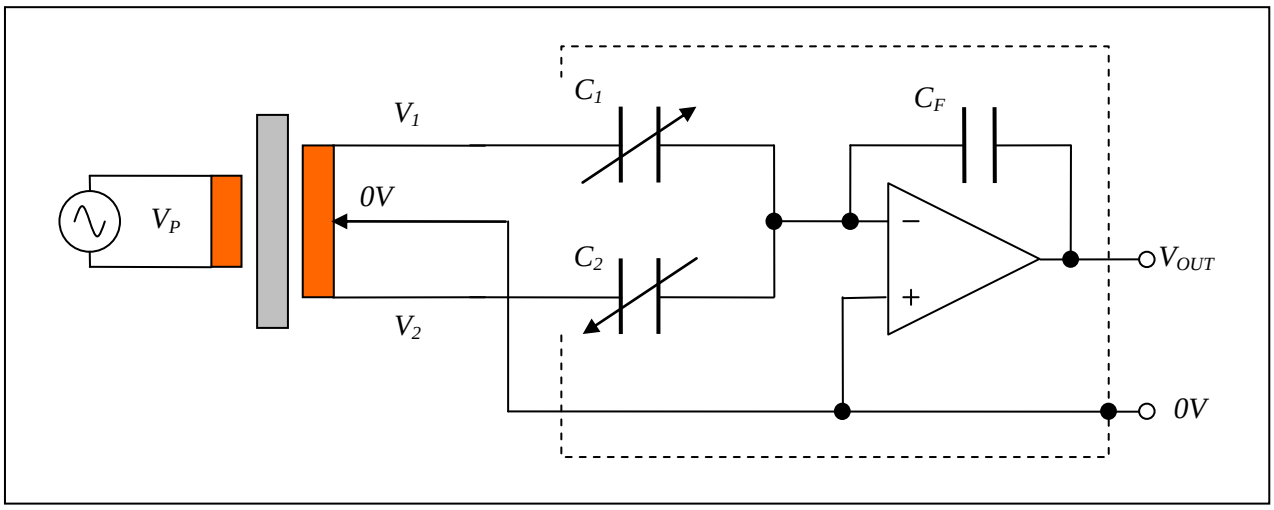


Fig. 3.4 Differential capacitance bridge

Define a reference voltage $\left(V_R = \frac{V_1 - V_2}{2}\right)$ and the ratio, *n*, such that: -

$$V_1 = V_R(1-n) \quad \text{and} \quad V_2 = -V_R(1+n)$$

N.B. When the transformer tap is at the centre ($n = 0$): $V_1 = V_R$ and $V_2 = -V_R$

As n increases the tap moves towards the V_1 end and AC voltage V_1 reduces while V_2 becomes increasingly negative.

At balance:
$$V_{OUT} = \frac{V_1 C_1 + V_2 C_2}{C_F} = 0$$

$$\Rightarrow \quad V_R(1-n)C_1 - V_R(1+n)C_2 = 0 \quad \Rightarrow \quad n = \frac{C_1 - C_2}{C_1 + C_2}$$

An alternative method has practical benefits. The centre of the centre-tapped ratio transformer is driven with a voltage, V_C, which is derived from the reference voltage $V_C = -nV_R$ using an IVD or RT so that: -

$$V_1 = V_R + V_C = V_R(1-n) \qquad V_2 = -V_R + V_C = -V_R(1+n)$$

As above, at null balance:
$$n = \frac{C_1 - C_2}{C_1 + C_2}$$

This method has found application with a dual differential rotary capacitive transducer based servo system with sub arc second accuracy and excellent dynamic response. For more detail see the monograph "A 16 bit binary differential capacitance bridge" by the same author.

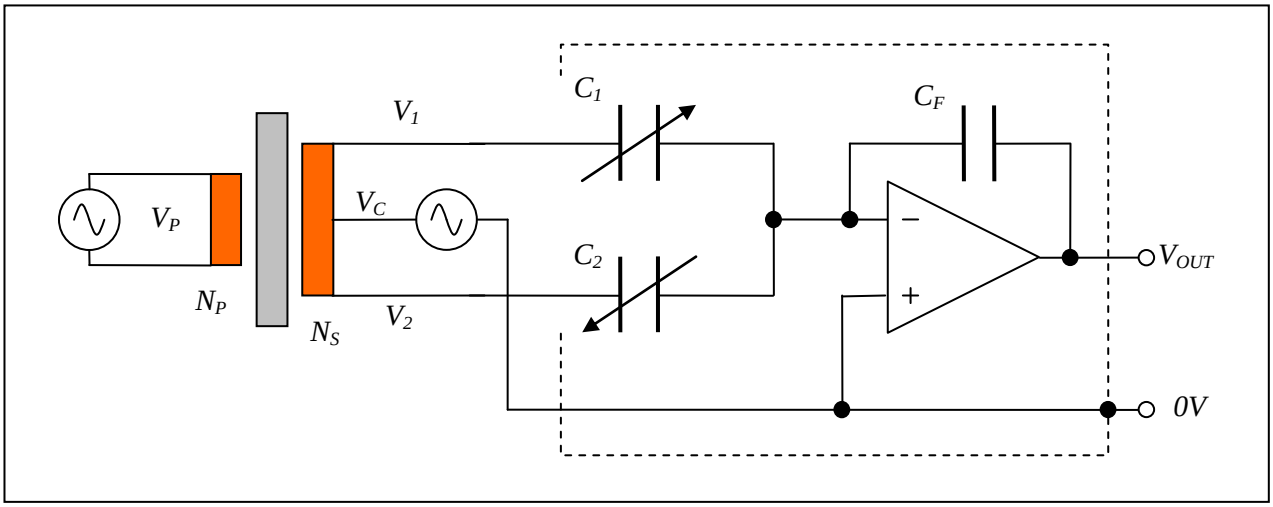


Fig. 3.5 An alternative differential capacitance bridge

For more detail see the monograph "Single stage IVDs and RTs" by the same author [2].

1. Part 3, monograph 8: "A 16 bit binary differential capacitance bridge".
2. Part 3, monograph 2: "Single stage inductive voltage dividers and ratio transformers".

4. The three-stage ratio transformer (e.g. ASL model F17)

The principle of a separate energising core can be taken further with a second energising stage. The bottom core provides most of the flux and the second energising core makes up for most of the rest. The result is a massive increase in input impedance of the ratio winding and much higher accuracy when used as a transformer. The current flowing through the ratio winding is now so low that much more resistance can be tolerated and thinner wire may be used. This allows for either a longer, thinner rope (more turns) and/or a rope with many more strands. Some of the strands are connected in series to provide a primary winding while the remainder provide a tapped secondary and, usually, an extra strand for equalisation of lower decades.

The main advantage, compared to the most basic ratio transformer, is that the voltage ratio between primary and secondary is now very accurate. If the required operating voltage is low then smaller cores can be used and the design can be quite compact.

With the addition of voltage followers the result is an elegant design, which is easy to construct, with ratio accuracy better than 1ppm, even at low frequency (75 or 90Hz). The isolation between primary and secondary windings also provide for a simple, elegant bridge configuration, suitable for measurement of resistance ratio, with long connecting cables (e.g. for calibrating resistance thermometers).

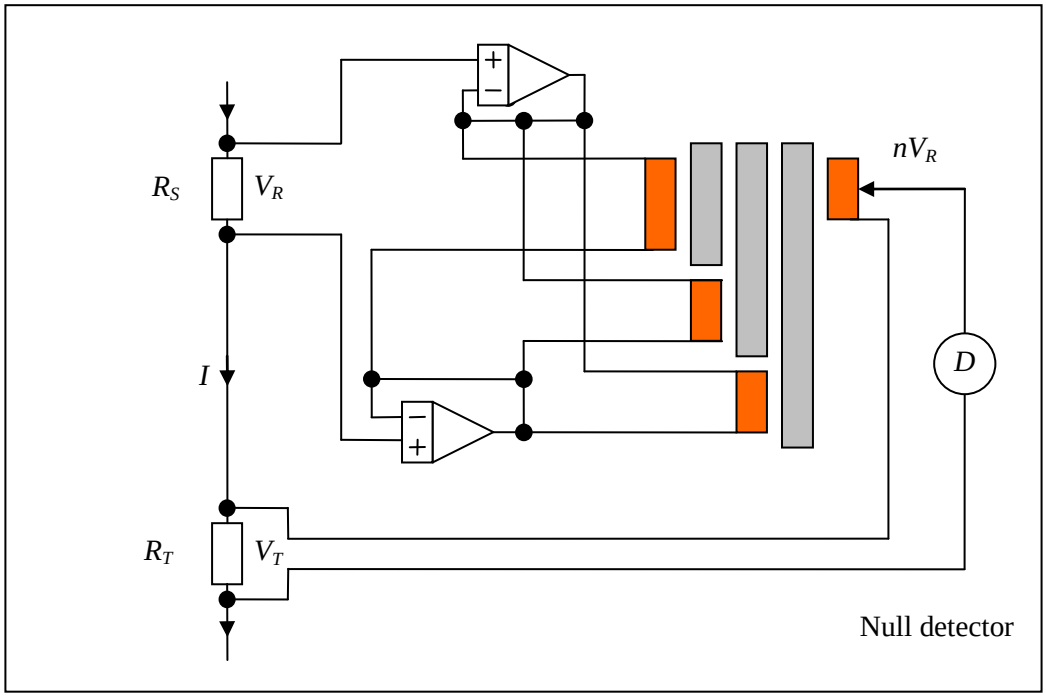


Fig. 3.1 A three-stage ratio transformer bridge (outline schematic)

The principle of operation is very simple: A current is passed through the transfer standard resistor, R_S, and the thermometer resistor, R_T. The current flowing through the resistors is the same and so the resistance ratio is the same as the voltage ratio. When the transformer ratio is adjusted for null balance: -

$$nV_R = V_T \qquad \Rightarrow \qquad \frac{R_T}{R_S} = \frac{V_T}{V_R} = n$$

The main disadvantage is the noise introduced by the followers. This limits speed and/or resolution to about 1ppm in 1 Hz of bandwidth – equivalent to approximately 0.25mK for platinum resistance thermometry. This is more than adequate for many commercial calibration requirements. For more see the monographs “Three-stage RTs” [1] and “An F17 type ratio transformer bridge” [2] by the same author.

1. Part 3, monograph 4: “Three-stage ratio transformers”.
2. Part 3, monograph 6: “An F17 type ratio transformer bridge”

5. An advanced three-stage ratio transformer (e.g. ASL model F18)

In resistance thermometry the self heating of the thermometer can be a major issue. This limits the current that can be used and, therefore, the voltage signal generated. This is why resistance thermometry is much more demanding on noise performance compared to the calibration of transfer standard resistors, where self heating is much less of a problem.

Similarly, for resistance thermometers with much lower resistance values (e.g. 0.25Ω at 0°C used for high temperature applications) the noise level becomes a major limitation. Fortunately it is possible to use followers to drive the energising windings only and still achieve sufficiently high input impedance on the ratio winding primary so that it can be connected directly to the transfer standard resistor.

To take advantage of the much lower noise contribution of the followers, one of the consequences is the need to reduce the noise level of the null detector also. This is possible, for low values of source resistance, by transformer matching to a low noise preamplifier. The ASL type F18, for example, has provision for matching to 1Ω, 10Ω or 100Ω source resistance.

The result is a versatile bridge with better than 0.1ppm accuracy and resolution limited by Johnson noise of the reference resistor and/or detector.

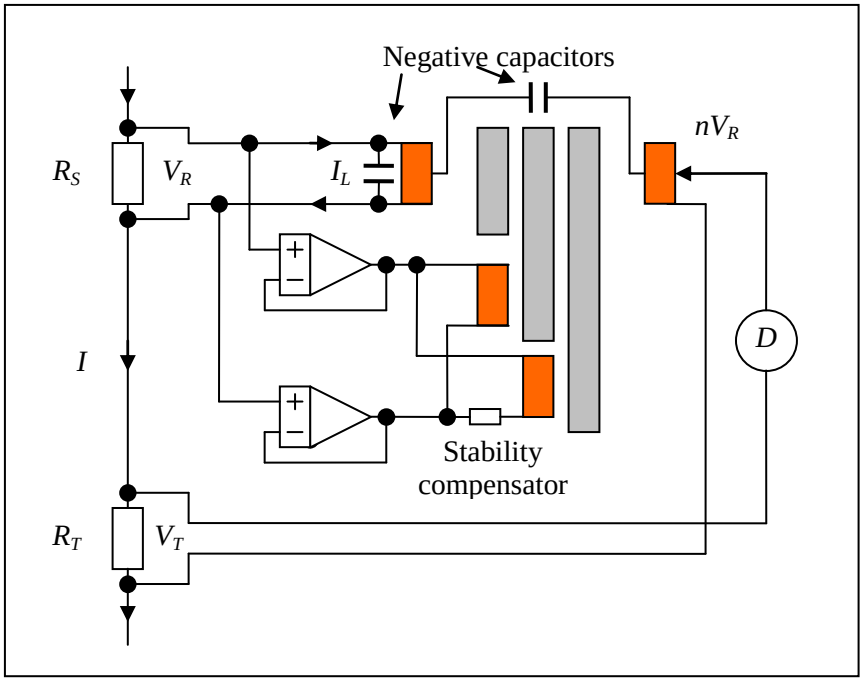


Fig. 4.1 An advanced three-stage ratio transformer bridge configuration

Historical note: A number of other workers produced multi-stage designs using a series of reference resistors to drive the energising windings. Robert Cutkosky of NBS, for example, produced a five-stage design, for low frequency operation (15Hz), with 10Ω reference resistors, for use with the recently developed 0.25Ω SPRT [1]. It was necessary to go to 5 stages due to the input capacitances interacting with the source resistance – not a problem with the very low output impedance of the voltage follower. He was also "forced" to adopt a doubly shielded fifth stage. JDY got round the problem in a simple and elegant way, obviating the need for shielding – negative capacitors, which cancel the inter-winding capacitance. Passively driven multi-stage transformers are, therefore, of historical interest only – there are no advantages and numerous disadvantages compared to actively driven transformers. For more detail see the monographs "Three-stage RTs" and "An F18 type ratio transformer bridge" and "A simulated negative capacitor circuit" by the same author.

1. Cutkosky, R, D: "An Automatic Resistance Thermometer Bridge". IEEE Trans. on Instrumentation and Measurement, Vol. IM-29, No. 4, Dec. 1980.

6. A three-stage current transformer (NPL's "Knight" bridge)

A well constructed three-stage ratio transformer works equally well as a current transformer. In an interesting case of symmetry, compared to a voltage transformer bridge, the current source and detector are interchanged. The same "bootstrapping" principle applies though, in the case of a multi-stage current transformer, the flux flowing in core T_3 is now very small. The following example is a three-stage current bridge originally developed at NPL [1] and made available by Tinsley Ltd. As mentioned above passive drive is of historical interest only, though with an advertised accuracy specification of ±2ppm with a 25Ω or 100Ω thermometer it is worth a mention.

N.B. For an ideal current transformer the flux cutting the ratio windings is zero and the currents are in the ratio of the number of turns - the total magneto-motive force (MMF) generated by the ampere-turns cancel.

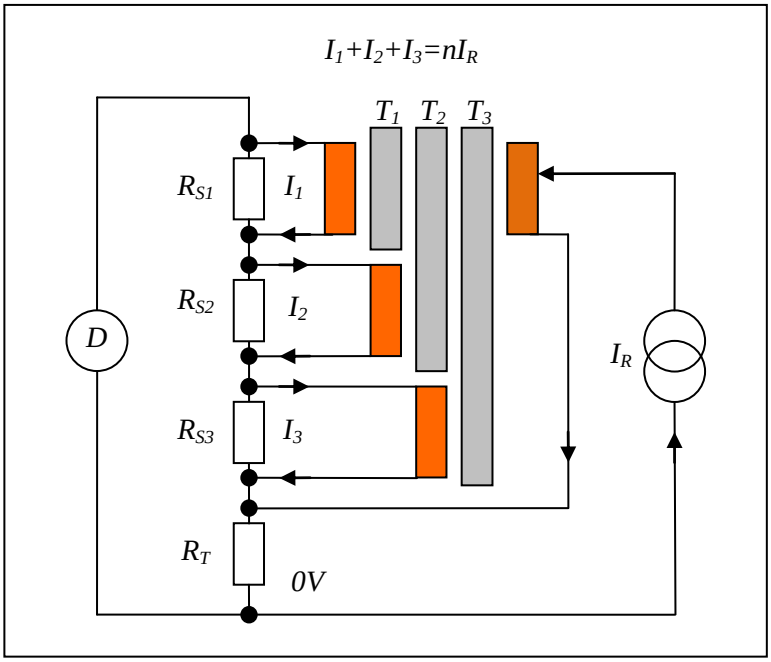


Fig. 5.1 A three-stage current transformer bridge (simplified schematic)

The bridge consists of three reference resistors which are ideally of the same value (= R_S). R_{S1} is the main standard resistor and must be very stable (<1ppm). R_{S2} is less critical (typically <100ppm) and R_{S3} is the least critical (<1%). The current passing through the variable ratio winding results in currents I_1, I_2 and I_3 in the fixed ratio windings. Current I_1 flows though R_{S1}, I_2 through R_{S2} etc.

To a very good approximation: $$I_1 + I_2 + I_3 = nI_R$$

At balance: $$I_R R_T - I_3 R_{S3} - I_2 R_{S2} - I_1 R_{S1} = 0$$

To a good approximation, therefore: $$I_R R_T = (I_1 + I_2 + I_3) R_S \quad \Rightarrow \quad n = \frac{R_T}{R_S}$$

The main disadvantage of this approach is the source resistance presented to the primary windings (compared to using followers) making it necessary to operate at the relatively high frequency of 400Hz. Also, in the commercial version, the reference resistors are internal to the instrument and can't be easily changed (e.g. for other applications and for checks on bridge accuracy with a resistance bridge calibrator).

1. Knight R. B. D. "A precision bridge for resistance thermometry using a single inductive current divider". IEE Conference No 152 (Euromeas 77).

7. Kusters' comparator (e.g. Guildline model 6622T)

Originally designed for calibrating transfer standard resistors [1] at relatively high measuring current this technique was then developed for resistance thermometry [2]. Whereas the title of the original paper characterises it as a "direct current" (DC) instrument it would be more accurate to describe it as operating at a very low frequency - in practice the DC current has to be reversed to eliminate the effect of DC offsets (e.g. thermal EMFs). For the model 6622T, for example, the reversal rate is selectable from 2 – 1,637 seconds (i.e. a square wave of frequency 0.25Hz or less). It is stated in [2], with unintended irony: "The reversing cycle of the bridge should be as short as possible in order to be able to cope with changing thermal EMFs."

The original electronic circuits and semi-automatic balance mechanism were also inept – introducing unnecessary extra noise. The null detector, for example, operates for only half of the available time, waiting for transients to settle after the "DC" current is reversed.

The main advantage is that it provides a means of comparing DC and AC measurements for resistance standards and sensors. This is less important recently as designers now understand the importance of materials and construction to minimise AC effects (e.g. insulators that do not exhibit dielectric absorption). The technique is also well suited to high accuracy measurements over a very wide range of resistance values (up to 100kΩ as standard and higher with a suitable extender unit and reference resistors).

The main disadvantage is the inevitably high noise level (mainly due to detector "1/f noise" and fluctuating thermal EMFs) compared to higher frequency operation (e.g. 10 – 400Hz).

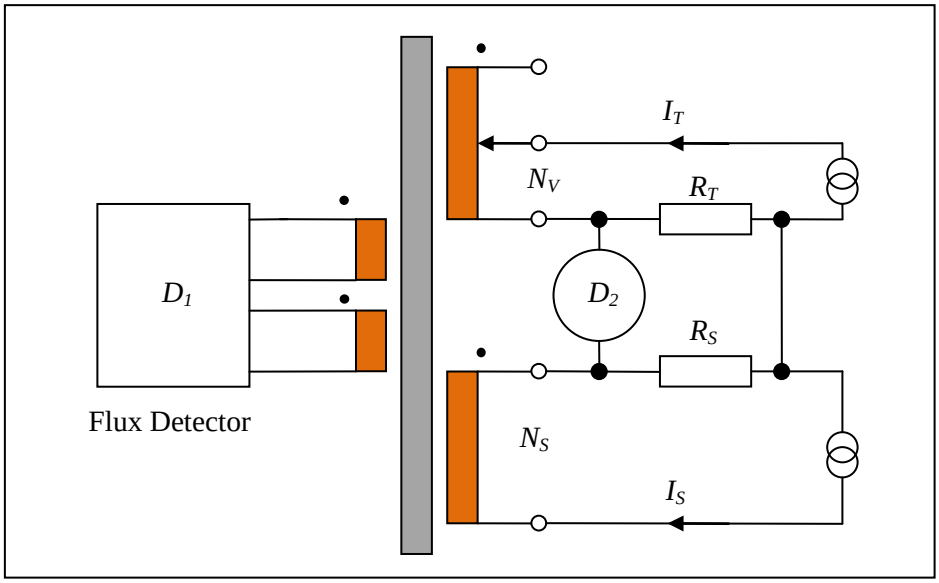


Fig. 7.1 The Kusters current comparator bridge (simplified schematic)

The transformer is a single stage device with a very large number of turns. The sensitive flux detector, D_1, is used to establish zero flux in the magnetic core. The magneto-motive force (MMF) generated by the fixed and variable windings are adjusted for a null balance: -

$$N_V I_T - N_S I_S = 0$$

The current I_S is also adjusted for null balance at the detector D_2: $I_T R_T - I_S R_S = 0$

To a very good approximation, therefore: $\frac{R_T}{R_S} = \frac{I_S}{I_T} = \frac{N_V}{N_S}$

1. MacMartin M. P. and Kusters N.L.: "Direct-Current Comparator Ratio Bridge for four-terminal resistance measurements" IEEE Trans. On Instrum. And Meas., vol. IM-15, pp 212-220, Dec 1966
2. Kusters N.L. and MacMartin M. P.: "Direct-Current Comparator Bridge for Resistance thermometry". IEEE Trans. on Instrum. And Meas., vol. IM-19, No. 4, Nov 1970

8. The double balanced potentiometer (e.g. The ASL "Cryo-bridge" [1])

Certain types of sensor (e.g. Germanium, used for cryogenic temperature measurements) have not only high resistance values but also high resistance in the connecting leads. Currents flowing in the voltage sensing leads must be negligible requiring an input resistance of the order $10^{12}\Omega$. High impedance circuits employing JFETs introduce too much noise - the signal level is severely constrained by heat dissipation in the sensor. Johnson noise of the sensor and connections is low because they are at very low temperature. One practicable solution is to use a ratio transformer with two sets of variable windings and a double balance. JFET input detectors can be transformer matched to provide optimum noise performance [2]. Despite a relatively low detector input impedance null balance ensures that the current passing through the detectors is negligible.

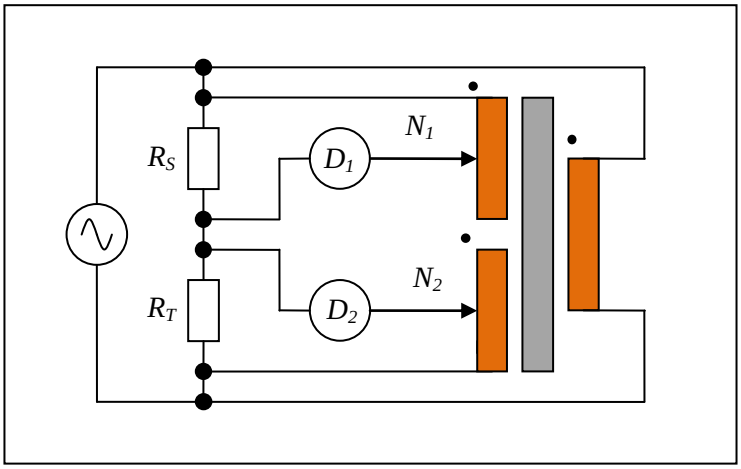


Fig. 8.1 The double balanced potentiometer (simplified schematic)

At balance: $$\frac{R_T}{R_S} = \frac{I_S}{I_T} = \frac{N_1}{N_2}$$

In practice the bridge measuring current was made sufficiently stable for a single variable winding and detector to be switched between the two resistors. Cable and inter-winding capacitances are significant issues, however, and a low operating frequency of 25Hz was therefore chosen. In the original "Cryo-bridge" [1] the transformer primary and secondary windings also had to be wound separately, reducing ratio accuracy to about ±1ppm. Quadrature is quite high (due to cable capacitance) and an accurate quadrature servo proved necessary. The main challenge is producing two sine waves with precisely 90 degrees phase difference (typically to within ±1 mRad) [3].

1. Wolfendale, P. C. F. "A precise automatic AC potentiometer for low temperature resistance thermometry". Journal of Scientific Instruments (Journal of Physics E) 1969, Series 2, vol. 2.
2. Part 3, monograph 5: "Noise matching transformers".
3. Part 5, monograph 1: "Null detectors – the basics".

Single stage inductors and transformers

1. Introduction

Inductors based on high permeability materials (typically μ_R > 50,000) exhibit significant non-linearity and hysteresis. At high levels of flux density (> 0.5 Tesla) the material saturates and the incremental permeability drops sharply. An applied sinusoidal voltage, for example, results in a distorted current waveform (and vice versa - see figs.1.1 and 1.2). Fortunately this does not prevent us from constructing and employing high value inductors (> 1H) as we normally rely only on a minimum inductance. This is readily achieved by designing magnetic circuits with a sufficient cross-sectional area so that the maximum flux density remains well below saturation level.

If one assumes a linear model with a minimum value for permeability one can at least predict minimum inductances and, therefore, maximum errors, which is the most important thing. Fortunately, measurements confirm the validity of this approach for flux density levels up to about 0.5 Tesla.

The other main consideration is winding resistance. This adds a real component to the mainly inductive (imaginary) impedance and is a major factor in the design of inductors and transformers. Whereas the resistance increases in direct proportion to the number of turns the inductance increases according to the number squared.

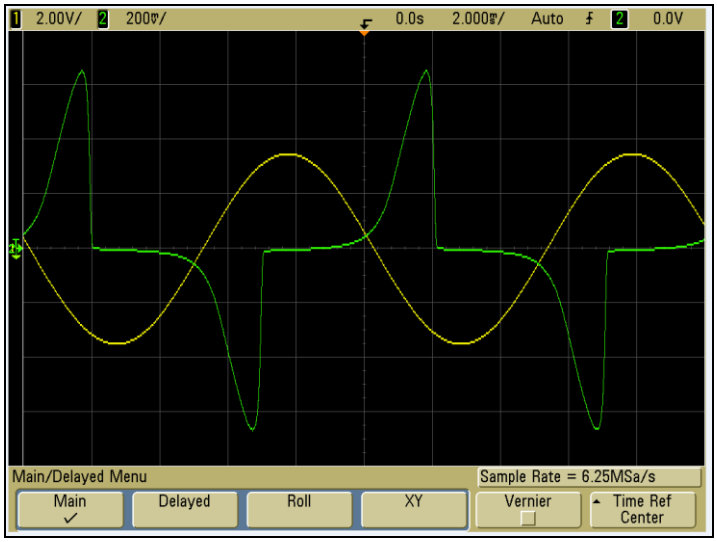


Fig. 1.1 Voltage and current traces for a high permeability core winding at high flux level

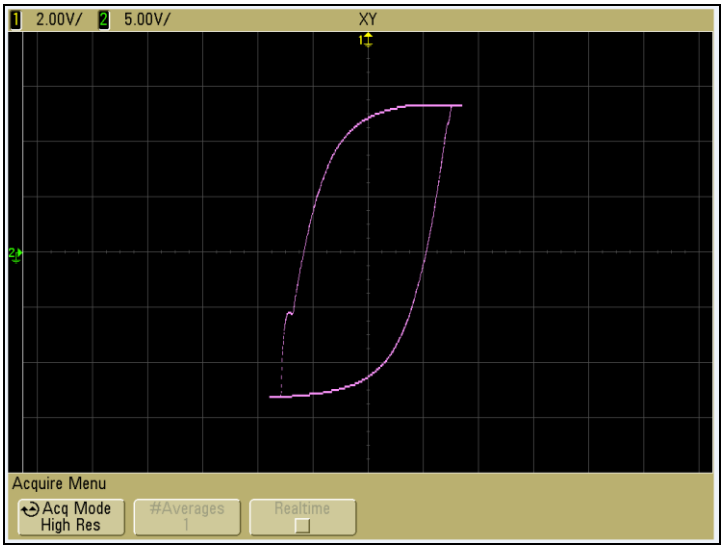


Fig. 1.2 B-H trace for a high permeability core

Rule of thumb for a good quality inductor and transformer design: high inductance to resistance ratio and low flux level (see sections 2 and 3).

2. Calculating impedance from basic parameters

In the following analysis I shall assume that the core material is linear with a constant value of permeability. I also assume that the effect of inter-winding capacitance is negligible and that the applied voltage is a sine wave.

Fig. 2.1 defines the main parameters of a basic inductor with winding resistance: -

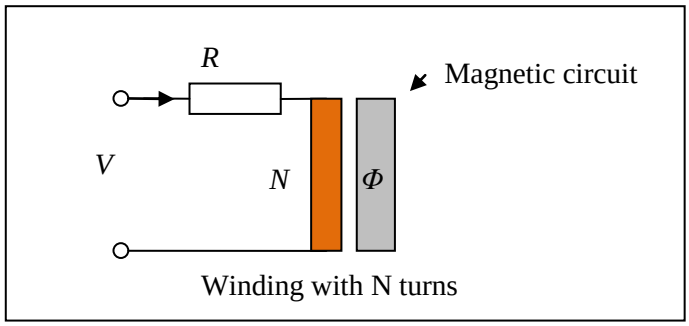


Fig. 2.1 An inductor with series resistance

The magneto-motive force (MMF) generated by the current flowing through the winding is defined as:-

$$MMF = NI \qquad \text{(units: "Ampere-turns")}$$

The resulting magnetic flux is determined by the reluctance R_L of the magnetic circuit.
(c.f. Ohm's law: V=IR): -

$$MMF = \Phi R_L$$

Reluctance depends on the cross-sectional area, length and permeability of the magnetic circuit: -

$$R_L = \frac{L_E}{\mu_0 \mu_R A_E}$$

N.B. The formula for reluctance can be compared with that for resistance of a conductor length L_E, cross-sectional area, A_E, and conductivity, σ: -

$$R = \frac{L_E}{\sigma A_E}$$

Clearly permeability is analogous to conductivity.

In practice the parameter provided in most data sheets for soft magnetic components is the reciprocal of reluctance. I call this the "permittance" [1]: -

$$A_L = \frac{1}{R_L} = \frac{\mu_0 \mu_R A_E}{L_E}$$

The flux is, therefore: $\Phi = NIA_L$

1. According to Wikipedia this term was originally used by Oliver Heaviside for the imaginary part of admittance, the reciprocal of impedance. That parameter is now widely referred to as the "susceptance".

The changing flux induces a "back EMF", according to Faraday's Law: -

$$V_B = N\frac{d\Phi}{dt} = N^2 A_L \frac{dI}{dt}$$

If one assumes the applied voltage is sinusoidal then using the complex representation: $\frac{dI}{dt} = j\omega I = sI$

$$V_B = N^2 A_L sI = sLI$$

Where L is the "self inductance": $L = N^2 A_L$

The applied voltage is balanced by the back-EMF plus the resistive voltage drop: -

$$V = IR + V_B = I(R + sL)$$

The total impedance is, therefore, the sum of winding resistance and inductive impedance: -

$$Z = \frac{V}{I} = R + sL$$

The reader will not be surprised that a key parameter is the relative value of the resistance and inductive impedance. A useful measure of this is the characteristic frequency at which the magnitude of the inductive impedance is the same as the resistance: -

$$\omega_{RL} = 2\pi f_{RL} = \frac{R}{L}$$

In virtually all cases a lower characteristic frequency is better – lots of turns of very thick copper wire. The art of inductor/transformer design, in this context, is to find the best practical compromise – minimum size and number of turns that is guaranteed to achieve the accuracy required.

3. High levels of flux density and core saturation

The effective cross-sectional area of the magnetic circuit, A_E, and the number of turns, N, determine the maximum voltage that can be applied at a given frequency. The operating frequency is usually quite low (typically 25Hz, 75Hz or 400Hz) and saturation can be a major consideration. If B_{MAX} is the maximum flux density then the maximum amount of flux is $A_E B_{MAX}$. The maximum (sine wave) voltage, at frequency, ω, is, therefore: -

$$V_{MAX} = N\frac{d\Phi_{MAX}}{dt} = \omega N A_E B_{MAX}\,.$$

In practical RT and IVD design the B_{MAX} is often set quite low (though not in all cases), compared to power transformer applications. As a rule of thumb a B_{MAX} of 500mT is the most one can tolerate for mumetal and similar alloys.

See appendix 1: data for a range of commercially available magnetic cores including the maximum voltage per turn per hertz assuming a B_{MAX} of 500mT.

3.1 Example calculation

The following applies to a simple resistance bridge (see section 4.6).

A size 3A core with 128 turns at 400Hz: $V_{MAX} = 2\pi f N(A_E B_{MAX}) \approx 23V$

This is well above the operating voltage anticipated (≈ 100mV) and the core operates well below saturation.

4. Basic (single-stage) transformer theory

Consider a transformer with an ideal (linear) magnetic circuit but with winding resistances, driving load impedance Z_L. Fig. 4.1 defines the main parameters: -

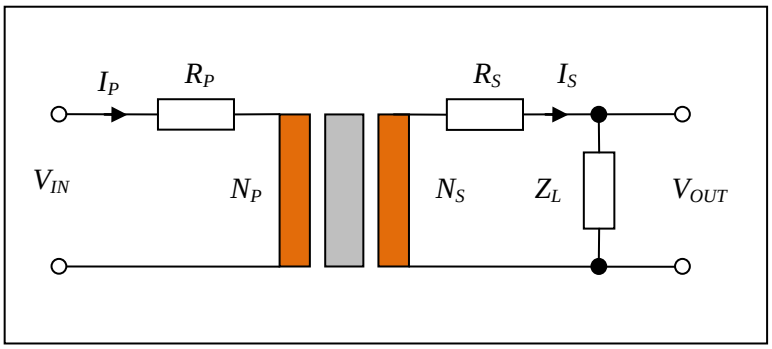


Fig. 4.1 Model of a basic transformer with load impedance

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $(s = j\omega)$: -

$$V_{IN} = I_P R_P + sN_P\Phi \qquad V_{OUT} = sN_S\Phi\frac{Z_L}{R_S + Z_L} \qquad \text{and} \qquad V_{OUT} = I_S Z_L$$

The flux generated by the net magneto-motive force is proportional to permittance of the magnetic circuit: -

$$\Phi = (N_P I_P - N_S I_S)A_L$$

4.1 Current transformers

From the above one can deduce: -

$$V_{OUT} = sN_S(N_P I_P - N_S I_S)A_L\frac{Z_L}{R_S + Z_L} = I_S Z_L$$

With a little algebra the current ratio is: -

$$\frac{I_S}{I_P} = \frac{sN_S N_P A_L}{R_S + Z_L + sN_S^2 A_L}$$

If $R_S = Z_L = 0$ the result is an ideal current transformer: $\dfrac{I_S}{I_P} = \dfrac{N_P}{N_S}$

Otherwise: $\dfrac{I_S}{I_P} = \dfrac{N_P}{N_S}\left(\dfrac{\alpha s}{1+\alpha s}\right)$ with $\alpha = \dfrac{N_S^2 A_L}{R_S + Z_L} = \dfrac{L_S}{R_S + Z_L}$

Where: L_S is the inductance of the secondary winding. This is the value of inductance that would be measured with the primary winding open circuit.

There is a practical limit to how low one can make the winding resistance but it is possible to make the load impedance very small (e.g. with active circuitry), in which case $\alpha = \frac{L_S}{R_S}$ and is real. I call this the time constant of the secondary winding. The result is the transfer function with a first order high-pass characteristic: -

$$\frac{I_S}{I_P} = \frac{N_P}{N_S}\left(\frac{\tau s}{1+\tau s}\right) \qquad \text{with} \qquad \tau = \frac{L_S}{R_S}$$

At high frequency, to a very good approximation: $|\tau s| >> 1 \;\Rightarrow\; \frac{I_S}{I_P} \approx \frac{N_P}{N_S}\left(1 - \frac{1}{\tau s} + \frac{1}{\tau^2 s^2}\right)$

More conveniently with frequency in Hz: -

$$f >> f_N \;\Rightarrow\; \frac{I_S}{I_P} \approx \frac{N_P}{N_S}\left(1 + j\frac{f_N}{f} - \left(\frac{f_N}{f}\right)^2\right) \quad \text{with} \quad f_N = \frac{R_S}{2\pi L_S}$$

The first error term is quadrature and inversely proportional to frequency. The in-phase error is inversely proportional to frequency squared and is, in practice, too large for low frequency high accuracy applications (typically 25Hz or 75Hz).

4.2 Voltage transformers

From above the voltage ratio is: -

$$\frac{V_{OUT}}{V_{IN}} = \frac{sN_S\Phi}{I_P R_P + sN_P\Phi} \times \frac{Z_L}{R_S + Z_L}$$

With no load current and zero primary winding resistance the result is an ideal voltage transformer: -

$$|Z_L| = \infty \qquad \text{and} \qquad R_P = 0 \;\Rightarrow\; \frac{V_{OUT}}{V_{IN}} = \frac{N_S}{N_P}$$

Neither is possible in practice and so one needs an approximation. One can eliminate the flux and current by noting, from above: -

$$\frac{I_S}{I_P} = \frac{N_P}{N_S}\left(\frac{\alpha s}{1+\alpha s}\right) \;\Rightarrow\; \Phi = \left(N_P I_P - N_S I_S\right)A_L = \left(N_P I_P - N_P I_P \frac{\alpha s}{1+\alpha s}\right)A_L = \frac{N_P I_P A_L}{1+\alpha s}$$

$$\text{with} \qquad \alpha = \frac{L_S}{R_S + Z_L}$$

$$\Rightarrow \quad \frac{V_{OUT}}{V_{IN}} = \frac{N_S}{N_P} \times \frac{sL_P}{R_P(1+\alpha s) + sL_P} \times \frac{Z_L}{R_S + Z_L}$$

Where L_P is the inductance of the primary winding: $L_P = N_P^2 A_L$

N.B. This is the inductance that would be measured with the secondary open circuit.

4.2.1 Example calculation

Consider a simple step-down transformer, ratio 100:1 consisting of 200 turns on the primary and 2 turns on the secondary on a 5c SM100 toroidal core. The load is the primary of a similar transformer as often happens with a multi-decade transformer: -

Size 5c core: case OD 41mm and case ID 23mm. axial length 15mm.
Inner circumference: $=\pi\times 23=72mm$ with room for 200 turns in two layers of 0.7mm OD wire.
Outer circumference: $=\pi\times 41=128mm$ with room for 200 turns in one layer of 0.64mm OD wire.
The outer circumference is the more critical so choose 24SWG enamelled copper wire with 0.56mm OD and 71mΩm^{-1} resistance.

Length of wire per turn is approximately: $2\times 15+41-23=48mm$
For 200 turns: $200\times 48mm=9.6m$ of wire.
Primary resistance: $9.6m\times 71m\Omega m^{-1}\approx 0.7\Omega$
Primary inductance = $200\times 200\times 102\,\mu H/turn^2\approx 4H$
Primary impedance at 75Hz = $2\pi\times 75\times 4=1.88k\Omega$

The secondary resistance is mainly due to the connecting leads. If one assumes a 50cm long twisted pair then the resistance is $1.096m\times 71m\Omega m^{-1}\approx 78m\Omega$
The secondary inductance is: $2\times 2\times 102\,\mu H/turn^2\approx 400\mu H$
From above, repeated for convenience: -

$$\frac{V_{OUT}}{V_{IN}}=\frac{N_S}{N_P}\times\frac{sL_P}{R_P(1+\alpha s)+sL_P}\times\frac{Z_L}{R_S+Z_L}\qquad\text{with}\qquad\alpha=\frac{L_S}{R_S+Z_L}$$

$$|Z_L|>>R_S\ \Rightarrow\ \alpha s=\frac{sL_S}{R_S+Z_L}\approx\frac{sL_S}{Z_L}\approx\frac{sL_S}{sL_P}\approx 10^{-4}$$ i.e. the ratio of turns squared and negligible.

The loading effect on the secondary is very small and substantially quadrature: -

$$\frac{Z_L}{R_S+Z_L}\approx 1-\frac{R_S}{Z_L}+\left(\frac{R_S}{Z_L}\right)^2\approx 1+j\frac{78m\Omega}{1.88k\Omega}-\left(\frac{78m\Omega}{1.88k\Omega}\right)^2=1+j4.1\times 10^{-5}-1.7\times 10^{-9}$$

The errors are 41ppm of quadrature and a negligible in-phase component. It is reasonable to assume, therefore, at low frequency: -

$$|\alpha s|<<1\ \Rightarrow\ \omega<<10^4\ \text{ and }\ |Z_L|>>R_S\ \Rightarrow\ \frac{V_{OUT}}{V_{IN}}=\frac{N_S}{N_P}\times\frac{sL_P}{R_P+sL_P}$$

This is a simple high pass filter characteristic due to the primary resistance and inductance and the main limitation on accuracy in most cases. From above, in more convenient form (frequency in Hz) over a useful range of frequency: -

$$f_N<<f<<\frac{10^4}{2\pi}(1.6kHz)\ \Rightarrow\ \frac{V_{OUT}}{V_{IN}}\approx\frac{N_S}{N_P}\left(1+j\frac{f_N}{f}-\left(\frac{f_N}{f}\right)^2\right)\quad\text{with}\quad f_N=\frac{R_P}{2\pi L_P}=0.028Hz$$

$$\text{At 75Hz:}\quad\Rightarrow\ \frac{V_{OUT}}{V_{IN}}\approx\frac{N_S}{N_P}\left(1+j3.7\times 10^{-4}-1.4\times 10^{-7}\right)$$

The main error is 370ppm of quadrature.

4.3 Input impedance

From above: -

$$V_{IN} = I_P R_P + sN_P\Phi \quad \text{and} \quad \Phi = \frac{N_P I_P A_L}{1+\alpha s} \quad \Rightarrow \quad Z_P = \frac{V_{IN}}{I_P} = R_P + \frac{sN_P^2 A_L}{1+\alpha s} = R_P + \frac{sL_P}{1+\alpha s}$$

$$\text{with} \qquad \alpha = \frac{L_S}{R_S + Z_L}$$

If the load impedance is infinite (open circuit secondary) the input impedance is, as expected, the primary resistance in series with its inductance: -

$$Z_L = \infty \quad \Rightarrow \quad \alpha = 0 \quad \Rightarrow \quad Z_P = R_P + sL_P$$

Otherwise, the load impedance is reflected back to the primary, acting in parallel with the primary inductance. As above the input impedance is: -

$$Z_P = R_P + \frac{sL_P}{1+\alpha s}$$

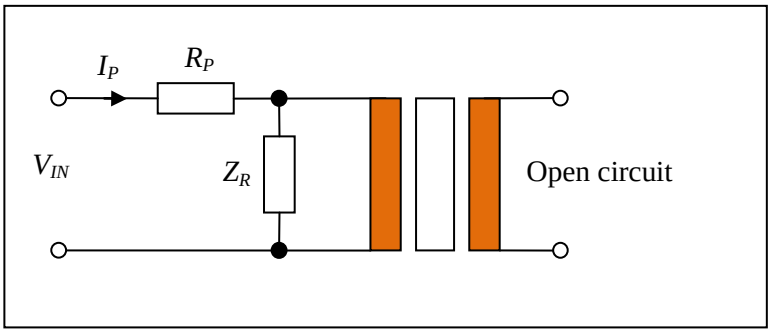


Fig. 4.3.1Equivalent circuit – load impedance reflected back to the primary

Expressed as admittance (the reciprocal of impedance) the primary winding inductance in parallel with the reflected load impedance is: -

$$\frac{1}{Z_P - R_P} = \frac{1+\alpha s}{sL_P} = \frac{1}{sL_P} + \frac{1}{L_P}\frac{L_S}{(R_S + Z_L)}$$

The first component is the admittance of the primary winding, the second is, therefore the reflected admittance. The reflected impedance is, therefore: -

$$Z_R = \frac{L_P}{L_S}(R_S + Z_L)$$

The ratio of inductance is the turns ratio squared: -

$$\Rightarrow \quad Z_R = \left(\frac{N_P}{N_S}\right)^2 (R_S + Z_L)$$

Very often the load impedance is much greater than the secondary winding resistance and, therefore: -

$$|Z_L| >> R_S \qquad \Rightarrow \quad Z_R \approx \left(\frac{N_P}{N_S}\right)^2 Z_L$$

The square law is an important characteristic of ratio transformers and inductive voltage dividers.

4.4 Output impedance

One can model the effect, solely due to loading, as follows: -

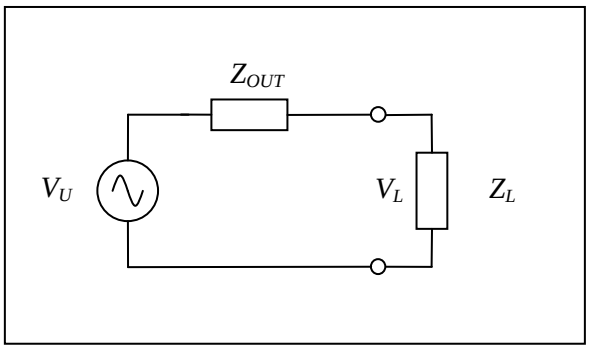


Fig. 4.4.1 Circuit model for output impedance

From above, the loaded and unloaded output voltages are, respectively: -

$$V_L = V_{IN}\frac{N_S}{N_P}\frac{sL_P}{R_P(1+\alpha s)+sL_P}\times\frac{Z_L}{R_S+Z_L} \qquad \text{and} \qquad V_U = V_{IN}\frac{N_S}{N_P}\frac{sL_P}{R_P+sL_P}$$

$$\text{With: } \alpha = \frac{L_S}{R_S+Z_L}$$

The effect of loading due to the secondary output impedance is: -

$$\frac{V_L}{V_U} = \frac{Z_L}{Z_{OUT}+Z_L} \Rightarrow \quad Z_{OUT} = Z_L\left(\frac{V_U}{V_L}-1\right)$$

$$\Rightarrow \quad Z_{OUT} = Z_L\left(\frac{R_P(1+\alpha s)+sL_P}{R_P+sL_P}\times\frac{R_S+Z_L}{Z_L}-1\right)$$

$$\Rightarrow \quad Z_{OUT} = \frac{R_P(R_S+Z_L)+\alpha s R_P(R_S+Z_L)+sL_P(R_S+Z_L)-Z_L(R_P+sL_P)}{R_P+sL_P}$$

Two terms cancel. Eliminate $\alpha = \frac{L_S}{R_S+Z_L}$ for the simple result: -

$$Z_{OUT} = \frac{R_P R_S + sL_S R_P + sL_P R_S}{R_P+sL_P} = R_S + \frac{sR_P L_S}{R_P+sL_P}$$

As one might expect the output impedance is the secondary winding resistance in series with an extra term reflected from the primary. It is the primary winding resistance in parallel with the winding inductance, multiplied by the ratio of inductances. The latter is, again, the turns ratio squared: -

$$Z_R = \frac{sR_P L_S}{R_P+sL_P} = \frac{L_S}{L_P}\frac{sL_P R_P}{R_P+sL_P} = \left(\frac{N_S}{N_P}\right)^2\frac{sL_P R_P}{R_P+sL_P}$$

This is also an important characteristic of ratio transformers and inductive voltage dividers.

This can be represented as follows: -

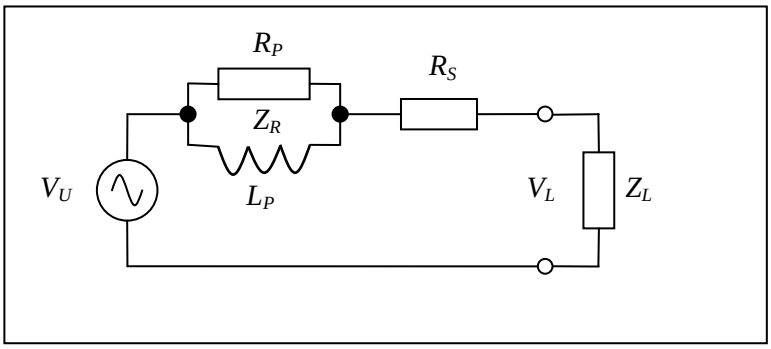


Fig. 4.4.2 Equivalent circuit showing the reflected impedance, Z_R

At a sufficiently high frequency (greater than the characteristic frequency of the winding) the inductive impedance of the primary is much greater than the winding resistance and the expression simplifies further: -

$$\left|sL_P\right| >> R_P \quad \Rightarrow \quad Z_{OUT} \approx R_S + \left(\frac{N_S}{N_P}\right)^2 R_P$$

4.5 Capacitive load

In some cases (e.g. a capacitance bridge) the load impedance is capacitive, (e.g. due to interconnecting cables) with a magnitude much greater than the winding resistance:-

$$Z_L = \frac{1}{sC_L} \qquad \text{and} \qquad sR_SC_L << 1 \quad \Rightarrow \quad \alpha = \frac{sL_SC_L}{1+sR_SC_L} \approx sL_SC_L$$

Also, if the operating frequency is much lower than the characteristic frequency of the secondary inductance/load capacitance, then the value of α is also very small and, to a very good approximation: -

$$\left|sL_SC_L\right| << 1 \quad \Rightarrow \quad \frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P} \times \frac{sL_P}{R_P + sL_P} \times \frac{1}{1+sR_SC_L}$$

The resulting transfer function can be summarised as a first order high-pass and low-pass characteristic: -

$$\frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P} \times \frac{\beta s}{1+\beta s} \times \frac{1}{1+\gamma s} \qquad \text{with} \qquad \beta = \frac{L_P}{R_P} \qquad \text{and} \qquad \gamma = R_SC_L$$

It is often the case that one effect dominates and it is best to keep the effects separate. To a very good approximation: -

$$\left|\beta s\right| >> 1 \qquad \text{and} \qquad \left|\gamma s\right| << 1 \Rightarrow \qquad \frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P}\left(1 - \frac{1}{\beta s} + \frac{1}{\beta^2 s^2}\right) \times \left(1 - \gamma s + \gamma^2 s^s\right)$$

In more convenient form: $\qquad f = \frac{1}{2\pi}\frac{R_P}{L_P} >> f_{RL} \qquad$ and $\qquad f = \frac{1}{2\pi R_SC_L} << f_{RC}$

$$\Rightarrow \quad \frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P}\left(1 + j\frac{f_{RL}}{f} - \left(\frac{f_{RL}}{f}\right)^2\right) \times \left(1 - j\frac{f}{f_{RC}} - \left(\frac{f}{f_{RC}}\right)^2\right)$$

4.5.1 Example calculation

The output of a ratio transformer is connected to a remote capacitive sensor via 10m of coaxial cable (100pFm^{-1}). The other side of the bridge is connected to a local reference capacitor, hence a significant imbalance.

The transformer includes a 20 turn primary and 200 turn secondary (ten strands with 20 turns) of 24SWG copper wire on a 5c SM100 toroidal core. The primary consists of 5 sets of 20 turns (BNNL) connected in parallel (to ensure a uniform sheet of current around the core). From the previous example, very approximately (the length of wire and resistance is slightly higher due to the rope construction): -

Secondary inductance = $200 \times 200 \times 102\,\mu H/turn^2 \approx 4H$

Secondary resistance: $\approx 1\Omega$

Primary inductance = $20 \times 20 \times 102\,\mu H/turn^2 \approx 40mH$

Primary resistance $\approx 0.02\Omega$

N.B. to maintain a very low primary resistance the negative feedback connection (to the driving stage) is taken from as close as possible to the transformer terminals, avoiding "tails". From above: -

$$f >> f_{RL} \quad \text{and} \quad f << f_{RC} \Rightarrow \quad \frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P}\left(1 + j\frac{f_{RL}}{f} - \left(\frac{f_{RL}}{f}\right)^2\right) \times \left(1 - j\frac{f}{f_{RC}} - \left(\frac{f}{f_{RC}}\right)^2\right)$$

$$\text{with} \quad f_{RL} = \frac{1}{2\pi}\frac{R_P}{L_P} \approx 0.08Hz \qquad \text{and} \quad f_{RC} = \frac{1}{2\pi R_S C_L} \approx 160MHz$$

$$\frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P}\left(1 + j5 \times 10^{-5} - 2.5 \times 10^{-9}\right) \times \left(1 - j10^{-5} - 10^{-10}\right)$$

$$\frac{V_{OUT}}{V_{IN}} \approx \frac{N_S}{N_P}\left(1 + j4 \times 10^{-5} - 2.6 \times 10^{-9}\right)$$

40ppm of quadrature can be easily rejected with a suitable quadrature servo or accurate signal conditioner/phase sensitive detector. The in-phase component is negligible.

4.6 A single stage resistance bridge

The relatively low input impedance and limited accuracy (primary to secondary), at low frequency, make single stage RTs of limited application. The performance of IVDs and RTs is much improved with extra energising stages [1 and 2].

The following, however, can be used as part of a simple (low cost) resistance thermometry bridge (accuracy ≈ 10mK), operating at 400Hz, by employing a high accuracy voltage follower [3] with semiconductor switches. The follower forces the reference voltage across a standard resistor. The resulting current flows through the thermometer resistor which creates the second voltage. At null balance the ratio of voltages is the ratio of resistances.

1. Part 3, monograph 3: "Two-stage IVDs and RTs".
2. Part 3, monograph 4: "Three-stage RTs".
3. Part 4, monograph 2: "High accuracy voltage followers". See section 6.3.

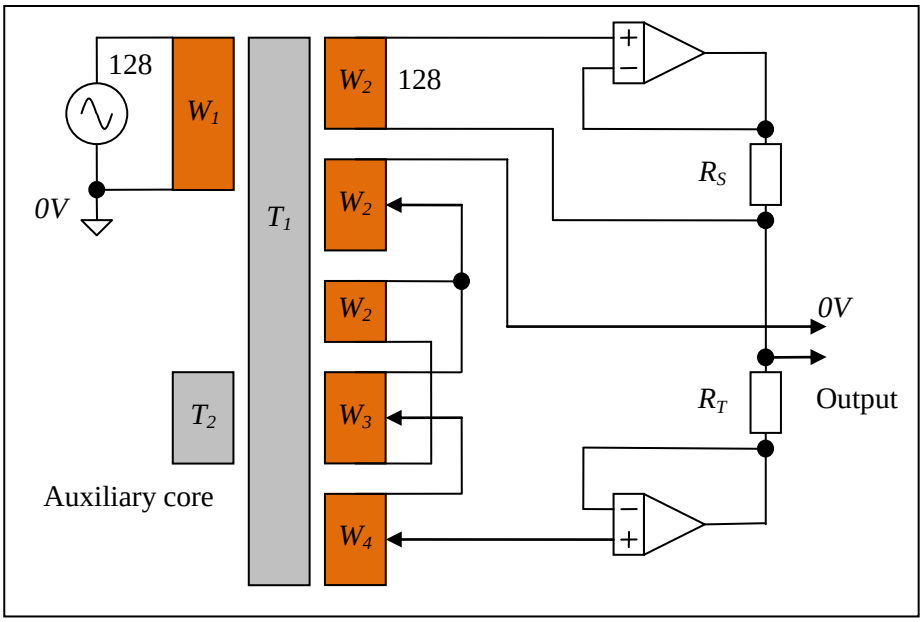


Fig. 4.6.1 A single-stage two-terminal pair resistance bridge

Winding1: 128 turns 0.315mm enamelled copper wire wound BNNL [1].
This carries the current to energise the main core. The precise number of turns is not important.

Winding 2: 128 turns of a 5 strand rope of 0.2mm wire (BNNL).
One of the strands provides the fixed reference voltage, V_S. Another is connected across winding 3 which, with the extra turns around the auxiliary core, ensures an accurate voltage match between V_S and the outputs of winding 3 ("equalisation": see section 6). The other 3 are connected in series and switched to provide the first (most significant) two bits of the (binary) divider.

Winding 3: 8 turns (NNL) of a 16 strand rope of 0.2mm wire around the main core (size 3A mumetal) plus 3 turns around the auxiliary core (size 1A mumetal). The 16 strands are connected in series and switched (0 – 15) to provide the next four bits of the divider. The voltage developed across all 16 strands (total 128) may not be quite the same as V_S, due to flux leakage, hence the equalisation winding and auxiliary core.

Winding 4: 1 turn of 8 strand rope. The 8 strands are connected in series and switched (0 – 7) to provide the next three bits of the divider. In this case the accuracy is sufficient not to require equalisation with winding 3.

The variable winding thus varies from 0 – 512 turns and a ratio in the range of 0 – 4 and more than sufficient to cover the range of resistance of platinum resistance thermometers with a suitable standard resistor (typically 25Ω or 100Ω).

The auxiliary core ensures ratio accuracy between the two main ropes. See section 6 for details.

The ratio transformer provides the first 9 bits of a null balance bridge. Further accuracy and resolution could be provided by an R-2R ladder 8 or 12 bit multiplying DAC or by interpolation with an analogue to digital converter.

5. Inductive voltage dividers (auto-transformers)

The analysis for single stage IVDs, often referred to as auto-transformers, is exactly the same. The result is a device with usefully high input impedance and low output impedance, making multi-decade dividers relatively easy to construct (i.e. the Kelvin-Varley method). See the monograph "IVDs and RTs – the basics" by the same author [2].

1. Part 3, monograph 1: "IVDs and RTs – the basics". For balanced no-net-loop (BNNL) etc see section 1.3.
2. Ibid. For the Kelvin-Varley method see section 2.

6. Equalising windings

A second decade variable winding can usually be provided by a second rope with one tenth the number of turns of the first decade and employed as an IVD. Being part of the same rope the voltage of each strand should accurately match. Unfortunately it is possible that flux leakage, particularly at the start and end of the magnetic strip that makes up the "clock-spring" core, can result in a few parts per million (ppm) of difference between decades 1 and 2 because they are not on the same rope. The most practical solution is to include an extra strand to the first decade rope (the "equalising" winding) and connect it in parallel with ten strands of the second decade. Extra inductance is added in series with the decade 2 rope by winding extra turns around an auxiliary core. This is often represented, employing the usual convention, as follows: -

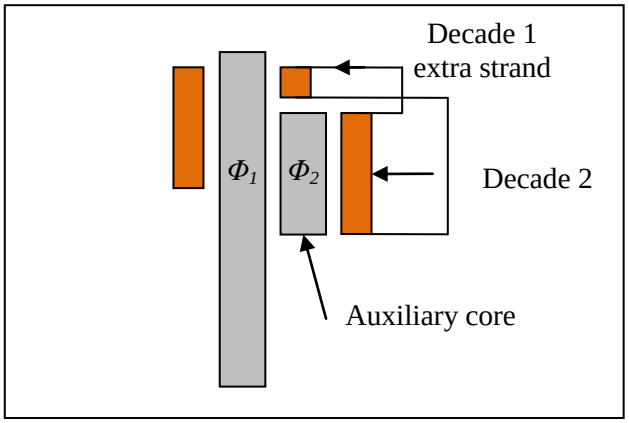


Fig. 6.1 Equalising voltages between the first and second decade

Unfortunately this representation is not particularly revealing and a more representative diagram is as follows: -

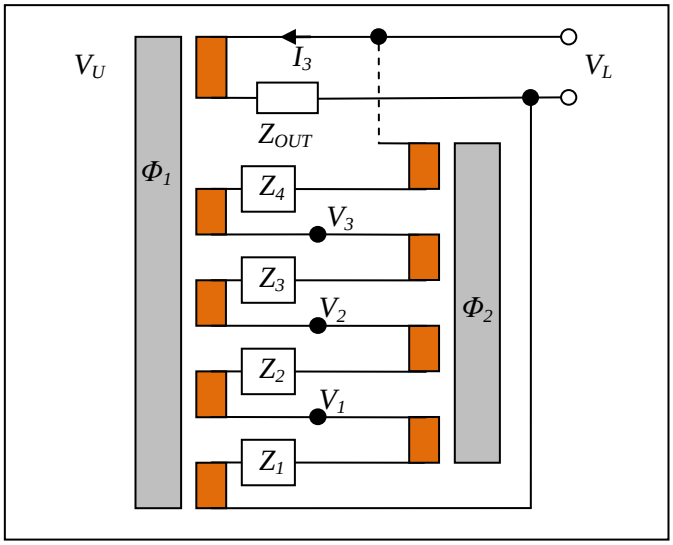


Fig. 6.2 Detailed diagram of an equalising scheme (only four strands shown)

The flux through the main core is Φ_1. The flux in the auxiliary core is $\Phi_2 = N_A I_3 A_{L2}$

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $(s = j\omega \equiv d/dt)$: -

The unloaded voltage induced in the equalising winding is: $V_U = sN_S\Phi_1$

The loaded voltage includes the error voltage due to the flowing current: $V_L = V_U + I_3 Z_{OUT}$

$$\Rightarrow \quad sN_S\Phi_1 = V_L - I_3 Z_{OUT}$$

The total voltage in the rest of the circuit consists of the induced voltages, including the flux leakage represented as a relative factor, δ, and the voltage dropped across all ten winding resistances and any reflected impedance from the primary side. These are represented (in fig 6.2) as lumped impedances in series with each strand. As these are very small (mainly the auxiliary winding resistance) and closely matched the second decade IVD remains very linear: -

$$V_L = sN_S\Phi_1(1+\delta) - sN_A\Phi_2 - I_3Z_A \qquad \text{with} \qquad Z_A = Z_1 + Z_2 + ... + Z_{10}$$

$$N_S\Phi_1 = V_L - I_3Z_{OUT} \qquad \text{and} \qquad \Phi_2 = N_AI_3A_{L2} \quad \Rightarrow \quad V_L = (V_L - I_3Z_{OUT})(1+\delta) - I_3(sL_A + Z_A)$$

The small voltage $\delta I_3 Z_{OUT}$ is negligible and the error, expressed as a ratio, is, therefore: -

$$\frac{I_3Z_{OUT}}{V_L} = \delta\frac{Z_{OUT}}{(Z_{OUT} + sL_A + Z_A)}$$

The impedance of the auxiliary windings is much greater than the output impedance of the equalising winding and certainly much larger than the total impedances Z_A. To a good approximation, therefore: -

$$\frac{I_3Z_{OUT}}{V_L} \approx \delta\frac{Z_{OUT}}{sL_A}$$

6.1 Example calculation

A typical (competitive) target accuracy is an overall error of 0.1ppm and a reasonable target for the second decade contribution is, therefore, also 0.1ppm (even though its effect is reduced by a factor of 10). With a worst case mismatch of 10ppm due to flux leakage one can conclude: -

$$\left|\frac{I_3Z_{OUT}}{V_L}\right| < 10^{-7} \quad \text{and} \qquad \delta \approx 10^{-5} \quad \Rightarrow \quad sL_A > 100Z_{OUT}$$

With a maximum source resistance of 100Ω which, when reflected to the first decade secondary, appears as an output resistance of 1Ω. The auxiliary impedance at the operating frequency needs to be greater than 100Ω. This is easily achieved with a sufficiently large auxiliary core with very few turns. The choice often reduces to one of practicality and aesthetics (a nice fit that sits on top).

If the resistance of the energising winding becomes an issue then a simple solution is to include two or three extra strands to the decade 1 rope and connect them in parallel.

Appendix 1: Toroidal core data courtesy Telcon

Type	Axial code	Case OD	Case ID	Case axial	Circ./turn	Core OD	Core ID	Core axial	L_E	A_E	A_L μH/turn^2	V_{MAX}/turn/Hz
0	a	18	11	5	18	16	13	3.2	45	5	14	1.6E-05
	b	18	11	8	24	16	13	6.4	45	10	28	3.2E-05
1	a	21	11	5	21	19	13	3.2	50	10	26	3.2E-05
	b	21	11	7	24	19	13	4.8	50	15	38	4.8E-05
	c	21	11	8	27	19	13	6.4	50	20	51	6.4E-05
2	a	24	12	7	26	22	14	4.8	57	19	41	5.9E-05
	b	24	12	8	29	22	14	6.4	57	25	55	7.9E-05
	c	24	12	10	32	22	14	7.9	57	31	69	9.9E-05
3	a	31	17	7	29	29	19	4.8	75	23	38	7.1E-05
	b	31	17	9	32	29	19	6.4	75	30	51	9.5E-05
	c	31	17	10	36	29	19	7.9	75	38	63	1.2E-04
4	a	36	20	9	34	33	22	6.4	87	35	51	1.1E-04
	b	36	20	10	37	33	22	7.9	87	44	63	1.4E-04
	c	36	20	12	40	33	22	9.5	87	53	76	1.7E-04
5	a	41	23	9	36	38	25	6.4	100	40	51	1.3E-04
	b	41	23	12	42	38	25	9.5	100	60	76	1.9E-04
	c	41	23	15	48	38	25	12.7	100	81	102	2.5E-04
6	a	50	29	7	36	48	32	4.8	125	38	38	1.2E-04
	b	50	29	10	42	48	32	7.9	125	63	63	2.0E-04
	c	50	29	12	45	48	32	9.5	125	75	76	2.4E-04
	d	50	29	15	52	48	32	12.7	125	100	101	3.2E-04
7	a	60	36	7	38	57	38	4.8	150	45	38	1.4E-04
	b	60	36	12	48	57	38	9.5	150	91	76	2.9E-04
	c	60	36	15	54	57	38	12.7	150	121	102	3.8E-04
8	a	69	42	10	48	67	45	7.9	174	88	64	2.8E-04
	b	69	42	14	54	67	45	11.1	174	123	89	3.9E-04
	c	69	42	17	61	67	45	14.3	174	159	115	5.0E-04
9	a	79	49	14	58	76	51	11.1	200	141	89	4.4E-04
	b	79	49	17	64	76	51	14.3	200	182	114	5.7E-04
	c	79	49	15	61	76	51	12.7	200	161	101	5.1E-04
10	a	90	54	15	65	86	57	11.1	224	158	89	5.0E-04
	b	90	54	18	72	86	57	14.3	224	204	114	6.4E-04
	c	90	54	23	81	86	57	19.1	224	272	153	8.6E-04
11	a	112	66	16	78	108	70	12.7	279	242	109	7.6E-04
	b	112	66	19	85	108	70	15.9	279	303	136	9.5E-04
	c	112	66	23	91	108	70	19.1	279	364	164	1.1E-03
12	a	132	78	18	90	127	83	14.3	329	317	121	1.0E-03
	b	132	78	23	100	127	83	19.1	329	424	162	1.3E-03
	c	132	78	29	112	127	83	25.4	329	564	215	1.8E-03
13	a	164	110	18	90	159	114	14.3	429	322	94	1.0E-03
	b	164	110	23	100	159	114	19.1	429	430	126	1.4E-03
	c	164	110	29	112	159	114	25.4	429	572	167	1.8E-03
14	a	28	17	6	23	25	19	3.2	70	10	18	3.1E-05
	b	28	17	9	29	25	19	6.4	70	20	36	6.3E-05

Notes: -

1. All dimensions in mm or mm^2
2. A_L Assumes μ_R=100,000.
3. V_{MAX} for sinusoidal peak flux density of 0.5T, one turn at 1Hz

Appendix 2:Copper winding data

Imperial standard wire gauge (SWG)

No.	16	18	20	22	24	26	28
Diameter (mm)	1.626	1.219	0.914	0.711	0.559	0.457	0.376
$m\Omega m^{-1}$	8.3	14.8	26.0	43.5	70.5	105	155

No.	30	32	34	36	38	40
Diameter (mm)	0.315	0.274	0.234	0.193	0.152	0.122
$m\Omega m^{-1}$	222	293	404	590	950	1480

Metric wire data

Diameter (μm)	50	63	80	100	125	160	200	250	315	400	500
Ωm^{-1}	8.78	5.53	3.43	2.20	1.41	0.86	0.55	0.35	0.22	0.137	0.088

Two-stage IVDs and RTs

1. Two-stage IVDs

The most practical way to increase input impedance and accuracy of an IVD ratio winding (at low frequency) is to use an extra core with a separate energising winding. The energising winding is around one core, the ratio winding around both. The basic idea is that the energising winding carries most of the current and the energising core most of the flux. The voltage induced in the ratio winding opposes the applied voltage and thus "bootstraps" its input impedance. This is represented diagrammatically in fig. 1.1

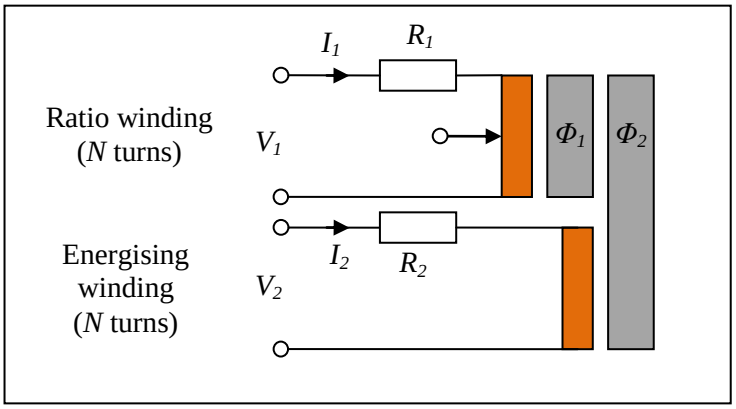


Fig. 1.1 A two-stage inductive voltage divider

I shall assume negligible flux leakage and a linear model for permeability.

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $\left(s = j\omega \equiv d/dt\right)$: -

$$V_1 = I_1R_1 + sN\left(\Phi_1 + \Phi_2\right) \qquad V_2 = I_2R_2 + sN\Phi_2$$

For ideal magnetic circuits the fluxes are: -

$$\Phi_1 = NI_1A_{L1} \qquad \Phi_2 = N\left(I_1 + I_2\right)A_{L2}$$

A_{L1} = permittance of core 1. A_{L2} = permittance of core 2.

From the above: $V_1 = I_1R_1 + sN^2\left(I_1A_{L1} + I_1A_{L2} + I_2A_{L2}\right)$ $V_2 = I_2R_2 + sN^2\left(I_1A_{L2} + I_2A_{L2}\right)$

Define winding inductances: $L_1 = N^2A_{L1}$ and $L_2 = N^2A_{L2}$

Then: $V_1 = I_1\left(R_1 + sL_1 + sL_2\right) + I_2sL_2$ and $V_2 = I_1sL_2 + I_2\left(R_2 + sL_2\right)$

In vector-matrix notation: $\mathbf{V} = \mathbf{ZI}$ or $\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} R_1 + sL_1 + sL_2 & sL_2 \\ sL_2 & R_2 + sL_2 \end{pmatrix}\begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$

The inverse of a 2×2 matrix is: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

The inverse is: $\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$ or $\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} R_2 + sL_2 & -sL_2 \\ -sL_2 & R_1 + sL_1 + sL_2 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$

Where $|\mathbf{Z}|$ is the determinant of the input impedance matrix: -

$$|\mathbf{Z}| = \begin{vmatrix} R_1 + sL_1 + sL_2 & sL_2 \\ sL_2 & R_2 + sL_2 \end{vmatrix} = R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2$$

If both inputs are connected to the same voltage source, V_{IN} then: -

$$I_1 = \frac{R_2}{|\mathbf{Z}|}V_{IN} \quad \text{and} \quad I_2 = \frac{R_1 + sL_1}{|\mathbf{Z}|}V_{IN}$$

The ratio of the currents is an interesting result: $$\frac{I_1}{I_2} = \frac{R_2}{R_1 + sL_1}$$

At a typical operating frequency the magnitude of the inductance is much greater than the winding resistances and the current in the ratio winding is, therefore, much smaller than the current in the energising winding.

Also, one can conclude that the flux in the energising core is much greater than the flux in the ratio core: $\Phi_2 >> \Phi_1$ This is advantageous – a low flux level ensures the highest permeability.

The input impedances are: -

Energising winding: $$Z_2 = \frac{V_{IN}}{I_2} = \frac{|\mathbf{Z}|}{R_1 + sL_1}$$

Ratio winding: $$Z_1 = \frac{V_{IN}}{I_1} = \frac{|\mathbf{Z}|}{R_2} = Z_2 \times \frac{R_1 + sL_1}{R_2}$$

At very low frequency (including DC) the result is just the winding resistances: -

$$|s| << \frac{R}{L} \quad \Rightarrow \quad Z_1 \approx R_1 \quad \text{and} \quad Z_2 \approx R_2$$

At high frequency the energising winding looks like a simple inductor but the ratio winding looks like a large (frequency dependent) negative resistor (increasing as frequency squared): -

$$|s| >> \frac{R}{L} \qquad \Rightarrow \qquad Z_2 \approx sL_2 \qquad \text{and} \qquad Z_1 \approx \frac{s^2L_1L_2}{R_2}$$

Without the energising winding one would expect $Z_1 \approx sL_1$ and so the impedance has been increased by a large "bootstrap" factor. Put simply: -

$$|s| >> \frac{R}{L} \quad \Rightarrow \quad Z_1 = sL_1 \times \frac{sL_2}{R_2}$$

As with single-stage inductors a general rule of thumb for the energising stage is to maximise the ratio of inductance to resistance [1].

1. Part 3, monograph 2: "Single-stage inductors and transformers". See sections 2 and 3.

The full analysis, in normalised form $\left(s = j\,\omega/\omega_N\right)$is: -

$$Z_1 = R_1\left(1 + as + s^2\right) \quad \text{and} \quad Z_2 = R_2\,\frac{1 + as + s^2}{1 + bs}$$

$$\text{With} \quad \omega_N = \sqrt{\frac{R_1 R_2}{L_1 L_2}} \quad a = \left(\frac{L_2}{R_2} + \frac{L_1}{R_1} + \frac{L_2}{R_1}\right)\omega_N \quad \text{and} \quad b = \sqrt{\frac{R_2 L_1}{R_1 L_2}}$$

Note that, as usual, the natural frequency is the geometric mean of the two characteristic frequencies.

To a very good approximation, retaining both in-phase and quadrature components, therefore: -

$$|s| >> \frac{R}{L} \Rightarrow \qquad Z_1 \approx R_1\left(as + s^2\right)$$

In more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \qquad Z_1 \approx R_1\left(ja\,\frac{f}{f_N} - \left(\frac{f}{f_N}\right)^2\right) \qquad \text{with} \qquad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}}$$

The phase of Z_1 is: $\theta = \arctan\left(\frac{\text{Im}}{\text{Re}}\right) = \arctan\left(-a\,\frac{f_N}{f}\right) \approx \pi - a\,\frac{f_N}{f}$ (radians).

I.e. just less than 180 degrees (predominantly negative real).

If one assumes that the inductances are the same (same size/permeability cores) but R_1 is twice R_2 (the most basic design - the top winding is longer because it is around both cores): -

$$R_1 = 2R_2 \Rightarrow \quad a = \frac{2R_2 + R_2 + R_2}{\sqrt{2R_2 R_2}} = 2\sqrt{2} \qquad b = \frac{1}{\sqrt{2}} \qquad \text{and} \qquad f_N = \frac{\sqrt{2}}{2\pi}\,\frac{R_2}{L}$$

With a phase angle: $\theta \approx \pi - 2\sqrt{2}\,\frac{f_N}{f}$ when expressed in radians.

Rule of thumb: A higher inductance to resistance ratio means a lower natural frequency and a lower phase error (imaginary component), compared to a pure (negative) resistance.

This model works well for an operating frequency up to a few hundred Hz. At higher frequency the interwinding capacitance becomes significant.

1.1 Example calculation

The following illustrates what is possible with 400 turns on a pair of medium sized SM100 cores for operation at 75Hz. This is fairly simple to construct by employing the BNNL winding technique [1].

Both cores type 7c (see appendix 1): -
OD = *60mm* ID = *36mm* axial = *15mm.*
The length of wire per turn on the energising winding is approximately: *15 × 2 + 60 – 36 = 54mm*

Inner circumference = 36mm × π = 113mm can accommodate 400 turns in two layers of 0.565mm OD wire.
Outer circumference = 60mm × π = 188mm can accommodate 400 turns of 0.47mm OD in a single layer.
The outer circumference is the limiting dimension: the largest wire diameter that is a comfortable fit is 26SWG (enamelled) at 0.457mm diameter with a resistance of 105mΩm^{-1}.

The energising winding requires: $400 \times 54mm \approx 22m$ of wire.
Resistance R_2 would be approx: $R_2 \approx 22m \times 105m\Omega m^{-1} \approx 2.3\Omega$
Permittance of both cores: $A_L = 102\mu H / turn^2$
Inductance of both windings: $L_1 = L_2 = N^2 A_L = 400 \times 400 \times 102 \times 10^{-6} \approx 16H$

The ratio winding consists of a rope with ten strands wound 40 times (also BNNL) of the same wire.
With 10 strands the rope is approximately 2mm in diameter. The inner diameter is reduced slightly: -
Inner circumference = 30mm × π = 94mm can accommodate 40 turns in a single layer (2.35mm/turn).
Outer circumference = 60mm × π = 188mm can easily accommodate 40 turns also in a single layer.

As a ball park estimate the length of wire, for the ratio winding, is roughly twice that of the energising winding, partly because it is wound around both cores and partly because the strands are twisted into a rope. If one assumes that the source resistance to both energising and ratio windings are negligible compared to the winding resistances the natural frequency is: -

$$R_1 \approx 2R_2 = 4.6\Omega \quad \Rightarrow \quad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}} = 0.032Hz$$

At *75Hz* the inductive reactance of the energising winding is: $X_L = 2\pi fL \approx 7.5k\Omega$ (inductive)

The bootstrapped input impedance of the ratio winding is: $Z_1 \approx \frac{s^2 L_1 L_2}{R_2} \approx -25M\Omega$ (negative resistive)

The result is a much smaller device which is easier to wind than a comparable single stage IVD. In practice it would be possible to experiment with thicker wire for the rope winding, in order to ensure a neat construction, with an evenly distributed double layer on the inner circumference, and even better results.

N.B. The two-stage IVD has been investigated thoroughly and it is stated (in the synopsis) that "Single decade dividers having in-phase errors of less than 5ppb of the input over the frequency band 40 – 400Hz have been constructed". The limiting factor is stated as being "due to inequalities in the core flux effectively linking each of the ten sections of the divider" [2].

Historically the two-stage IVD led to a number of bridge configurations but these are now obsolete thanks to active drive of the energising winding (see section 4) and the development of the three-stage ratio transformer [3].

1. Part 3, monograph 1: "IVDs and RTs – the basics". For balanced no-net-loop technique see section 1.3.
2. Hill, J. J. and Deacon, T.A. "Two stage inductive voltage dividers" Proc. IEE 1968 vol. 115 part 6, pp888 – 892.
3. Part 3, monograph 4: "Three-stage RTs"

With the inductance and resistance values above a spreadsheet model confirms the predicted behaviour (real component increasing second order and imaginary component first order with respect to frequency: -

$$f >> f_N \Rightarrow \qquad Z_1 \approx R_1\left(ja\frac{f}{f_N} - \left(\frac{f}{f_N}\right)^2 \right) \qquad \text{with} \qquad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}}$$

Fig. 1.1.1 Real (blue) and imaginary (green) components of the input impedance (Ohms on a log scale).

2. Two-stage transformers

The ratio winding rope can easily accommodate extra strands for use as an isolated secondary winding. The reduction of current in the ratio primary results in a significant improvement in ratio accuracy between primary and secondary. In the following analysis I shall assume that no current is flowing in the secondary and, therefore, the input impedances are the same as for the two-stage IVD.

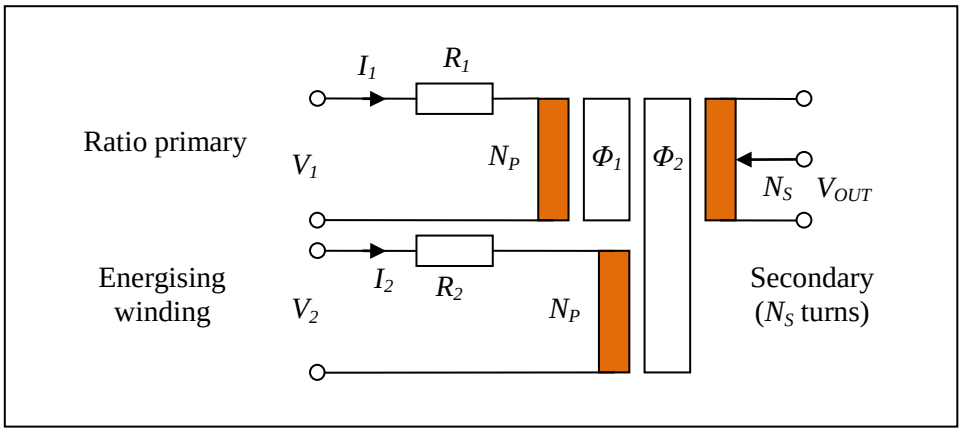


Fig. 2.1 A two-stage ratio transformer

According to Faraday's law the output voltage is: $V_{OUT} = sN_S(\Phi_1 + \Phi_2)$

I shall assume that both inputs are connected to the same voltage source, V_{IN}. As before the input voltage is: -

$$V_{IN} = I_1R_1 + sN_P(\Phi_1 + \Phi_2)$$

The transfer function is, therefore: $$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{sN_S(\Phi_1 + \Phi_2)}{I_1R_1 + sN_P(\Phi_1 + \Phi_2)}$$

Eliminate the fluxes with: $\Phi_1 = N_P I_1 A_{L1}$ $\Phi_2 = N_P(I_1 + I_2)A_{L2}$

$$\Rightarrow \quad T(s) = \frac{sN_SN_P(I_1A_{L1} + I_1A_{L2} + I_2A_{L2})}{R_1 + sN_P^2(I_1A_{L1} + I_1A_{L2} + I_2A_{L2})}$$

From section 1: $$\frac{I_2}{I_1} = \frac{R_1 + sL_1}{R_2}$$

Also, by definition, the inductances are: $L_1 = N_P^2A_{L1}$ and $L_2 = N_P^2A_{L2}$

With a little algebra (divide top and bottom by I_1 and multiply by R_2) the transfer function is: -

$$\Rightarrow \quad T(s) = \frac{N_S}{N_P}\left(\frac{s(R_2L_1 + R_2L_2 + R_1L_2) + s^2L_1L_2}{R_1R_2 + s(R_2L_1 + R_2L_2 + R_1L_2) + s^2L_1L_2}\right)$$

In normalised form $(s = j\omega/\omega_N)$:

$$\frac{V_{OUT}}{V_{IN}} = \frac{N_S}{N_P}\left(\frac{as + s^2}{1 + as + s^2}\right)$$

With $\omega_N = \sqrt{\frac{R_1 R_2}{L_1 L_2}}$ and $a = \left(\frac{L_1}{R_1} + \frac{L_2}{R_1} + \frac{L_2}{R_2}\right)\omega_N$

Note, again, the natural frequency is the geometric mean of the two characteristic frequencies.

Apart from the turns ratio this is exactly the same transfer function I derived for a two stage high-pass filter [1]. One can use the same approximation at high frequency: -

$$\omega >> \omega_N \Rightarrow \qquad T(s) \approx 1 - \frac{1}{s^2} + \frac{a}{s^3}$$

In a more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \qquad \text{with} \quad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}}$$

As with a basic (single-stage) transformer the in-phase error is second order and usually too large for the very highest accuracy applications. The quadrature error, however, is now third order and very small at the operating frequency. A two-stage transformer could be used, therefore, where only low phase error is required. (e.g. as a step-down injector as part of a quadrature null balance system). They are also useful in the lower (less critical) decades of a multi-decade bridge. At higher frequency two-stage transformers also prove useful for multi-decade capacitance bridges by allowing a reduction in the required number of turns and a corresponding reduction in inter-winding capacitance.

2.1 Example calculation

As with the previous example calculation (section 1.1) I shall assume the most basic design (same size/permeability cores), R_1 is twice R_2 (the ratio winding is longer because it is around both cores and twisted into a rope) and both primary windings are connected to the same low resistance source: -

$$R_1 = 2R_2 \text{ and } L_1 = L_2 \quad \Rightarrow \quad a = \left(\frac{L_1}{R_1} + \frac{L_2}{R_1} + \frac{L_2}{R_2}\right)\omega_N = \frac{R_2 L_1 + R_2 L_2 + R_1 L_2}{\sqrt{R_1 R_2 L_1 L_2}} = 2\sqrt{2} = 2.8$$

The result is a small peak near the natural frequency. Also: $f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}} = \frac{\sqrt{2}}{2}\frac{R_2}{L_2}$

With a pair of 7c SM100 cores with 400 turns of 26SWG wire for the windings: -

$$L_1 = L_2 \approx 16H \qquad R_2 \approx 2.3\Omega \qquad f_N = 0.032Hz$$

$$\text{at } 75Hz \;\Rightarrow\; T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 = 1 + 1.8\times10^{-7} + j2.2\times10^{-10}$$

With up to 0.2ppm the in-phase error this is not quite good enough for some applications but the phase error is negligible. For lower in-phase error see [2].

1. Part 2, monograph 1: "Two-stage filters". See section 4.
2. Part 3, monograph 4: "Three-stage RTs".

3. The matrix method and the equivalent circuit

If one compares the impedance matrix of the two-stage IVD (page 1) and the two-stage RC low-pass filter one observes that replacing the capacitors with inductors results in the required transfer function [1]. For example: -

For the low-pass RC filter: $$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} R_1 + X_1 + X_2 & X_2 \\ X_2 & R_2 + X_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \qquad X_1 = \frac{1}{sC_1} \text{ etc}$$

From page 1: $$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} R_1 + sL_1 + sL_2 & sL_2 \\ sL_2 & R_2 + sL_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \qquad X_1 = sL_1 \text{ etc}$$

This suggests the equivalent circuit: -

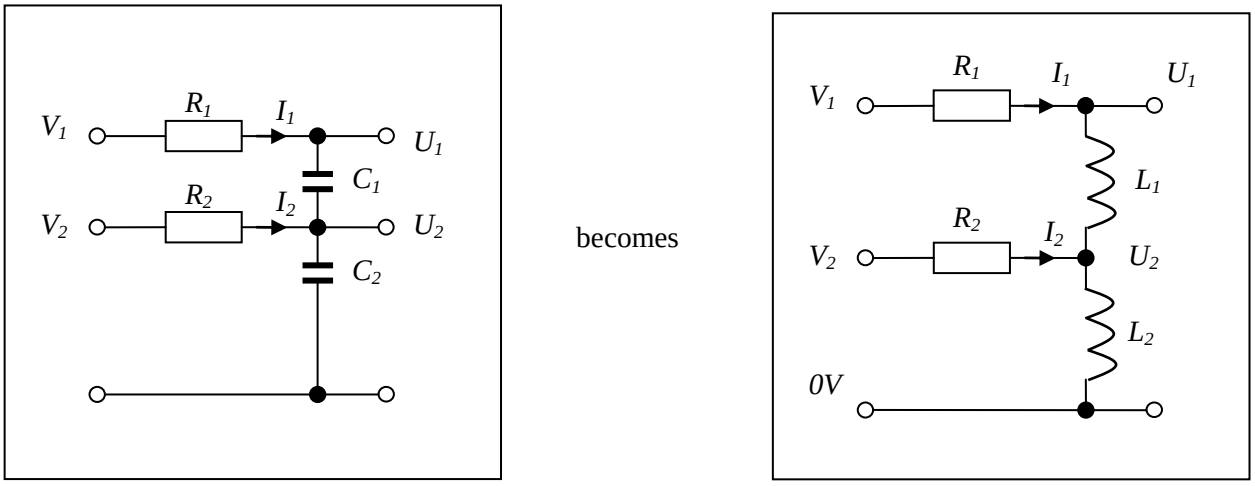


Fig. 3.1 Equivalent circuit for a two-stage voltage transformer

The bootstrap effect of the energising core is just the same as for two-stage filters. This proves useful for modelling multi-stage transformers, especially when it comes to stability analysis of active drives. One can also re-use the matrix analysis of high accuracy filters. According to Ohm's and Kirchhoff's Laws the output voltages are: -

$$U_2 = sL_2(I_1 + I_2) \qquad U_1 = sL_2(I_1 + I_2) + sL_1 I_1$$

I shall again employ the general symbols $(X_1 = sL_1)$ etc so that: -

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 + X_2 & X_2 \\ X_2 & X_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \text{ i.e. } \mathbf{U = XI}$$

This defines the output impedance matrix $\mathbf{X}$.

The input impedance matrix equation can also be constructed by noting: -

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} + \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} R_1 + X_1 + X_2 & X_2 \\ X_2 & R_2 + X_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \qquad \text{i.e. } \mathbf{V = ZI}$$

This defines the input impedance matrix $\mathbf{Z}$.

1. Part 2, monograph 1: "Two-stage filters". See section 1.

The inverse is often useful (for calculating input currents and impedances): -

$$\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$$

The inverse of a 2×2 matrix is: $$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\mathbf{Z}^{-1} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} R_2 + X_2 & -X_2 \\ -X_2 & R_1 + X_1 + X_2 \end{pmatrix}$$

$|\mathbf{Z}|$ is the determinant: $$|\mathbf{Z}| = \begin{vmatrix} R_1 + X_1 + X_2 & X_2 \\ X_2 & R_2 + X_2 \end{vmatrix} = R_1R_2 + R_1X_2 + X_1R_2 + X_1X_2 + R_2X_2$$

Combine with the output impedance matrix equation for the matrix equivalent of the transfer function equation (outputs versus inputs): -

$$\mathbf{U} = \mathbf{XI} = \mathbf{XZ}^{-1}\mathbf{V} = \mathbf{TV}$$

Where $\mathbf{T}$ is the transfer function matrix: $\mathbf{T} = \mathbf{XZ}^{-1}$

So that: $$\mathbf{T} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} X_1 + X_2 & X_2 \\ X_2 & X_2 \end{pmatrix}\begin{pmatrix} R_2 + X_2 & -X_2 \\ -X_2 & R_1 + X_1 + X_2 \end{pmatrix}$$

Multiply the matrices: -

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} X_1R_2 + X_2R_2 + X_1X_2 & X_2R_1 \\ X_2R_2 & X_2R_1 + X_2X_1 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

This defines the matrix $\mathbf{\Omega}$ so that: $\mathbf{\Omega} = |\mathbf{Z}|\mathbf{T} = |\mathbf{Z}|\mathbf{XZ}^{-1}$

Replace the general symbols $(X_1 = sL_1)$ etc so that: -

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} s(L_1 + L_2)R_2 + s^2L_1L_2 & sL_2R_1 \\ sL_2R_2 & sL_2R_1 + s^2L_1L_2 \end{pmatrix}$$

To find any of the output voltages, in terms of the inputs, it is now a simple case of calculating only the required components of $\mathbf{\Omega}$ and the determinant $|\mathbf{Z}|$. Specifically: -

$$U_1 = \frac{\Omega_{11}V_1 + \Omega_{12}V_2}{|\mathbf{Z}|} \quad \text{and} \quad U_2 = \frac{\Omega_{21}V_1 + \Omega_{22}V_2}{|\mathbf{Z}|}$$

A spreadsheet with sliders for the main parameters and graphs for the frequency response and error terms (in-phase and quadrature) is ideally suited to computer modelling. The calculations are reduced to bite-size chunks (e.g. matrix elements) using complex number functions (IMSUM, IMPRODUCT, IMDIVIDE etc.) and then combined.

4. Two-stage active drive and stability

4.1 Active drive for both primary windings

With two-stage IVDs and ratio transformers both ratio and energising windings can be driven with, for example, high accuracy voltage followers, with no stability problems. See fig. 4.1.1. Note the order of the negative feedback connection – the larger current flowing through the energising winding should not flow through any part of the connection to the ratio winding (i.e. it is a zero-Ohm junction, type 1 [1]). The followers need very low DC offset in order to avoid DC current flowing through the (low resistance) windings.

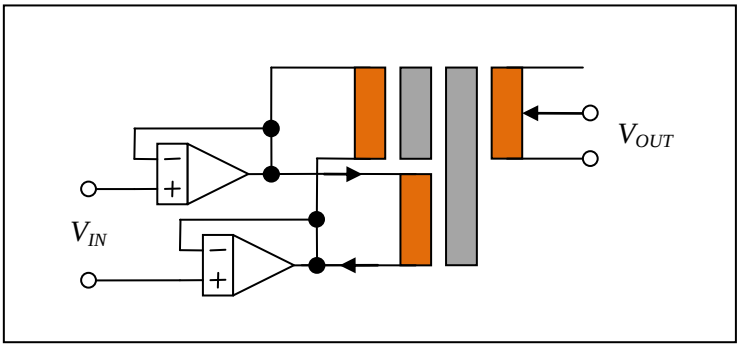


Fig. 4.1.1 A two-stage ratio transformer with high accuracy followers

The main advantage is the very high input impedance and extremely low output impedance of the followers. The main disadvantage is extra noise due to the followers.

In this case both inputs are connected to the same low resistance source and the output (ratio 1:1) is: -

$$V_{OUT} = U_1 = \frac{(\Omega_{11}V_1 + \Omega_{12}V_2)}{|\mathbf{Z}|} = \frac{(\Omega_{11} + \Omega_{12})}{|\mathbf{Z}|}V_{IN}$$

$$|\mathbf{Z}| = R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2$$

The transfer function required is, therefore: $$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{R_2sL_1 + R_2sL_2 + sL_1sL_2 + R_1sL_2}{R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2}$$

Reorganise into powers of *s*: $$T(s) = \frac{s((L_1 + L_2)R_2 + R_1L_2) + s^2L_1L_2}{R_1R_2 + s((L_1 + L_2)R_2 + R_1L_2) + s^2L_1L_2}$$

Divide top and bottom by R_1R_2 and express in normalised form $(s = j\omega/\omega_N)$. This agrees with the previous result (pages 5 and 6) as long as one re-introduces the turns ratio: -

$$T(s) = \frac{N_S}{N_P}\left(\frac{as + s^2}{1 + as + s^2}\right)$$

With $\omega_N = \sqrt{\frac{R_1R_2}{L_1L_2}}$ and $a = \left(\frac{L_1}{R_1} + \frac{L_2}{R_1} + \frac{L_2}{R_2}\right)\omega_N$

The error analysis is the same as section 2.

1. Part 1, monograph 1: “High accuracy resistors”. See section 3.4.

4.2 Active drive for the energising winding only [1]

The problem of voltage follower voltage noise can largely be avoided by driving only the energising winding. The follower current noise, however, flows through the source resistance and best performance is achieved when the follower noise resistance is substantially greater than the source resistance. The main advantages of two-stage construction (increase in input impedance and low phase error) are retained.

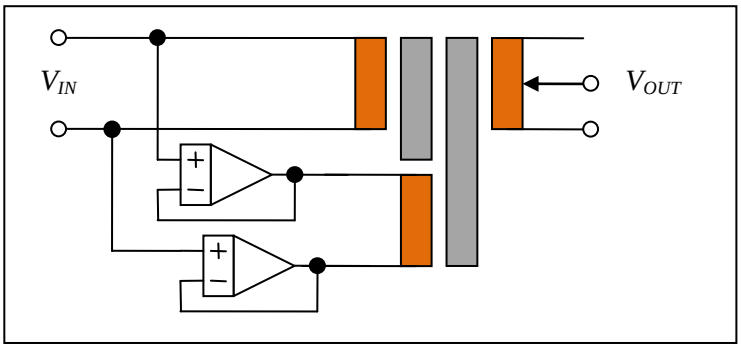


Fig. 4.2.1 Active drive of the energising winding only

The equivalent circuit again proves useful as a model (see section 3). If one assumes a transformer ratio setting of 1:1 the equivalent circuit is an actively driven two-stage high-pass filter: -

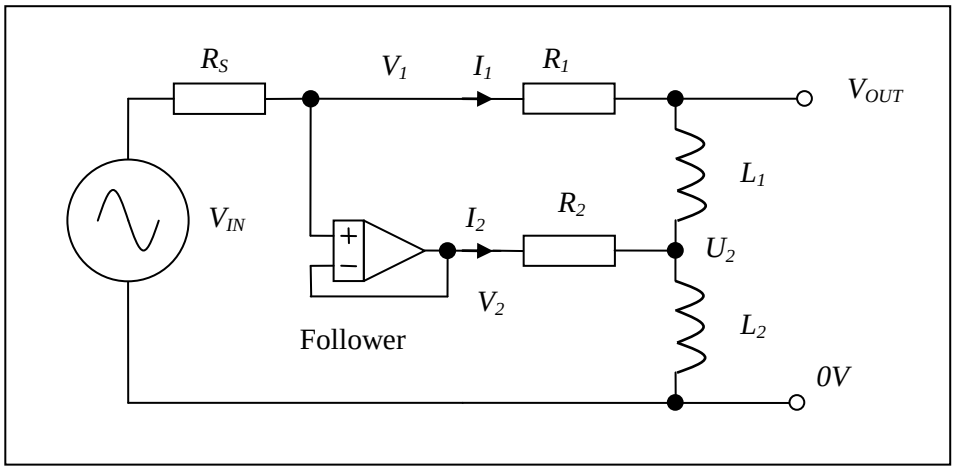


Fig. 4.2.2 Active drive of the energising winding only – equivalent circuit

The only problem with this approach is the peak in the frequency response at around the natural frequency due to positive feedback. The size of the peak depends mainly on the source resistance.

If the voltage source is set to zero the feedback network consists of the energising winding (high-pass characteristic) followed by the ratio winding and source resistance (low-pass characteristic). See fig. 4.2.3.

1. Gibbins, D. L. H.: "A circuit for reducing the exciting current of inductive devices". Proc. IEE vol. 108B (1961) pp 339-343.
Gibbins was an early contributor to the development of active drive of two-stage IVDs though he did not seem to understand the problem of (low frequency) feedback and stability. In his paper, cited by many other experts, the first sentence begins "A negative feedback device is described…" Later, however, he does identify "The technical difficulties of the device, regarded as an exercise in feedback-amplifier design, reside in the fact that the external source impedance, which is not under the designer's control, appears in the feedback loop between two transformers (fig. 2)". The diagram to which he refers should have provided a strong clue – the feedback is positive! He achieved an increase in input impedance by a very useful factor of 130 at 52.5Hz but there is no mention of stability margin. He does anticipate, however, the possibility of accurate three-stage ratio transformers but finally admits: "Whether it is possible to achieve a higher overall multiplication by this means before the feedback problems become insuperable has not been investigated."

The result is positive feedback with a band-pass characteristic whose peak magnitude can approach unity, depending on the relative cut-off frequencies: -

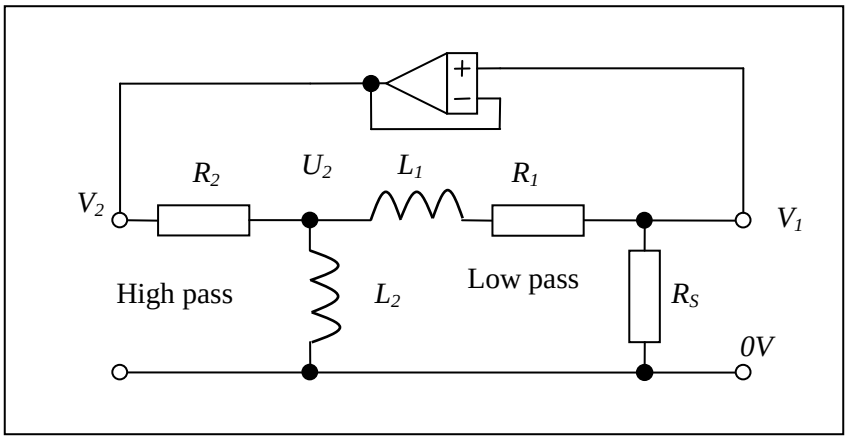


Fig. 4.2.3 The feedback network (input voltage $V_{IN}=0$)

For adequate stability margin the high-pass cut-off frequency needs to be about the same or higher than the low-pass cut-off: -

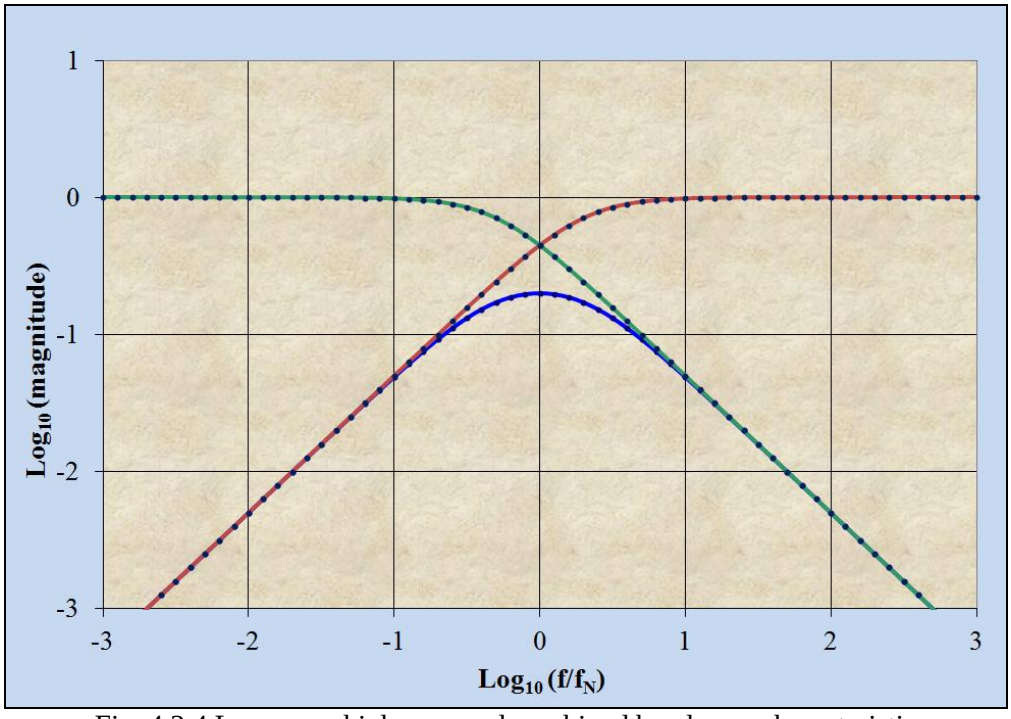


Fig. 4.2.4 Low-pass, high-pass and combined band-pass characteristic

Apart from the interaction between the high pass and low pass sections the cut-off frequencies are: -

$$\omega_{HP} = \frac{R_2}{L_2} \qquad \omega_{LP} = \frac{R_S + R_1}{L_1} \qquad \text{required: } \omega_{HP} \geq \omega_{LP} \quad \Rightarrow \quad \frac{R_2}{L_2} \geq \frac{R_S + R_1}{L_1}$$

With a high source resistance (R_S > 10Ω) it would seem that one needs to increase R_2 with an extra component resistor and/or increase L_1 substantially relative to L_2. Fortunately there is an ingenious solution which makes it possible to retain a low natural frequency (high accuracy) as well as a good stability margin: a "compensator" - see later. This proves particularly necessary for three-stage ratio transformers [1] where the source resistance can be up to 100Ω. For higher source resistance (≈1kΩ) see the monograph "Noise matching transformers" [2].

1. Part 3, monograph 4: "Three-stage RTs". See section 3.
2. Part 3, monograph 5: "Noise matching transformers". See section 3.

4.2.1 Network analysis

The matrix equations for the transformer network are repeated for convenience (see section 3): -

$$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} R_2 + X_2 & -X_2 \\ -X_2 & R_1 + X_1 + X_2 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \quad \text{with} \quad X_1 = sL_1 \text{ etc.}$$

$$|\mathbf{Z}| = R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2$$

If one assumes an ideal voltage follower ($V_2 = V_1$) then I_1 simplifies to: -

$$I_1 = \frac{R_2}{|\mathbf{Z}|}V_1 = \frac{R_2}{R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2}V_1$$

But $V_{IN} = V_1 + I_1R_S = V_1\left(1 + \frac{R_2R_S}{|\mathbf{Z}|}\right)$ so that $V_1 = V_2 = \frac{|\mathbf{Z}|}{|\mathbf{Z}| + R_2R_S}V_{IN}$

From above: $V_{OUT} = \frac{(\Omega_{11}V_1 + \Omega_{12}V_2)}{|\mathbf{Z}|} \Rightarrow V_{OUT} = \frac{(\Omega_{11} + \Omega_{12})}{|\mathbf{Z}|} \times \frac{|\mathbf{Z}|}{|\mathbf{Z}| + R_2R_S}V_{IN}$

Resulting in: $V_{OUT} = \frac{\Omega_{11} + \Omega_{12}}{|\mathbf{Z}| + R_2R_S}V_{IN}$

From above: $\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} s(L_1 + L_2)R_2 + s^2L_1L_2 & sL_2R_1 \\ sL_2R_2 & sL_2R_1 + s^2L_1L_2 \end{pmatrix}$

Also from above: $|\mathbf{Z}| = R_1R_2 + R_1X_2 + X_1R_2 + X_1X_2 + R_2X_2$

The closed loop transfer function for the equivalent circuit is, therefore: -

$$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{s((L_1 + L_2)R_2 + R_1L_2) + s^2L_1L_2}{R_2(R_S + R_1) + s((L_1 + L_2)R_2 + R_1L_2) + s^2L_1L_2}$$

In normalised form $(s = j\omega/\omega_N)$: $T(s) = \frac{as + s^2}{1 + as + s^2}$

With $\omega_N = \sqrt{\frac{R_2(R_S + R_1)}{L_1L_2}}$ and $a = \left(\frac{L_1 + L_2}{R_S + R_1} + \frac{R_1L_2}{R_2(R_S + R_1)}\right)\omega_N$

The result is similar to the previous case (see pages 5 and 6) with two important differences. The resistance R_1 is replaced by $R_1 + R_S$ in the formula for the natural frequency and in the denominator for parameters *a* and *b*.

Increasing source resistance results, therefore, in increasing natural frequency (lower accuracy) and decreasing values of *a* and *b* (less stability margin).

4.2.2 Stability analysis (approximate)

This is the same transfer function as the two-stage high pass filter with the same trade-off [1]. A value of the coefficient *a* between 2 and 4 results in peaking which is tolerable in most applications. A lower value results in a higher resonant peak. If the inductances are the same, R_1 is twice R_2 (see example calculation page 4) and if one assumes the minimum value of $a = 2$: -

$$a = 4\sqrt{\frac{R_2}{R_S + 2R_2}} \approx 2 \quad \text{then the maximum source resistance is:} \quad R_S \approx 2R_2$$

Since a typical energising winding is ≈ 2Ω (or lower) this is only possible for some applications (e.g. for an IVD based system with low source resistance). It would not be possible, for example, to connect the ratio winding directly to a standard resistor or sensor or via a semiconductor switch with series resistance greater than 4Ω.

Increasing source resistance increases the cut-off frequency of the low pass section of the feedback network, increasing the overlap with the high pass characteristic. Stability could be restored with a higher value for R_2 thus increasing the frequency of the high pass characteristic and reducing the overlap. Unfortunately this also increases the natural frequency of the overall transfer function and reduces accuracy. Reminder: from above, the in-phase error is only second order with respect to frequency: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \qquad \text{with} \qquad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}}$$

If, for example, one adds a component resistor in series with R_2 so that the sum is the same value as the maximum expected source resistance: -

$$R_2 \to R = R_S \Rightarrow \qquad a = 4\sqrt{\frac{R}{R_S + 2R}} = \frac{4}{\sqrt{3}} \qquad \text{and} \qquad R_S = 0 \Rightarrow \qquad a = 4\sqrt{\frac{R}{0 + 2R}} = \frac{4}{\sqrt{2}}$$

$$\text{or} \qquad 2.3 \le a \le 2.8$$

The result is reasonably stable though the increase in natural frequency could be a problem.

Rule of thumb: To improve stability add extra resistance in series with the energising winding. This should be approximately the same as the source resistance.

It is also possible to add extra turns and/or select a larger, higher permeability core to increase the inductance L_1 but the latter improvement is modest (with much more expensive RM200 cores the inductance would increase only by a factor of two). Doubling the number of turns (on the top core only), on the other hand, increases the inductance, L_1, by a factor of four (see section 4.4). Fortunately there is also a simple way of reducing the overall natural frequency: a "compensator" (see section 4.3). This largely restores the bootstrap effect at the operating frequency but retains stability at lower frequency.

Rule of thumb: To improve stability employ a thicker (axially) and higher permeability top core and add extra turns to the ratio winding (around the top core only).

Rule of thumb: To improve accuracy add a compensator in series with the energising winding.

1. Part 2, monograph 1: "Two-stage filters". See section 3.

4.2.3 Example calculation

This example illustrates the problem of source resistance with the most basic two-stage IVD or RT with active drive for the energising winding. It employs the same basic transformer design (size 7c SM100 cores with 400 turns. See section 1.1): -

Inductance of both windings: $L_1 = L_2 \approx 16H$

Ratio (primary) winding resistance: $R_1 \approx 4.6\Omega$

Energising winding resistance: $R_2 \approx 2.3\Omega$

With three values of source resistance: $R_S \approx 1\Omega$, $R_S \approx 10\Omega$ and $R_S \approx 100\Omega$

The transfer function is, in normalised form $(s = j\omega/\omega_N)$: $T(s) = \dfrac{as + s^2}{1 + as + s^2}$

With $\omega_N = \sqrt{\dfrac{R_2(R_S + R_1)}{L_1 L_2}}$ and $a = \left(\dfrac{L_1 + L_2}{R_S + R_1} + \dfrac{R_1 L_2}{R_2(R_S + R_1)}\right)\omega_N$

The respective natural frequencies (in radians/s) are: $\omega_N \approx 0.224$ $\omega_N \approx 0.362$ and $\omega_N \approx 0.97$

The normalised parameters are: $a \approx 2.6$ (lowest peak) $a \approx 1.6$ and $a \approx 0.59$ (highest peak)

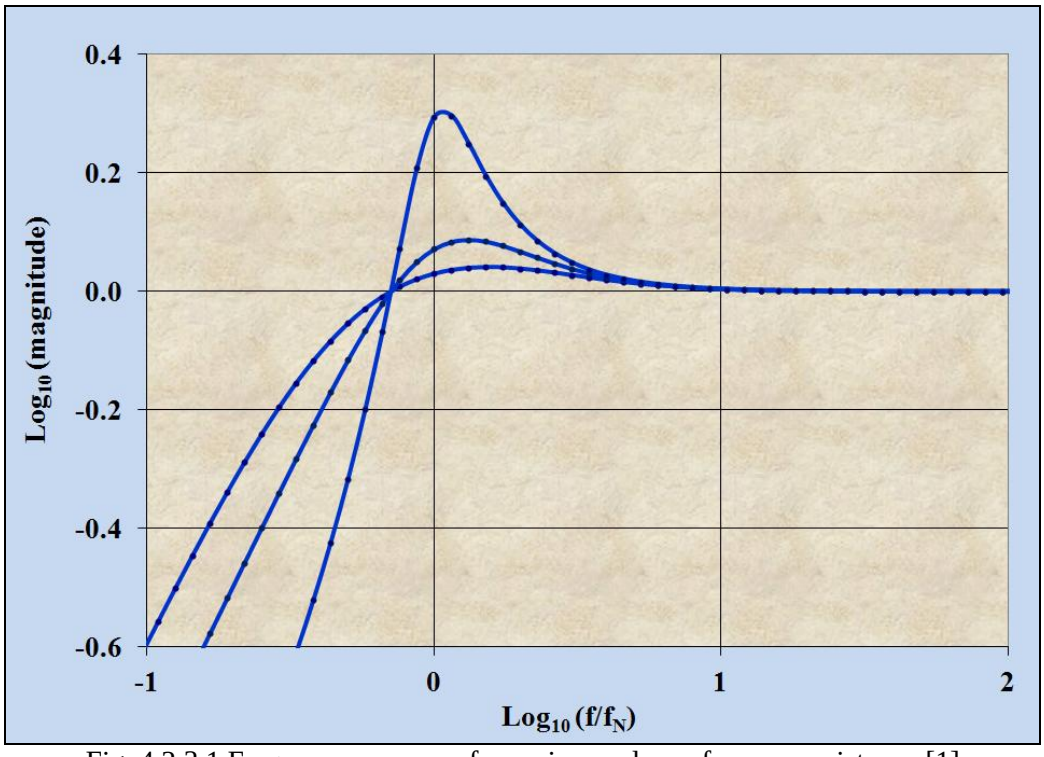


Fig. 4.2.3.1 Frequency response for various values of source resistance [1]

Clearly the highest peak corresponds to the highest value of source resistance (100Ω) and is borderline stable for most applications. The results for 1Ω and 10Ω are tolerable.

The conclusion is that the most basic two-stage IVD or RT needs a low source resistance (R_S < 10Ω). For applications with a higher source resistance see sections 4.3 and 4.4.

1. Spreadsheet: "A two-stage ratio transformer".

4.2.4 Errors in the energising follower

The following analysis investigates the effect on performance due to any error or noise of the follower (driving only the energising winding). Define parameter $\delta << 1$ to describe the difference between the voltage applied to the ratio winding and the voltage applied to the energising winding, such that: -

$$V_2 = (1+\delta)V_1$$

N.B. δ can be a complex number to represent in-phase and/or quadrature errors or a random number to represent noise. From the matrix analysis above: -

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} s(L_1+L_2)R_2 + s^2 L_1 L_2 & sL_2R_1 \\ sL_2R_2 & sL_2R_1 + s^2 L_1 L_2 \end{pmatrix}$$

$$V_{OUT} = U_1 = \frac{(\Omega_{11}V_1 + \Omega_{12}V_2)}{|\mathbf{Z}|} = \frac{(\Omega_{11} + \Omega_{12}(1+\delta))}{|\mathbf{Z}|}V_{IN}$$

$$|\mathbf{Z}| = R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2$$

The transfer function required is, therefore: -

$$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{R_2sL_1 + R_2sL_2 + sL_1sL_2 + R_1sL_2(1+\delta)}{R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2}$$

Divide top and bottom by R_1R_2 and multiply each term by the natural frequency to the appropriate power so that, in normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{(a+\varepsilon)s + s^2}{1 + as + s^2} = \frac{as + s^2}{1 + as + s^2} + \frac{\varepsilon s}{1 + as + s^2}$$

With $\omega_N = \sqrt{\frac{R_1R_2}{L_1L_2}}$ $\quad a = \left(\frac{L_1}{R_1} + \frac{L_2}{R_2} + \frac{L_2}{R_2}\right)\omega_N$ and $\quad \varepsilon = \frac{L_2}{R_2}\delta\omega_N = \delta\sqrt{\frac{L_2R_1}{R_2L_1}}$

At high frequency the extra term in the transfer function represents a small additional error: -

$$|s| >> 1 \;\Rightarrow\; \frac{\varepsilon s}{1 + as + s^2} \approx \frac{\varepsilon}{s}$$

The result is an extra first order error (in terms of frequency) but since δ can be made very small it is often negligible at the operating frequency. From the previous error analysis: -

$$|s| >> 1 \;\Rightarrow\; T(s) \approx 1 + \frac{\varepsilon}{s} - \frac{1}{s^2} + \frac{a}{s^3}$$

In a more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 - j\varepsilon\left(\frac{f_N}{f}\right) + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \qquad \text{with} \qquad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1R_2}{L_1L_2}}$$

N.B. Remember that ε can be imaginary so that the extra error could be in-phase. With a little care the follower errors can be made less than 1ppm and so the overall effect, at the operating frequency, is usually negligible.

If one follows the previous example (section 1.1) where the inductances are the same and R_1 is twice R_2 a typical value would be: -

$$\varepsilon = \delta\sqrt{\frac{L_2 R_1}{R_2 L_1}} \quad \Rightarrow \qquad \varepsilon \approx \delta\sqrt{2}$$

For a single op-amp with a gain-bandwidth product f_B the follower error is mainly quadrature and proportional to frequency [1]: -

$$\delta = -j\frac{f}{f_B} \qquad \Rightarrow \qquad -j\varepsilon\left(\frac{f_N}{f}\right) = \sqrt{2}\,\frac{f_N}{f_B}$$

With a transformer natural frequency less than 1Hz and a gain-bandwidth product of 5MHz the error contribution is less than 1ppm. A two stage follower would result in a truly negligible contribution.

4.2.5 Follower noise

If δ is random (i.e. noise due to the follower) then the extra noise appearing at the output is also significantly reduced: -

$$\delta = \frac{V_{NOISE}}{V_{IN}} \qquad \Rightarrow \qquad V_{OUT} = T(s)V_{IN} \approx \left(1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3\right)V_{IN} - j\varepsilon\left(\frac{f_N}{f}\right)$$

In the limit that the input signal, V_{IN}, tends to zero (noise remaining): -

$$\delta = \frac{V_{NOISE}}{V_{IN}} \quad \text{and} \quad \varepsilon \approx \delta\sqrt{2} \quad \Rightarrow \quad V_{OUT} = \sqrt{2}\left(\frac{f_N}{f}\right)V_{NOISE}$$

Where V_{NOISE} is a random number representing the follower noise. The contribution due to the active drive is thus much reduced. The follower current noise, however, flows through the source resistance and appears across the ratio primary. **Best performance is achieved, therefore, when the noise resistance of the followers is significantly greater than the source resistance.**

4.2.6 The effect on input impedance

From section 1:
$$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} R_2 + sL_2 & -sL_2 \\ -sL_2 & R_1 + sL_1 + sL_2 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \quad \text{with} \quad V_2 = (1+\delta)V_1$$

$$\Rightarrow \quad I_1 = \frac{V_1}{|\mathbf{Z}|}\left(R_2 + sL_2 - sL_2(1+\delta)\right) = \frac{V_1}{|\mathbf{Z}|}\left(R_2 - \delta sL_2\right)$$

N.B. Unfortunately the value of L_2 is not well defined (only the minimum is guaranteed) and so it is not possible to exploit this property to reduce the input current to zero. To ensure that the ratio winding input impedance remains high, therefore, one must rely entirely on the bootstrapping effect. The input impedance is, therefore: -

$$Z_1 = \frac{V_1}{I_1} = \frac{|\mathbf{Z}|}{R_2 - \delta sL_2} \quad \text{and, to a good approximation} \quad Z_1 \approx \frac{|\mathbf{Z}|}{R_2}\left(1 + \delta\frac{sL_2}{R_2}\right)$$

For the effect to be negligible one needs to ensure: $|\delta| << \left|\frac{R_2}{sL_2}\right|$

As above this is easily achieved in practice.

1. Part 4, monograph 1: “High gain blocks”. See section 3: “Some operational amplifier circuits”.

4.3 Active drive with compensator (resistor/parallel capacitor)

Stability margin can be improved by adding a compensator in series with the energising winding. The capacitor is chosen to have little effect at low frequency but bypasses the extra resistor at the higher (operating) frequency. This substantially restores the bootstrap effect and ensures high input impedance and low phase error.

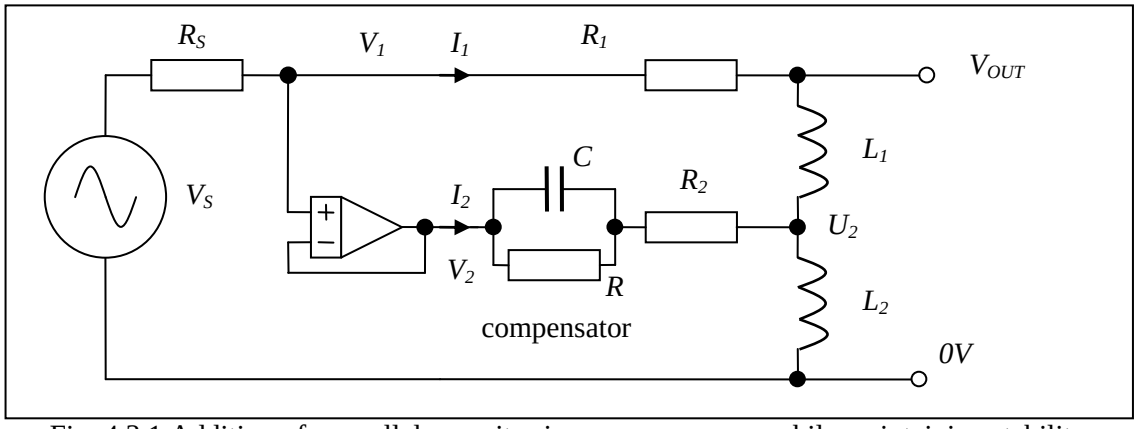


Fig. 4.3.1 Addition of a parallel capacitor improves accuracy while maintaining stability

From above, repeated for convenience, the transfer function (without compensating *R* and *C*) is: -

$$T(s)=\frac{s(R_2L_1+R_2L_2+R_1L_2)+s^2L_1L_2}{R_2(R_S+R_1)+s(R_2L_1+R_2L_2+R_1L_2)+s^2L_1L_2}$$

The new transfer function can be obtained by replacing the symbol for resistance, R_2, with Z, the symbol for the total series impedance including compensator: -

$$R_2 \rightarrow Z = R_2 + \frac{R}{1+sRC}$$

The result is a third order transfer function (see appendix 2 for details): -

$$T(s)=\frac{((R+R_2)(L_1+L_2)+R_1L_2)s+(R_2RC(L_1+L_2)+R_1L_2RC+L_1L_2)s^2+L_1L_2RCs^3}{(R+R_2)(R_S+R_1)+[R_2RC(R_S+R_1)]s+((R+R_2)(L_1+L_2)+R_1L_2)s+(R_2RC(L_1+L_2)+R_1L_2RC+L_1L_2)s^2+L_1L_2RCs^3}$$

In normalised form $(s = j\omega/\omega_N)$, in which I anticipate that δ can be made small compared to *a*: -

$$T(s)=\frac{as+bs^2+s^3}{1+(a+\delta)s+bs^2+s^3}$$

$$\omega_N=\left(\frac{(R+R_2)(R_S+R_1)}{RCL_1L_2}\right)^{\frac{1}{3}} \qquad a=\frac{(R+R_2)(L_1+L_2)+R_1L_2}{(R+R_2)(R_S+R_1)}\omega_N$$

$$b=\frac{(R_2RC(L_1+L_2)+R_1L_2RC+L_1L_2)}{(R+R_2)(R_S+R_1)}\omega_N^2 \qquad \delta=\frac{R_2RC}{(R+R_2)}\omega_N$$

Once again the natural frequency is the geometric mean of three characteristic frequencies - this time including the *RC* time constant $(RC \neq 0)$. The transfer function is (very nearly) of the same form as a three-stage high-pass filter for which the analysis suggests ideal values for *a* and *b* between 2 and 4 [1]. For a computer model see [2].

1. Part 2, monograph 2: "Three-stage filters". See section 3.4.
2. Spreadsheet: "A two-stage ratio transformer with compensator"

A typical compensator consists of a large capacitor (1-100mF) in parallel with a small resistor (1 -100Ω). A typical design has: -

$$R = 10\Omega \quad \text{and} \quad C = 10mF$$

The extra 2.3Ω is due to the energising winding resistance: -

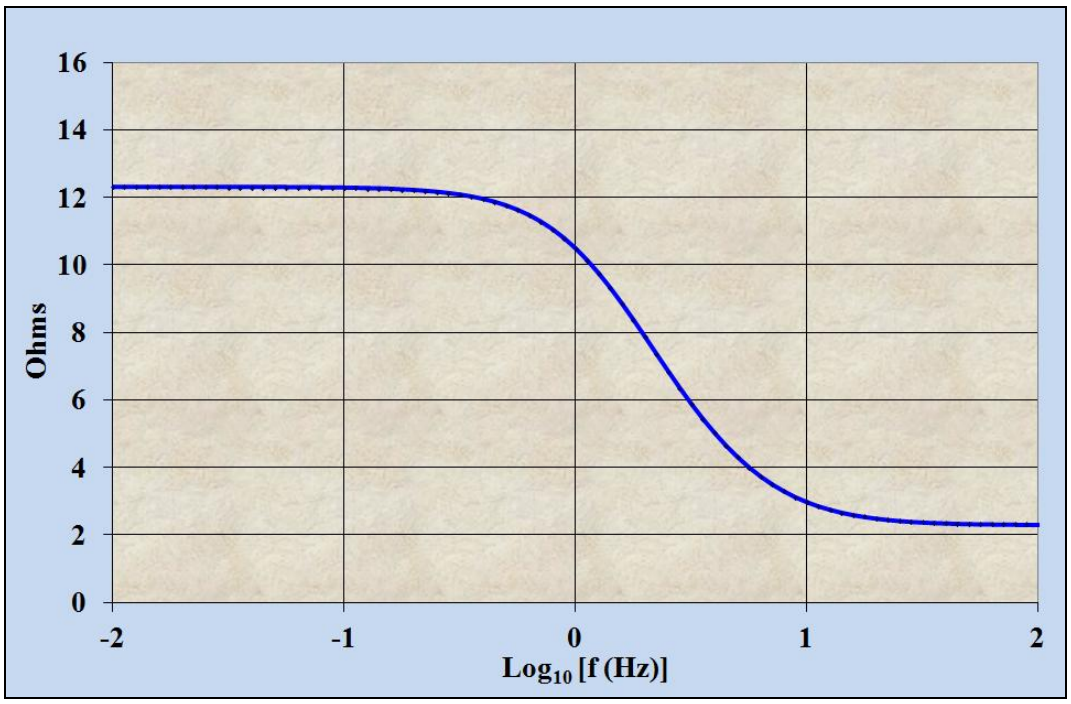


Fig. 4.3.2 Impedance (magnitude) of a typical compensator versus frequency

This is a rather large capacitor. A pair of back-back aluminium electrolytic type capacitors is not a realistic option. Fortunately, there is a practical solution: a simulated large capacitor circuit [1]: -

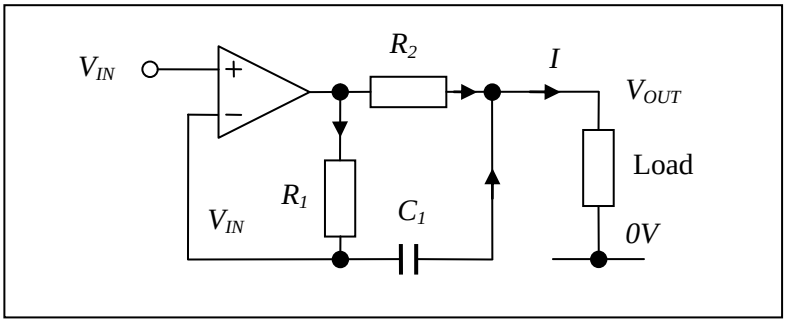


Fig. 4.3.3 A simulated large capacitor circuit [1]

Strictly speaking it is not just a large capacitor. The equivalent circuit also has the parallel resistance: -

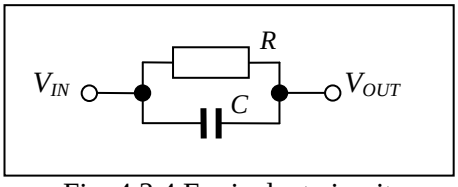


Fig. 4.3.4 Equivalent circuit

1. Part 6, monograph 1: “A simulated large capacitor circuit”.

4.3.1 Error analysis with compensator

In normalised form, repeated for convenience: $T(s)=\dfrac{as+bs^2+s^3}{1+(a+\delta)s+bs^2+s^3}$ $\delta=\dfrac{R_2RC}{(R+R_2)}\omega_N\approx R_2C\omega_N$

Divide top and bottom by $as+bs^2+s^3$ so that: -

$$T(s)=\left(1+\frac{1+\delta s}{as+bs^2+s^3}\right)^{-1}=\left(1+\left(\frac{1}{s^3}+\frac{\delta}{s^2}\right)\left(1+\frac{b}{s}+\frac{a}{s^2}\right)^{-1}\right)^{-1}$$

To a very good approximation at high frequency, neglecting terms smaller than s^{-4}: -

$$|s|>>1 \;\Rightarrow\; T(s)\approx\left(1+\left(\frac{1}{s^3}+\frac{\delta}{s^2}\right)\left(1-\frac{b}{s}\right)\right)^{-1}$$

Repeat the approximation: $$T(s)\approx 1-\left(\frac{1}{s^3}+\frac{\delta}{s^2}\right)\left(1-\frac{b}{s}\right)$$

It is now clear that δ results in a second order error (in terms of frequency) and one can neglect contributions of fourth order and higher as long as δ is not too small. The second order (in-phase) and third order (quadrature) analysis is, therefore: -

$$\delta b<<1 \qquad \text{and} \qquad |s|>>1 \;\Rightarrow\; T(s)\approx 1-\frac{\delta}{s^2}-\frac{1}{s^3}$$

In more convenient form: $$f>>f_N \;\Rightarrow\; T(s)\approx 1+\delta\left(\frac{f_N}{f}\right)^2-j\left(\frac{f_N}{f}\right)^3$$

4.3.2 Example calculation

If one applies a compensator of 10Ω and 1.6mF to the previous example with a source resistance of up to 100Ω: -

Inductance of both windings: $L_1=L_2\approx 16H$

Ratio (primary) winding resistance: $R_1\approx 4.6\Omega$

Energising winding resistance: $R_2\approx 2.3\Omega$

$$f_N=\frac{1}{2\pi}\left(\frac{(R+R_2)(R_S+R_1)}{RCL_1L_2}\right)^{\frac{1}{3}}\approx 1.1Hz \;\text{ and }\; \delta=\frac{R_2RC}{(R+R_2)}2\pi f_N\approx 0.020$$

$$f=75Hz \;\Rightarrow\; T(s)\approx 1+4.3\times10^{-6}-j3.2\times10^{-6}$$

The result is stable - a small resonant peak which increases with source resistance. Quadrature error is very small but the in-phase error (4.3ppm) is still too high for many applications.

4.3.3 Spreadsheet simulation

Algebraic analysis with a compensator is quite complicated and it is worth checking the conclusions by computer simulation. The simplest approach is to perform the calculations in bite-size chunks based on the matrix method. From above the inverse matrix and determinant, without a compensator, are: -

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} (sL_1 + sL_2)R_2 + s^2L_1L_2 & sL_2R_1 \\ sL_2R_2 & sL_2R_1 + s^2L_1L_2 \end{pmatrix}$$

$$|\mathbf{Z}| = R_1R_2 + sL_2R_1 + sL_1R_2 + sL_2R_2 + s^2L_1L_2$$

One simply needs to make the substitution: $R_2 \rightarrow Z = R_2 + \dfrac{R}{1+sRC}$

The matrix and determinant become: -

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} s(L_1 + L_2)Z + s^2L_1L_2 & sL_2R_1 \\ sL_2Z & sL_2R_1 + s^2L_1L_2 \end{pmatrix}$$

$$|\mathbf{Z}| = R_1Z + sL_2R_1 + s(L_1 + L_2)Z + s^2L_1L_2$$

Also, from section 4.2.1, the source resistance affects the feedback factor and $R_2 \rightarrow Z$ so that: -

$$V_1 = V_2 = \frac{|\mathbf{Z}|}{|\mathbf{Z}| + ZR_S}V_{IN} \quad \Rightarrow \quad V_{OUT} = \frac{\Omega_{11} + \Omega_{12}}{|\mathbf{Z}| + ZR_S}V_{IN}$$

$$\Rightarrow \quad T(s) = \frac{U_1}{V_{IN}} = \frac{\Omega_{11} + \Omega_{12}}{|\mathbf{Z}| + ZR_S}$$

For a typical application: $L_1 = 16H \quad L_2 = 2H \quad R_1 = 6.9\Omega \quad R_2 = 2.3\Omega \quad R = 10\Omega \; C = 1.6mF$

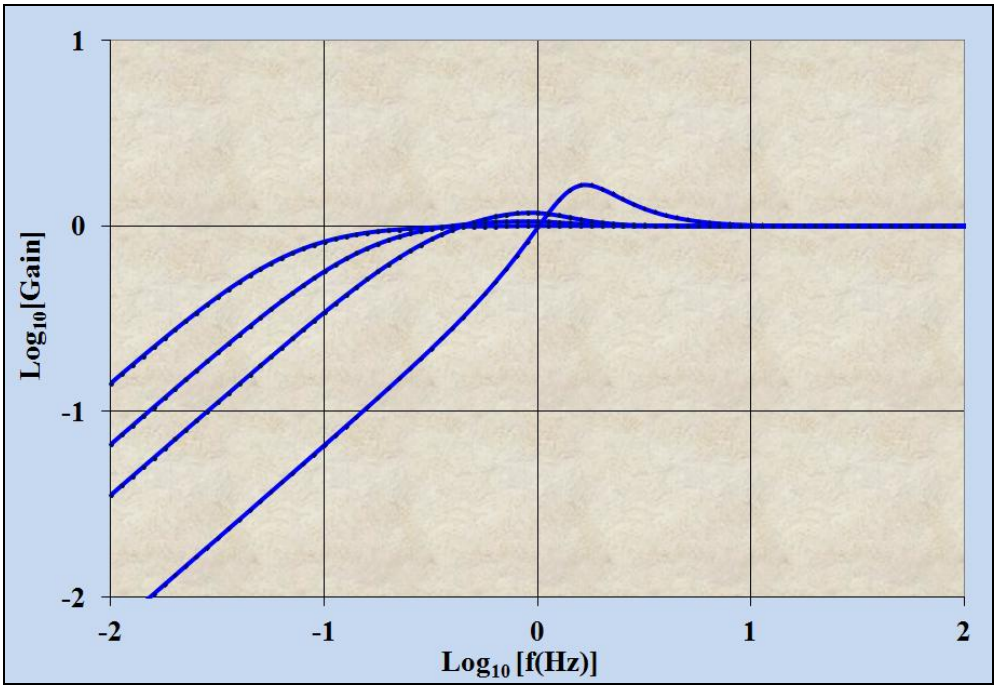


Fig. 4.3.3.1 Spreadsheet simulations for source resistance 1, 10, 25 and 100Ω

4.4 Extra turns around the top core

Adding extra turns around the top core can be represented, diagrammatically, according to the usual convention: windings horizontally adjacent to cores are wound around those cores, at least in part. In this case N_P turns of the ratio winding rope are wound around the top core only then another N_P turns are wound around both cores. The bootstrapping voltage induced in the ratio winding is the same as before. Ratio accuracy is also maintained as the secondary winding is part of the same rope as the primary and shares the same flux.

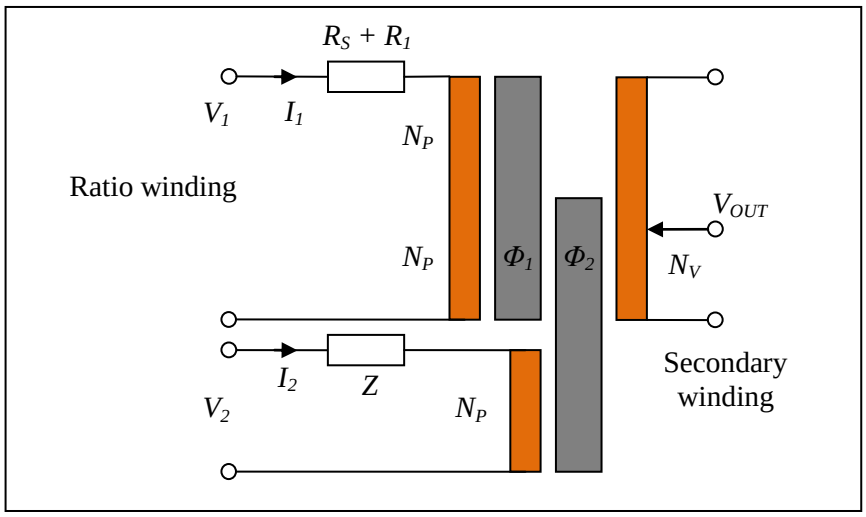


Fig. 4.4.1 Extra turns around the top (ratio) core

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $(s = j\omega \equiv d/dt)$: -

$$V_1 = I_1R_1 + s\left(2N_P\Phi_1 + N_P\Phi_2\right) \qquad V_2 = I_2R_2 + sN_P\Phi_2 \qquad \text{etc.}$$

Note that the extra factor of two is because of the extra turns around the top core. The expressions for V_2 and V_3 remain unaltered. Similarly, for ideal magnetic circuits, the magneto-motive force in the top core is also doubled so that the fluxes are: -

$$\Phi_1 = 2N_PI_1A_{L1} \qquad \Phi_2 = \left(N_PI_1 + N_PI_2\right)A_{L2}$$

A_{L1} = permittance of core 1. A_{L2} = permittance of core 2.

From the above: $$V_1 = I_1R_1 + s\left(4N_P^2I_1A_{L1} + N_P^2I_1A_{L2} + N_P^2I_2A_{L2}\right)$$

Previously I defined the winding inductances: $L_1 = N_P^2A_{L1}$ and $L_2 = N_P^2A_{L2}$

Then: $$V_1 = I_1\left(R_1 + 4sL_1 + sL_2\right) + I_2sL_2 \qquad \text{etc}$$

The rest of the analysis is exactly the same as before (see sections 1, 2 and 3) but with the transformation: $L_1 \rightarrow 4L_1$. One may as well retain the formulae for the transfer function and input impedances and calculate the inductance L_1 with twice the number of turns (i.e. the actual number of turns around the top core): -

$$L_1 = \left(2N_P\right)^2A_{L1}$$

A further factor of two can be had by using a higher (SM200) permeability top core: -

$$L_1 \approx 8L_2$$

The extra turns also increase the winding resistance by approximately 50%: $R_1 \approx 3 \times R_2$

4.4.1 Example calculation

If one incorporate these changes (SM200 top core with extra turns) into the previous example the result is: -

Winding inductances: $L_1 \approx 128H$ and $L_2 \approx 16H$

The increase in L_1 by a factor of 8 reduces the natural frequency by the factor of two $\left(8^{\frac{1}{3}}\right)$

Ratio (primary) winding resistance also increases $\left(R_1 \approx 3 \times R_2\right)$: $R_1 \approx 6.9\Omega$

Energising winding resistance: $R_2 \approx 2.3\Omega$

With three values of source resistance: $R_S \approx 1\Omega$, $R_S \approx 10\Omega$ and $R_S \approx 100\Omega$

Compensator resistance for worst case source resistance $\left(R_S \approx 100\Omega\right)$: $R_C \approx R_S \approx 100\Omega$

To give the compensator a reasonable chance to restore the bootstrap effect at operating frequency (typically 25Hz) the large parallel capacitor needs to be about the same as the compensator resistance at about 1Hz: -

$$\frac{1}{2\pi fC} \approx 100\Omega \text{ at } f = 1Hz \Rightarrow C \approx 1.6mF$$

With a spreadsheet model (with sliders) the results exhibit an increasing natural frequency with a small resonant peak for the worst case (R_S = 100Ω). Increasing the compensator capacitance reduces ratio errors but increases the size of the resonant peak.

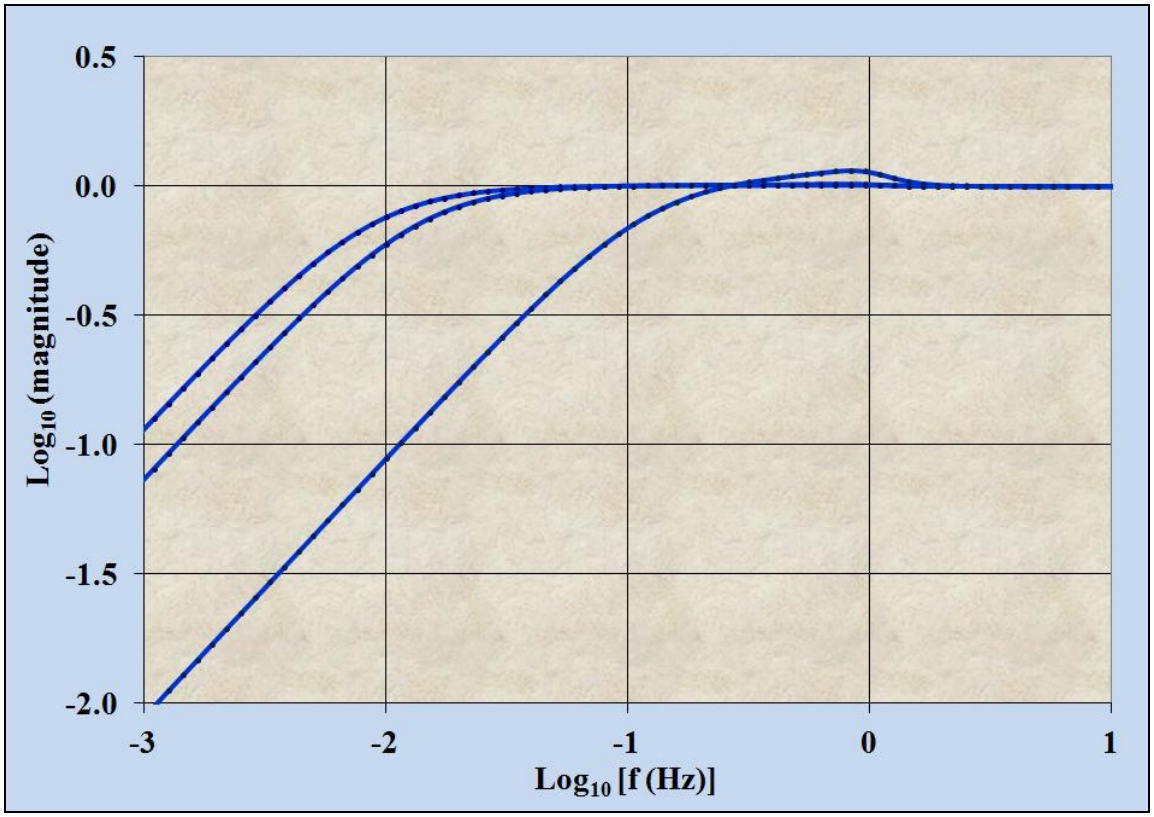


Fig. 4.4.1.1 Frequency response with 1, 10 and 100Ω source resistance [1]

1. Spreadsheet: “Two-stage with compensator MK1”.

As expected the quadrature error is small and third order with respect to frequency. In-phase error is second order (eventually) and also reduced because of the lower natural frequency.

$$f >> f_N \quad \Rightarrow \quad T(s) \approx 1 + \delta\left(\frac{f_N}{f}\right)^2 - j\left(\frac{f_N}{f}\right)^3$$

Real (blue) and imaginary (green)

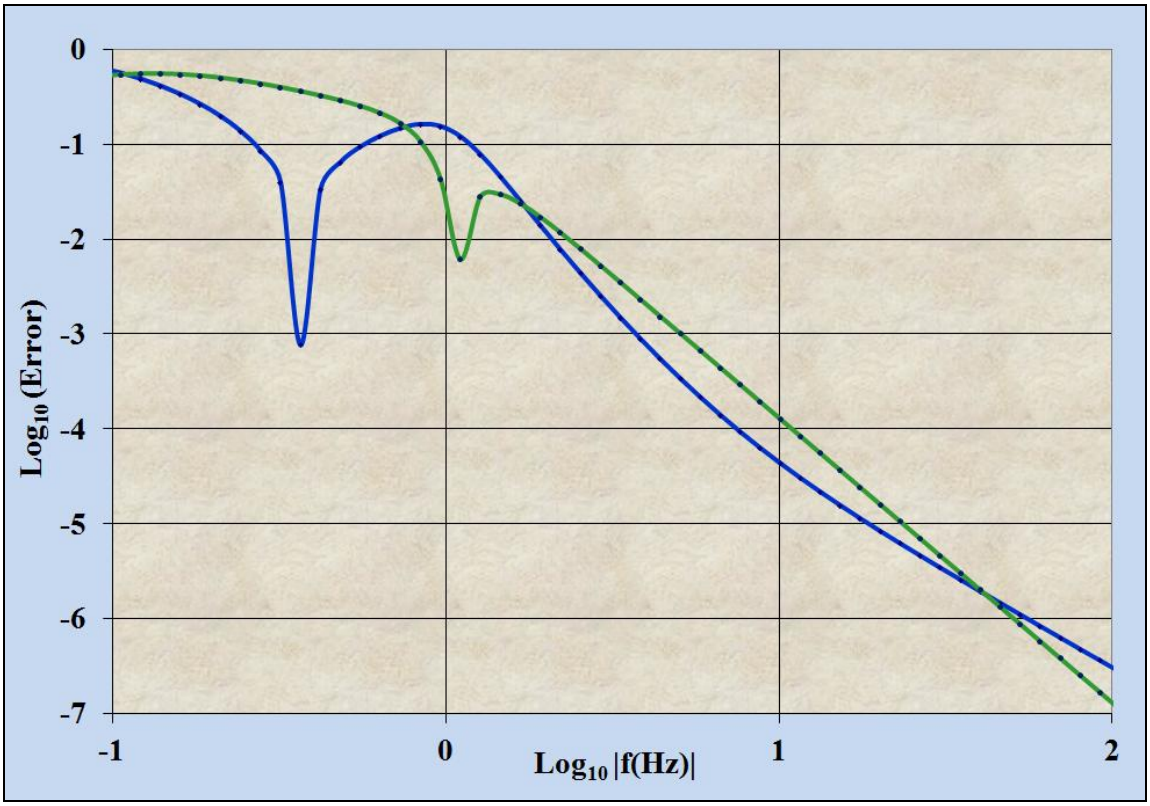


Fig.4.4.1.2 Ratio errors with a compensator (100Ω and 1.6mF) and 100Ω source resistance.

5. Loading effects

Whereas the main ratio secondary is seldom required to deliver current to a load it is frequently the case that a strand of the ratio winding is used to "equalise" the voltage of the next decade. This is often required because the next decade is formed on a separate rope or even a separate transformer or IVD. Similarly, the energising core, having the great majority of the flux, can be used, by means of a secondary winding, to drive the energising winding of a subsequent two-stage IVD or transformer. The load, reflected back to the primary windings, results in a small increase in current as shown in the case of single stage transformers [1]. The worst case is, not surprisingly, when there is a large source resistance and corresponding stability compensator (active drive of energising winding only). The analysis is much more complicated than for a single stage IVD/transformer but the algebraic approach is worth it. The general principles revealed can also be applied to more complex systems (e.g. three-stage RTs) without the need for detailed computer simulation.

5.1 Loading the energising core

In a typical application the third and fourth decades are obtained with a separate transformer or IVD. The second energising core (decades 3 & 4) is powered by a secondary winding on the first energising core (decades 1 & 2).

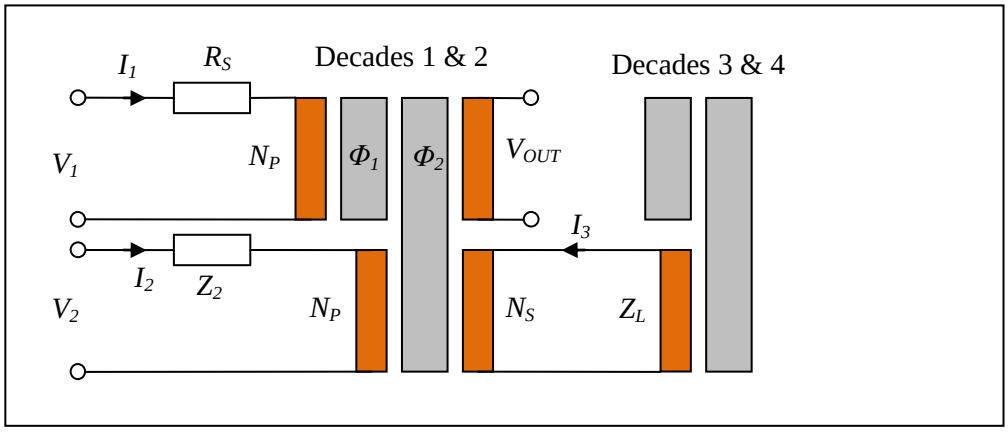


Fig. 5.1.1 The first energising core supplies a second energising core

5.1.1 The effect on energising accuracy of loading the energising secondary

Whereas decades 3 & 4 are less critical than decades 1 & 2 it is still important that the energising secondary output voltage should be accurate (typically the input voltage divided by 100). The source impedance of the secondary, consisting of the winding resistance, R_3, and, most likely, the source resistance and/or the compensator impedance reflected to the secondary, must be kept very low compared to the load impedance (i.e. decades 3 & 4 energising winding inductance). The best method for analysing the effect is to imagine applying a third voltage to the secondary which then requires a 3×3 matrix description.

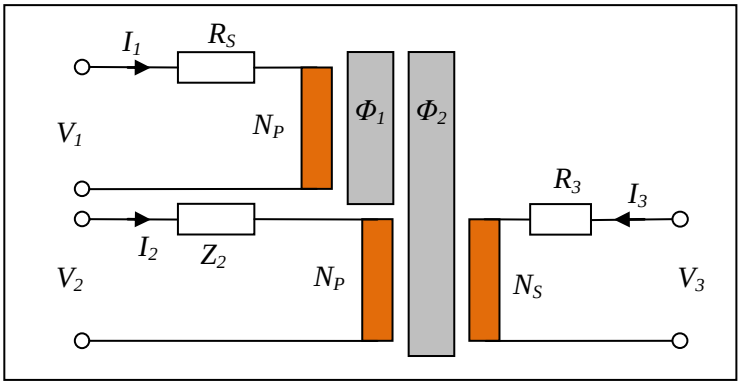


Fig. 5.1.1.1 Circuit model for analysing loading on the energising secondary

1. Part 3, monograph 2: "Single-stage IVD and RTs". See sections 4.4 and 4.5.

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $(s = j\omega \equiv d/dt)$: -

$$V_1 = I_1 R_S + sN_P(\Phi_1 + \Phi_2) \qquad V_2 = I_2 Z_2 + sN_P \Phi_2 \qquad V_3 = I_3 R_3 + sN_S \Phi_2$$

For ideal magnetic circuits the fluxes are: $\Phi_1 = N_P I_1 A_{L1}$ and $\Phi_2 = (N_P I_1 + N_P I_2 + N_S I_3) A_{L2}$

A_{L1} = permittance of core 1. A_{L2} = permittance of core 2.

From the above: $V_1 = I_1 R_S + s\left(N_P^2 I_1 A_{L1} + N_P^2 I_1 A_{L2} + N_P^2 I_2 A_{L2} + N_P N_S I_3 A_{L2}\right)$

$$V_2 = I_2 Z_2 + s\left(N_P^2 I_1 A_{L2} + N_P^2 I_2 A_{L2} + N_P N_S I_3 A_{L2}\right)$$

$$V_3 = I_3 R_3 + s\left(N_S N_P I_1 A_{L2} + N_S N_P I_2 A_{L2} + N_S^2 I_3 A_{L2}\right)$$

Define winding inductances: $L_1 = N_P^2 A_{L1}$ and $L_2 = N_P^2 A_{L2}$

Then: $V_1 = I_1(R_S + sL_1 + sL_2) + I_2 sL_2 + I_3 \dfrac{N_S}{N_P} sL_2$

$$V_2 = I_1 sL_2 + I_2(Z_2 + sL_2) + I_3 \frac{N_S}{N_P} sL_2$$

$$V_3 = I_1 \frac{N_S}{N_P} sL_2 + I_2 \frac{N_S}{N_P} sL_2 + I_3 \left(\frac{N_S}{N_P}\right)^2 sL_2 + I_3 R_3$$

Define a parameter (ratio of primary to secondary) to simplify the algebra $\rho = N_S / N_P$. In vector-matrix notation: $\mathbf{V} = \mathbf{ZI}$: -

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} R_S + s(L_1 + L_2) & sL_2 & \rho sL_2 \\ sL_2 & Z_2 + sL_2 & \rho sL_2 \\ \rho sL_2 & \rho sL_2 & R_3 + \rho^2 sL_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix}$$

The inverse is: $\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$ which one can calculate using Cramer's rule (see appendix 3) but first the determinant: -

$$|\mathbf{Z}| = (R_S + sL_1 + sL_2)\{(Z_2 + sL_2)(R_3 + \rho^2 sL_2) - \rho^2 s^2 L_2^2\}$$
$$+ sL_2\{\rho^2 s^2 L_2^2 - sL_2(R_3 + \rho^2 sL_2)\}$$
$$+ \rho sL_2\{\rho s^2 L_2^2 - (Z_2 + sL_2)\rho sL_2\}$$

Tackle in three parts and re-arrange into powers of *s* in order to identify terms which may cancel: -

$$|\mathbf{Z}| = Z_2 R_3 R_S + s\{Z_2 R_3(L_1 + L_2) + L_2 R_3 R_S + R_S Z_2 \rho^2 L_2\} + s^2\{L_2 R_3(L_1 + L_2) + Z_2 \rho^2 L_2(L_1 + L_2)\}$$
$$- s^2 L_2^2 R_3$$
$$- Z_2 \rho^2 s^2 L_2^2$$

$$|\mathbf{Z}| = Z_2 R_3 R_S + s\{Z_2 R_3(L_1 + L_2) + L_2 R_3 R_S + R_S Z_2 \rho^2 L_2\} + s^2\{L_1 L_2 R_3 + Z_2 \rho^2 L_1 L_2\}$$

Define $\mathbf{Y}$ such that: $\mathbf{Y} = |\mathbf{Z}|\mathbf{Z}^{-1}$ so that $\mathbf{I} = \dfrac{\mathbf{Y}}{|\mathbf{Z}|}\mathbf{V}$

If one assumes ideal followers ($V_1 = V_2$) then the current in the secondary is: $I_3 = \frac{1}{|\mathbf{Z}|}\left(Y_{13}V_1 + Y_{23}V_1 + Y_{33}V_3\right)$

From Cramer's rule (see appendix 3) the relevant components of **Y** can be calculated. The matrix is repeated for convenience: -

$$\mathbf{Z} = \begin{pmatrix} R_S + s(L_1 + L_2) & sL_2 & \rho sL_2 \\ sL_2 & Z_2 + sL_2 & \rho sL_2 \\ \rho sL_2 & \rho sL_2 & R_3 + \rho^2 sL_2 \end{pmatrix}$$

$$Y_{13} = \left(Z_{12}Z_{23} - Z_{13}Z_{22}\right) = \rho s^2 L_2^2 - \rho sL_2\left(Z_2 + sL_2\right) = -Z_2\rho sL_2$$

$$Y_{23} = \left(Z_{13}Z_{21} - Z_{11}Z_{23}\right) = \rho s^2 L_2^2 - \left(R_S + sL_1 + sL_2\right)\rho sL_2 = -\left(R_S + sL_1\right)\rho sL_2$$

$$Y_{33} = \left(Z_{11}Z_{22} - Z_{12}Z_{21}\right) = \left(R_S + sL_1 + sL_2\right)\left(Z_2 + sL_2\right) - s^2L_2^2$$

$$\Rightarrow \quad I_3 = \frac{1}{|\mathbf{Z}|}\left(\left\{\left(R_S + sL_1 + sL_2\right)\left(Z_2 + sL_2\right) - s^2L_2^2\right\}V_3 - \left\{Z_2\rho sL_2 + \left(R_S + sL_1\right)\rho sL_2\right\}V_1\right)$$

If one holds V_1 constant then the input admittance of the secondary is the (partial) derivative: -

$$\frac{\partial I_3}{\partial V_3} = \frac{\left(R_S + sL_1 + sL_2\right)\left(Z_2 + sL_2\right) - s^2L_2^2}{|\mathbf{Z}|}$$

The reciprocal of the admittance is the input impedance as seen from the source V_3. With a little algebra: -

$$Z_3 = \frac{Z_2R_3R_S + s\left\{Z_2R_3\left(L_1 + L_2\right) + L_2R_3R_S + R_SZ_2\rho^2L_2\right\} + s^2\left\{L_1L_2R_3 + Z_2\rho^2L_1L_2\right\}}{R_SZ_2 + s\left(L_2R_S + Z_2\left(L_1 + L_2\right)\right) + s^2L_1L_2}$$

At sufficiently low frequency this is simply the winding resistance of the secondary: -

$$\left|s^2\right| << \frac{|Z_2|R_S}{L_1L_2} \quad \Rightarrow \quad Z_3 \approx R_3$$

At sufficiently high frequency it is the winding resistance of the secondary plus the source impedance in series with the energising winding, reduced by the turns ratio squared: -

$$\left|s^2\right| >> \frac{|Z_2|R_S}{L_1L_2} \quad \Rightarrow \quad Z_3 \approx R_3 + \rho^2Z_2$$

It is a useful check that one can arrive at the same result in a different way. If one imagines a load connected to the secondary (for which one can think in terms of output impedance, Z_{OUT}) one has the following model: -

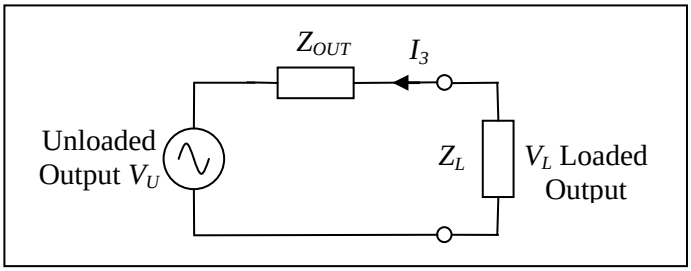


Fig. 5.1.1.2 Circuit model for output impedance

From above, the loaded output voltage is (note the direction the current was defined), assuming ideal voltage followers ($V_1 = V_2$): -

$$I_3 = \frac{1}{|\mathbf{Z}|}\left(Y_{13}V_1 + Y_{23}V_1 + Y_{33}V_3\right) \qquad \text{and} \qquad Z_L = -\frac{I_3}{V_3} \qquad \Rightarrow \qquad V_3 = V_L = -\frac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}}V_1$$

If the load impedance is infinite the result is the unloaded output: $Z_L \to \infty \quad \Rightarrow \quad V_3 = V_U = -\dfrac{\left(Y_{13}+Y_{23}\right)}{Y_{33}}V_1$

The effect of loading due to the secondary output impedance is: -

$$\frac{V_L}{V_U} = \frac{Z_L}{Z_{OUT} + Z_L} \quad \Rightarrow \quad Z_{OUT} = Z_L\left(\frac{V_U}{V_L} - 1\right) \quad \Rightarrow \quad Z_{OUT} = \frac{|\mathbf{Z}|}{Y_{33}}$$

This agrees with the result from the previous method (for admittance): $\dfrac{1}{Z_{OUT}} = \dfrac{\partial I_3}{\partial V_3} = \dfrac{Y_{33}}{|\mathbf{Z}|}$ QED.

One can conclude that the first energising core (decades 1 & 2) and its windings behave, at high frequency, very much like a single stage transformer with a load. The compensator impedance reflects to the output, reduced by the turns ratio squared, in series with the secondary resistance. The main source resistance (in series with the ratio winding) has a negligible contribution. This makes sense since most of the flux is in the energising core.

In a typical application the secondary voltage is one hundredth the primary and the reflected load in parallel with the first energising winding inductance is, therefore, 10^4 times the load impedance. As the load impedance, and its reflection, is substantially inductive (with a small series resistance) and usually of the same order of magnitude the effect is equivalent to a small reduction (typically < 0.01%) in the first energising winding inductance, L_2 [1].

Similarly, the source impedance (substantially Z_2) reflected to the secondary output [2] is reduced by a factor of 10^4 (Z_2 drops from a maximum of typically100Ω at low frequency to the winding resistance, R_2, at high frequency) and its contribution to the output impedance is negligible. Decade 3 & 4 IVD can be analysed, therefore, as if the energising winding were driven by a follower (i.e. very low source impedance). As the energising winding usually has a small number of turns (2, 3 or 4 at most) it is usually sufficient to only use the same gauge of wire as the subsequent energising primary.

The effect on the ratio primary input impedance/current and transformer accuracy is less obvious as the following demonstrates: -

5.1.2 Loading the energising core and its effect on ratio accuracy

If one adds a ratio secondary with the same number of turns as the primary (ratio 1:1) then the output is the input minus the voltage developed across the source resistance due to the current into the ratio primary. The transfer function is, therefore: -

$$V_{OUT} = V_{IN} - I_1 R_S \qquad \text{and} \qquad I_1 = \frac{1}{|\mathbf{Z}|}\left(Y_{11}V_1 + Y_{21}V_1 + Y_{31}V_3\right) \quad \Rightarrow \quad T_L(s) = 1 - \frac{R_S}{|\mathbf{Z}|}\left(Y_{11}+Y_{21}\right) + \frac{R_3}{|\mathbf{Z}|}\frac{V_3}{V_{IN}}Y_{31}$$

From the previous section: $\dfrac{V_3}{V_{IN}} = -\dfrac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}}$

The complete transfer function is, therefore: $T_L(s) = 1 - \dfrac{R_S}{|\mathbf{Z}|}\left(Y_{11}+Y_{21}\right) - \dfrac{R_S}{|\mathbf{Z}|}\dfrac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}}Y_{31}$

1. Part 3, monograph 2: “Single stage IVDs and RTs”. See section 4.3.
2. Ibid section 4.4.

From Cramer's method (see appendix 3): -

$$Y_{11}+Y_{21}=Z_2\left(R_3+\rho^2 s\left(L_1+L_2\right)\right)+\rho^2 s^2 L_1 L_2$$

$$Y_{31}=\rho s^2 L_2^2-\left(Z_2+sL_2\right)\rho s\left(L_1+L_2\right)$$

$$Y_{13}+Y_{23}=-\rho sL_2\left(Z_2+R_S+sL_1\right)$$

$$Y_{33}=\left(R_S+sL_1+sL_2\right)\left(Z_2+sL_2\right)-s^2L_2^2$$

$$\left|\mathbf{Z}\right|=Z_2R_3R_S+s\left\{Z_2R_3\left(L_1+L_2\right)+L_2R_3R_S+R_SZ_2\rho^2L_2\right\}+s^2\left\{L_1L_2R_3+Z_2\rho^2L_1L_2\right\}$$

It is useful to check against the previous result in the case with no load: -

$$\rho=0 \quad\Rightarrow\quad Y_{31}=0 \quad\Rightarrow\quad T_U(s)=1-\frac{R_S}{\left|\mathbf{Z}\right|}\left(Y_{11}+Y_{21}\right)$$

Also $Y_{11}+Y_{21}=Z_2R_3$ and $\left|\mathbf{Z}\right|=Z_2R_3R_S+s\left\{Z_2R_3\left(L_1+L_2\right)+L_2R_3R_S\right\}+s^2\left\{L_1L_2R_3\right\}$

$$T_U(s)=1-\frac{R_SZ_2R_3}{\left|\mathbf{Z}\right|}$$

$$\Rightarrow\quad T_U(s)=1-\frac{R_SZ_2}{Z_2R_S+s\left\{Z_2\left(L_1+L_2\right)+L_2R_S\right\}+s^2\left\{L_1L_2\right\}}$$

This is the same as the result obtained for the basic two-stage transformer (without source resistance and compensator: see section 3, page 8) if one makes the substitutions: -

$$Z_2\to R_2 \qquad \text{and} \qquad R_S\to R_1 \qquad \text{then} \qquad T(s)=\frac{s\left(R_2L_1+R_2L_2+R_1L_2\right)+s^2L_1L_2}{R_1R_2+s\left(R_2L_1+R_2L_2+R_1L_2\right)+s^2L_1L_2}$$

Apart from the term Z_2R_3 contributions from $Y_{11}+Y_{21}$ and $\left(Y_{13}+Y_{23}\right)Y_{31}$ include factors $\rho^2 s$ and $\rho^2 s^2$.

At high frequency, therefore, the terms in s^2 dominate so that, to a good approximation: -

$$Y_{11}+Y_{21}\approx Z_2R_3+\rho^2s^2L_1L_2 \qquad Y_{31}\approx\rho s^2L_1L_2 \quad Y_{13}+Y_{23}\approx-\rho s^2L_1L_2 \quad Y_{33}\approx s^2L_1L_2 \quad \left|\mathbf{Z}\right|\approx s^2\left\{L_1L_2R_3+Z_2\rho^2L_1L_2\right\}$$

N.B. The contribution from Z_2R_3 depends on the value of ρ^2 as well as frequency.

One can now substitute these into the full expression. At high frequency, to a good approximation: -

$$\text{From above:} \qquad T_L(s)=1-\frac{R_S}{\left|\mathbf{Z}\right|}\left(Y_{11}+Y_{21}\right)-\frac{R_S}{\left|\mathbf{Z}\right|}\frac{Z_L\left(Y_{13}+Y_{23}\right)}{\left|\mathbf{Z}\right|+Z_LY_{33}}Y_{31}$$

$$\left|s^2\right|<<\frac{\left|Z_2\right|R_S}{L_1L_2} \quad\Rightarrow\quad T_L(s)\approx1-\frac{R_SZ_2R_3}{\left|\mathbf{Z}\right|}-\frac{R_S}{\left|\mathbf{Z}\right|}\rho^2s^2L_1L_2+\frac{R_S}{\left|\mathbf{Z}\right|}\frac{Z_L\rho s^2L_1L_2}{\left|\mathbf{Z}\right|+Z_Ls^2L_1L_2}\rho s^2L_1L_2$$

The first two terms are the unloaded transfer function and the rest of the expression (keeping the symbol for $\left|\mathbf{Z}\right|$ for the time being) simplifies nicely: -

$$T_L(s)\approx T_U(s)-\frac{R_S}{\left|\mathbf{Z}\right|}\rho^2s^2L_1L_2\left(1-\frac{Z_Ls^2L_1L_2}{\left|\mathbf{Z}\right|+Z_Ls^2L_1L_2}\right)$$

The expression in brackets simplifies further: -

$$T_L(s)\approx T_U(s)-\frac{R_S}{\left|\mathbf{Z}\right|}\rho^2s^2L_1L_2\frac{\left|\mathbf{Z}\right|}{\left|\mathbf{Z}\right|+Z_Ls^2L_1L_2}$$

The determinant and factor $s^2 L_1 L_2$ now cancel to obtain: -

$$T_L(s) \approx T_U - \rho^2 \frac{R_S}{R_3 + \rho^2 Z_2 + Z_L}$$

The denominator is simply the total load impedance, including the actual load plus the winding resistance in series with the source impedance (to the energising winding) reflected from the primary. In practice the former is much larger than the latter two and one finds the final form: -

$$T_L(s) \approx T_U(s) - \rho^2 \frac{R_S}{Z_L}$$

One can interpret this as either reflected source impedance reduced or reflected load impedance increased by a factor of the turns ratio squared.

The effect is reflection to/from the ratio primary <u>as well as</u> the energising primary. Fortunately, in practice, the load impedance is likely to be mostly inductive and the error, therefore, mostly quadrature. It is important, however, <u>not</u> to have a resistive load.

5.2 Loading the ratio secondary

Taking current from the ratio secondary is also a little more complicated and can result in significant ratio errors, especially with high source resistance. The best method for analysing the effect is again (the same as loading the energising core) to imagine applying a third voltage to the ratio secondary.

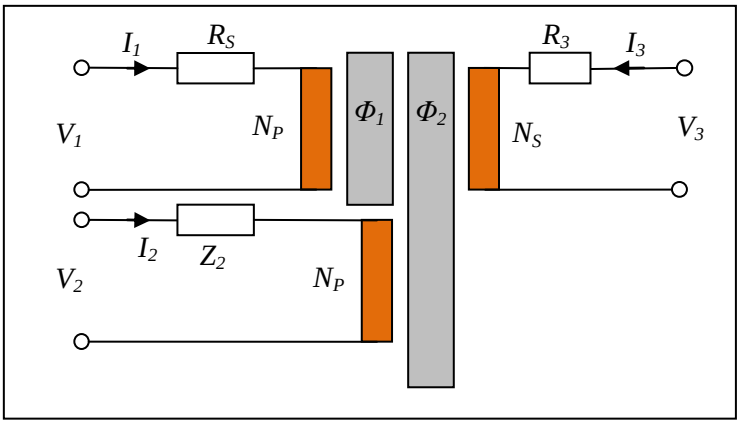


Fig. 5.2.1 Circuit model for analysing loading on the ratio secondary

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $(s = j\omega \equiv d/dt)$: -

$$V_1 = I_1 R_S + sN_P(\Phi_1 + \Phi_2) \qquad V_2 = I_2 Z_2 + sN_P \Phi_2 \qquad V_3 = I_3 R_3 + sN_S(\Phi_1 + \Phi_2)$$

For ideal magnetic circuits the fluxes are: -

$$\Phi_1 = (N_P I_1 + N_S I_3) A_{L1} \qquad \Phi_2 = (N_P I_1 + N_P I_2 + N_S I_3) A_{L2}$$

A_{L1} = permittance of core 1. A_{L2} = permittance of core 2.

From the above: $$V_1 = I_1 R_S + s\left(N_P^2 I_1 A_{L1} + N_P N_S I_3 A_{L1} + N_P^2 I_1 A_{L2} + N_P^2 I_2 A_{L2} + N_P N_S I_3 A_{L2}\right)$$

$$V_2 = I_2 Z_2 + s\left(N_P^2 I_1 A_{L2} + N_P^2 I_2 A_{L2} + N_P N_S I_3 A_{L2}\right)$$

$$V_3 = I_3 R_3 + s\left(N_S N_P I_1 A_{L1} + N_S^2 I_3 A_{L1} + N_S N_P I_1 A_{L2} + N_S N_P I_2 A_{L2} + N_S^2 I_3 A_{L2}\right)$$

Define winding inductances: $L_1 = N_P^2 A_{L1}$ and $L_2 = N_P^2 A_{L2}$

$$\text{Then:} \quad V_1 = I_1\left(R_S + sL_1 + sL_2\right) + I_2 sL_2 + I_3 \frac{N_S}{N_P}\left(sL_1 + sL_2\right)$$

$$V_2 = I_1 sL_2 + I_2\left(Z_2 + sL_2\right) + I_3 \frac{N_S}{N_P} sL_2$$

$$V_3 = I_1 \frac{N_S}{N_P}\left(sL_1 + sL_2\right) + I_2 \frac{N_S}{N_P} sL_2 + I_3\left(\frac{N_S}{N_P}\right)^2\left(sL_1 + sL_2\right) + I_3 R_3$$

Define parameter (ratio of primary to secondary) to simplify the algebra $\rho = \frac{N_S}{N_P}$.

In vector-matrix notation: $\mathbf{V} = \mathbf{ZI}$: -

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} R_S + s\left(L_1 + L_2\right) & sL_2 & \rho s\left(L_1 + L_2\right) \\ sL_2 & Z_2 + sL_2 & \rho sL_2 \\ \rho s\left(L_1 + L_2\right) & \rho sL_2 & R_3 + \rho^2 s\left(L_1 + L_2\right) \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix}$$

The inverse is: $\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$ which can be calculated using Cramer's rule (see appendix 3) but first the determinant: -

$$\begin{aligned} |\mathbf{Z}| &= \left(R_S + s\left(L_1 + L_2\right)\right)\left\{\left(Z_2 + sL_2\right)\left(R_3 + \rho^2 s\left(L_1 + L_2\right)\right) - \rho^2 s^2 L_2^2\right\} \\ &+ sL_2\left\{\rho^2 s^2 L_2\left(L_1 + L_2\right) - sL_2\left(R_3 + \rho^2 s\left(L_1 + L_2\right)\right)\right\} \\ &+ \rho s\left(L_1 + L_2\right)\left\{\rho s^2 L_2^2 - \left(Z_2 + sL_2\right)\rho s\left(L_1 + L_2\right)\right\} \end{aligned}$$

Tackle in three parts and re-arrange into powers of *s* in order to identify terms which may cancel. Note that for the first line there should be (2 × 2 × 2) + 2= 10 terms: -

$$\begin{aligned} |\mathbf{Z}| &= R_S Z_2 R_3 \\ &+ s\left\{R_S R_3 L_2 + Z_2 R_3\left(L_1 + L_2\right) + R_S Z_2 \rho^2\left(L_1 + L_2\right)\right\} \\ &+ s^2\left\{L_2 R_3\left(L_1 + L_2\right) + L_2 R_S \rho^2\left(L_1 + L_2\right) + Z_2 \rho^2\left(L_1 + L_2\right)^2 - R_S \rho^2 L_2^2\right\} \\ &+ s^3\left\{\rho^2 L_2\left(L_1 + L_2\right)^2 - \rho^2 L_2^2\left(L_1 + L_2\right)\right\} \\ &- s^2 R_3 L_2^2 \\ &- s^2\left\{Z_2 \rho^2\left(L_1 + L_2\right)^2\right\} \\ &+ s^3\left\{\rho^2 L_2^2\left(L_1 + L_2\right) - L_2 \rho^2\left(L_1 + L_2\right)^2\right\} \end{aligned}$$

There is some cancellation, including, as if by magic, all four terms in s^3: -

$$|\mathbf{Z}| = R_S R_3 Z_2 + s\left\{\left(L_1 + L_2\right)R_3 Z_2 + R_S R_3 L_2 + R_S \rho^2 Z_2\left(L_1 + L_2\right)\right\} + s^2\left\{R_3 L_1 L_2 + R_S \rho^2 L_2 L_1\right\}$$

If one assumes ideal followers ($V_1 = V_2$) then the current in the secondary is: $I_3 = \frac{1}{|\mathbf{Z}|}\left(Y_{13}V_1 + Y_{23}V_1 + Y_{33}V_3\right)$

From Cramer's rule the relevant components of **Y** are (see appendix 3): -

$$\mathbf{Z} = \begin{pmatrix} R_S + s(L_1 + L_2) & sL_2 & \rho s(L_1 + L_2) \\ sL_2 & Z_2 + sL_2 & \rho sL_2 \\ \rho s(L_1 + L_2) & \rho sL_2 & R_3 + \rho^2 s(L_1 + L_2) \end{pmatrix}$$

$$Y_{13} = (Z_{12}Z_{23} - Z_{13}Z_{22}) = \rho s^2 L_2^2 - \rho(sL_1 + sL_2)(Z_2 + sL_2)$$
$$Y_{23} = (Z_{13}Z_{21} - Z_{11}Z_{23}) = \rho s^2 (L_1 + L_2)L_2 - \rho sL_2(R_S + sL_1 + sL_2) = -\rho sL_2 R_S$$
$$Y_{33} = (Z_{11}Z_{22} - Z_{12}Z_{21}) = (R_S + sL_1 + sL_2)(Z_2 + sL_2) - s^2 L_2^2$$

$$\Rightarrow \quad I_3 = \frac{1}{|\mathbf{Z}|}\left(+\left\{(R_S + sL_1 + sL_2)(Z_2 + sL_2) - s^2 L_2^2\right\}V_3 - \left\{\rho(sL_1 + sL_2)(Z_2 + sL_2) + \rho sL_2 R_S - \rho s^2 L_2^2\right\}V_1\right)$$

If the voltage applied to the secondary is such that the current is zero, then one can check if the result is consistent with previous analysis: -

$$I_3 = 0 \;\Rightarrow\; \frac{V_3}{V_1} = -\frac{Y_{13} + Y_{23}}{Y_{33}} \;\Rightarrow\; \frac{V_3}{V_1} = \rho\frac{(sL_1 + sL_2)(Z_2 + sL_2) + sL_2 R_S - s^2 L_2^2}{(R_S + sL_1 + sL_2)(Z_2 + sL_2) - s^2 L_2^2}$$

This is the same as the result obtained for the basic two-stage transformer (without source resistance and compensator: see section 3, page 8) if one makes the substitutions: -

$$Z_2 \to R_2 \quad \text{and} \quad R_S \to R_1 \;\Rightarrow\; \frac{V_3}{V_1} = \rho\frac{s(L_2R_1 + L_1R_2 + L_2R_2) + s^2 L_1 L_2}{R_1R_2 + s(L_2R_1 + L_1R_2 + L_2R_2) + s^2 L_1 L_2}$$

If one holds the input voltages ($V_1 = V_2$) constant then the admittance of the secondary is the (partial) derivative: -

$$\frac{\partial I_3}{\partial V_3} = \frac{Y_{33}}{|\mathbf{Z}|} = \frac{(R_S + sL_1 + sL_2)(Z_2 + sL_2) - s^2 L_2^2}{|\mathbf{Z}|}$$

The reciprocal of this is the impedance of the secondary, including that reflected from the ratio primary and energising windings: -

$$Z_{OUT} = \frac{|\mathbf{Z}|}{(R_S + sL_1 + sL_2)(Z_2 + sL_2) - s^2 L_2^2}$$

The result is rather complicated but the main features are evident.

$$Z_{OUT} = \frac{R_S R_3 Z_2 + s\left\{(L_1 + L_2)R_3 Z_2 + R_S R_3 L_2 + R_S \rho^2 Z_2 (L_1 + L_2)\right\} + s^2\left\{R_3 L_1 L_2 + R_S \rho^2 L_2 L_1\right\}}{R_S Z_2 + s\left\{R_S L_2 + Z_2 (L_1 + L_2)\right\} + s^2 L_1 L_2}$$

At sufficiently low frequency this is simply the winding resistance of the secondary: -

$$\text{In the limit:} \qquad s \to 0 \;\Rightarrow\; Z_{OUT} \approx R_3$$

At sufficiently high frequency the output impedance simplifies to: -

$$s \to \infty \;\text{ and typically }\; |s^2| >> \left|\frac{R_S Z_2}{L_1 L_2}\right| \;\Rightarrow\; Z_{OUT} \approx R_3 + \rho^2 R_S$$

As with a single stage transformer the source resistance in series with the ratio primary is reflected to the output by the square of the ratio. The stability compensator in series with the energising winding has a much reduced effect.

As in the previous section one can deduce the transfer function (with load): -

$$I_3 = \frac{1}{|\mathbf{Z}|}\left(Y_{13}V_1 + Y_{23}V_1 + Y_{33}V_3\right) \qquad \text{and} \qquad I_3 = -\frac{V_3}{Z_L} \qquad \Rightarrow \qquad V_3 = V_L = -\frac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}}V_1$$

If the load impedance is infinite the result is the unloaded output: $Z_L \to \infty \quad \Rightarrow \quad V_3 = V_U = -\frac{\left(Y_{13}+Y_{23}\right)}{Y_{33}}V_1$

This is the same as the condition $\left(I_3 = 0\right)$ as in the previous section.

It is again useful to confirm the previous result by a different method. The effect of loading due to the secondary output impedance is: -

$$\frac{V_L}{V_U} = \frac{Z_L}{Z_{OUT} + Z_L} \quad \Rightarrow \quad Z_{OUT} = Z_L\left(\frac{V_U}{V_L} - 1\right) \quad \Rightarrow \quad Z_{OUT} = \frac{|\mathbf{Z}|}{Y_{33}}$$

This agrees with the result from the previous method (for admittance): $\frac{1}{Z_{OUT}} = \frac{\partial I_3}{\partial V_3} = \frac{Y_{33}}{|\mathbf{Z}|}$ QED.

As an extra check one can imagine the result in the limit that the energising current is zero.

In the limit $Z_2 \to \infty$ then $Z_{OUT} = \frac{R_S R_3 + s\left\{\left(L_1 + L_2\right)R_3 + R_S\rho^2\left(L_1 + L_2\right)\right\}}{R_S + s\left\{\left(L_1 + L_2\right)\right\}}$

In the limit $s \to 0$ the result is the resistance of the secondary $\Rightarrow \quad Z_{OUT} = R_3$

In the limit $s \to \infty$ the result is the resistance plus the source resistance reflected from the primary: -

$$Z_{OUT} = R_3 + R_S\rho^2$$

This is exactly what one would expect for a single stage transformer (i.e. without the energising winding).

The loaded transfer function is, from above: $T_L(s) = \frac{V_L}{V_1} = -\frac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}}$

The difference between the loaded and the unloaded transfer function represents the error due only to the loading effect: -

$$T_L(s) - T_U(s) = \frac{Y_{13}+Y_{23}}{Y_{33}} - \frac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}} \quad \Rightarrow \quad T_L(s) = T_U(s) + \frac{Y_{13}+Y_{23}}{Y_{33}} - \frac{Z_L\left(Y_{13}+Y_{23}\right)}{|\mathbf{Z}| + Z_L Y_{33}}$$

$$\Rightarrow T_L(s) = T_U(s) + \frac{Y_{13}+Y_{23}}{Y_{33}}\left(1 - \frac{Z_L Y_{33}}{|\mathbf{Z}| + Z_L Y_{33}}\right) \quad \Rightarrow \quad T_L(s) = T_U(s) + \frac{Y_{13}+Y_{23}}{Y_{33}}\left(\frac{|\mathbf{Z}|}{|\mathbf{Z}| + Z_L Y_{33}}\right)$$

Finally: $T_L(s) = T_U(s)\left(1 - \frac{|\mathbf{Z}|}{|\mathbf{Z}| + Z_L Y_{33}}\right)$

The factor in brackets includes the error due to loading only: $$E_L(s)=\frac{|\mathbf{Z}|}{|\mathbf{Z}|+Z_L Y_{33}}$$

As above, at a sufficiently high frequency, the relevant determinant and matrix element are, to a sufficiently good approximation: -

$$|\mathbf{Z}| \approx s^2\{R_3L_1L_2+R_S\rho^2L_2L_1\} \qquad Y_{33}\approx s^2L_1L_2$$

A factor of $s^2L_1L_2$ cancels to give: -

$$\frac{|\mathbf{Z}|}{|\mathbf{Z}|+Z_LY_{33}} \approx \frac{R_3+\rho^2R_S}{Z_L+R_3+\rho^2R_S}=\frac{Z_{OUT}}{Z_L+Z_{OUT}}$$

This confirms the previous results with: $$\frac{T_L(s)}{T_U(s)}=1-\frac{Z_{OUT}}{Z_L+Z_{OUT}}$$

With a load across the full secondary of, for example, a 1:1 transformer the load impedance must be very large indeed if high accuracy is required. **Even with a 10:1 reduction it is not possible to use an extra strand, for example, to drive a subsequent energising winding.**

Fortunately it is possible to fit two decades on the first transformer. With a reduction of 100:1 it is possible to employ a decade 2 ratio secondary strand to supply the ratio primary (bootstrapped) of the third decade IVD [1].

It is also possible to use a 10:1 extra strand, with a sufficiently large auxiliary inductive load, to "equalise" a small mismatch between the first and second decade rope windings as the next section describes.

6. Equalising windings

A second or subsequent decade can be "equalised" to the first decade in exactly the same way as with single-stage IVDs and RTs [2]. The loading effect on the ratio secondary is the same as for a single-stage transformer and so the analysis is exactly the same. This is often represented, employing the usual convention for a two-stage design, as follows: -

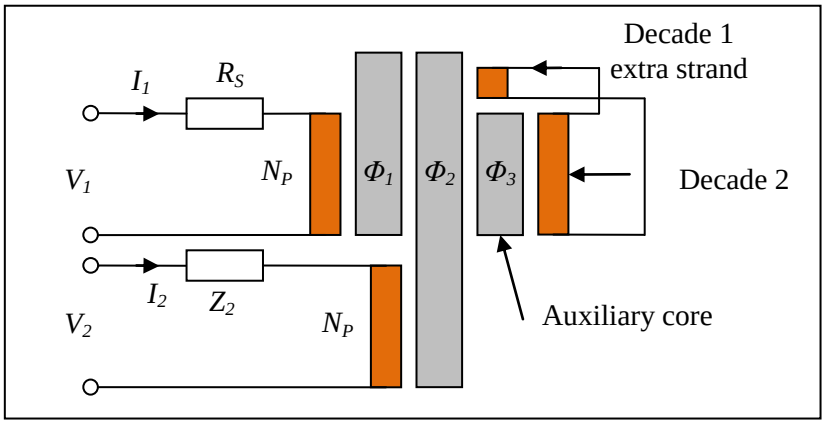


Fig. 6.1 Equalising voltages between the first and second decade

1. Part 3, Monograph 7: "An F18 type ratio transformer bridge". See section 3.
2. Part 3, monograph 2: "Single-stage IVDs and RTs". See section 6.

7. Energising from the secondary

A useful variant is to supply the energising current from the output side. The input is fully differential and the input impedance remains boosted but the output is relative to local 0V. A typical application is to provide the reference input to an R-2R multiplying DAC for the last three decades of a resistance bridge. The input is derived from high accuracy voltage followers [1]. In this case the source resistance is very low and stability is satisfactory with a simple voltage follower driving the energising winding directly. The voltage follower needs very good DC performance (low DC current through the low resistance of the energising winding) and a reasonably good gain-bandwidth product (e.g. OP-27 or equivalent). A composite op-amp with matched pair front end would be even better. With higher source resistance (e.g. a noise matching transformer) the current must be supplied through a compensator [2].

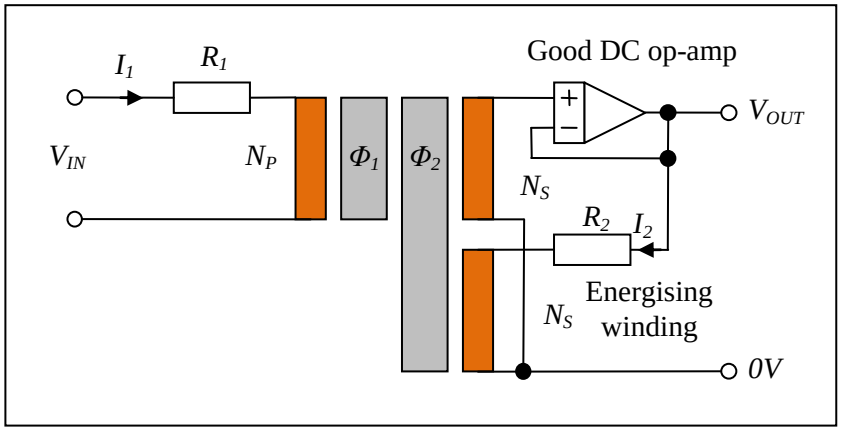


Fig. 7.1 Energising drive from the secondary (low source resistance)

The equivalent circuit is the same as an actively driven two-stage high-pass filter and one can obtain the transfer function by changing the capacitors into resistors and resistors into inductors [3]: -

$$X_1 \rightarrow R_1 \quad \text{and} \quad R_1 \rightarrow sL_1 \quad \text{etc.}$$

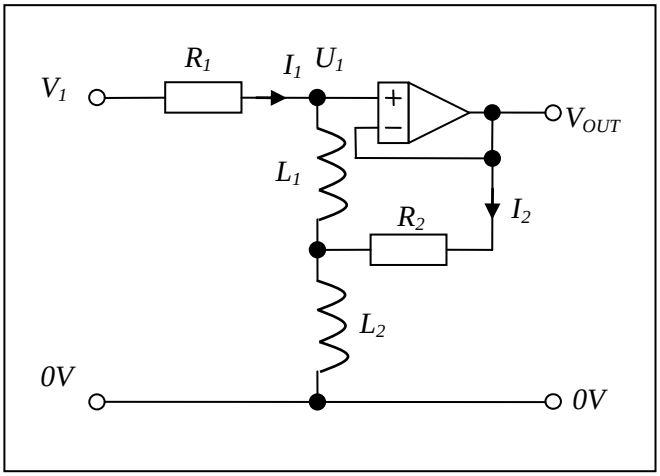


Fig. 7.2 The equivalent circuit

The transfer function becomes: $$T(s) = \frac{\{L_1R_2 + L_2R_2\}s + s^2L_1L_2}{R_1R_2 + \{L_1R_2 + L_2R_2\}s + s^2L_1L_2}$$

A useful parameter is the determinant: $$|\mathbf{Z}| = R_1R_2 + s\{L_2R_1 + (L_1 + L_2)R_2\} + s^2L_1L_2$$

1. Part 3, monograph 7: “An F18 type ratio transformer bridge”.
2. Part 3, monograph 5: “Noise matching transformers”.
3. Part 2, monograph 1: “Two-stage filters”. See section 5.

Divide top and bottom by R_1R_2 and it is easy to see that the normalised transfer function $(s = j\omega/\omega_N)$ is: -

$$T(s) = \frac{as + s^2}{1 + as + s^2} \quad \text{with natural frequency:} \qquad \omega_N = \sqrt{\frac{R_1}{L_1}\frac{R_2}{L_2}} \quad \text{and} \quad a = \left(\frac{L_1 + L_2}{R_1}\right)\omega_N$$

Parameter *a* has one less component compared to the case of energising from the input side (see section 4.2). With the same transformer (winding inductances and resistances are the same) the value of *a* is always lower and the stability margin is reduced slightly. The error analysis is the same: -

$$|s >> 1| \quad \Rightarrow \quad T(s) \approx 1 - \frac{1}{s^2} + \frac{a}{s^3}$$

In the more convenient form, with frequency in Hz: -

$$f >> f_N \quad \Rightarrow \quad T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \quad \text{with} \quad f_N = \frac{1}{2\pi}\sqrt{\frac{R_1R_2}{L_1L_2}}$$

The phase error (quadrature) is third order with respect to frequency and sufficiently low for most applications. The in-phase error is second order and acceptable for some applications.

7.1 Example calculation

A compact transformer with sufficient accuracy for the last three decades of a low frequency (25Hz) resistance bridge can be constructed with a pair of low cost 5c mumetal cores ($\mu_R \approx 50{,}000$). All three windings are 200 turns BNNL of 0.457mm enamelled copper wire (26SWG).

Both cores type 5c (see appendix 1): -
OD = *41mm* ID = *23mm* axial = *15mm.*
The length of wire per turn on the energising winding is approximately: *15 × 2 + 41 – 23 = 48mm*

Inner circumference = 23mm × π = 72mm can accommodate 200 turns in two layers of 0.72mm OD wire.
Outer circumference = 41mm × π = 129mm can accommodate 200 turns of 0.64mm OD in a single layer.
The outer circumference is the limiting dimension: A comfortable fit is 26SWG (enamelled) at 0.457mm diameter with a resistance of 105mΩm^{-1}.

The energising winding requires: $200 \times 48mm \approx 9.6m$ of wire.

Resistance R_2 would be approx: $R_2 \approx 9.6m \times 105m\Omega m^{-1} \approx 1\Omega$

Permittance of both cores: $A_L = 51\mu H/turn^2$

Inductance of both windings: $L_1 = L_2 = N^2 A_L = 200 \times 200 \times 51 \times 10^{-6} \approx 2H$

The ratio windings consist of 200 turns of a twisted pair around both cores (BNNL) of the same wire.
The inner diameter is reduced slightly but the twisted pair is still a comfortable fit with two layers on the inner circumference and a single layer on the outer.

1. Part 3, monograph 5: “Noise matching transformers”.

As a ball park estimate the length of wire, for the ratio winding, is roughly twice that of the energising winding, partly because it is wound around both cores and partly because the strands are twisted into a rope.

$$R_1 \approx 2R_2 = 2\Omega$$

Natural frequency: $f_N = \frac{1}{2\pi}\sqrt{\frac{R_1 R_2}{L_1 L_2}} \approx 0.11Hz$ and $a = \left(\frac{L_1 + L_2}{R_1}\right)\omega_N \approx 1.4$

The error analysis is the same: $T(f) \approx 1 + \left(\frac{f_N}{f}\right)^2 + ja\left(\frac{f_N}{f}\right)^3 \approx 1 + 19\times10^{-6} + j0.12\times10^{-6}$

The resulting frequency response has a slight peak close to the natural frequency [1]: -

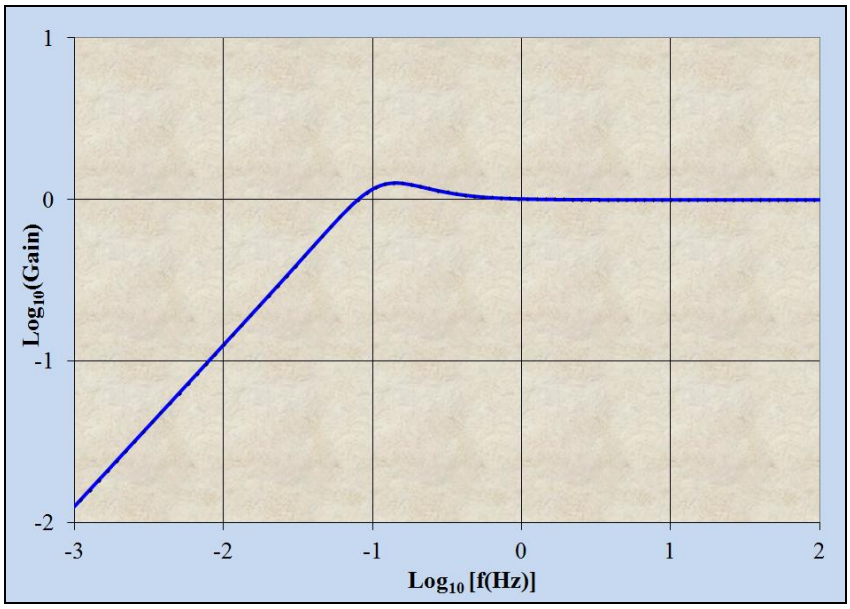


Fig. 7.1.1 Bode plot

The real (in-phase: blue) and imaginary (quadrature: green) errors are second and third order, respectively: -

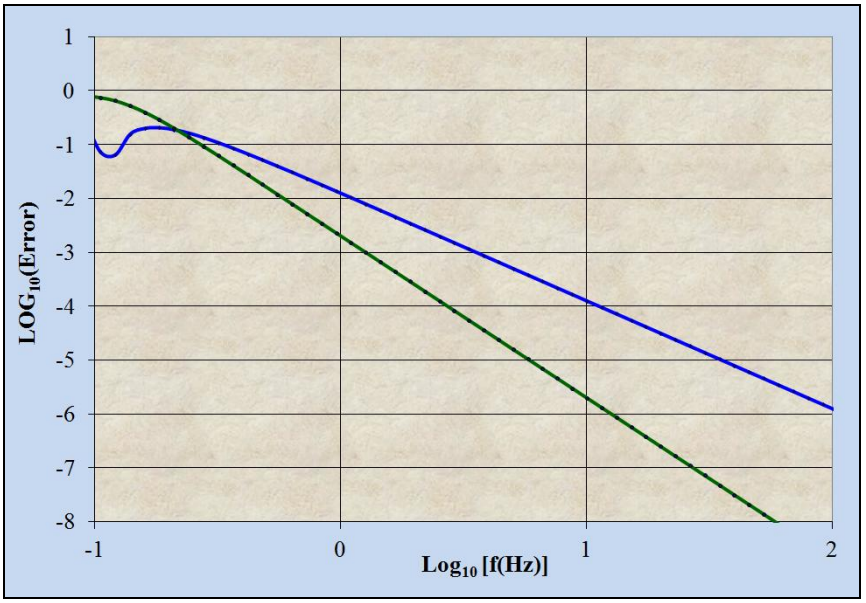


Fig. 7.1.2 Real (blue) and imaginary (green) error components

1. Spreadsheet: “Two-stage energised from output”.

Appendix 1: Toroidal core data courtesy Telcon

Type	Axial code	Case OD	Case ID	Case axial	Circ./turn	Core OD	Core ID	Core axial	L_E	A_E	A_L μH/turn^2	V_{MAX}/turn/Hz
0	a	18	11	5	18	16	13	3.2	45	5	14	1.6E-05
	b	18	11	8	24	16	13	6.4	45	10	28	3.2E-05
1	a	21	11	5	21	19	13	3.2	50	10	26	3.2E-05
	b	21	11	7	24	19	13	4.8	50	15	38	4.8E-05
	c	21	11	8	27	19	13	6.4	50	20	51	6.4E-05
2	a	24	12	7	26	22	14	4.8	57	19	41	5.9E-05
	b	24	12	8	29	22	14	6.4	57	25	55	7.9E-05
	c	24	12	10	32	22	14	7.9	57	31	69	9.9E-05
3	a	31	17	7	29	29	19	4.8	75	23	38	7.1E-05
	b	31	17	9	32	29	19	6.4	75	30	51	9.5E-05
	c	31	17	10	36	29	19	7.9	75	38	63	1.2E-04
4	a	36	20	9	34	33	22	6.4	87	35	51	1.1E-04
	b	36	20	10	37	33	22	7.9	87	44	63	1.4E-04
	c	36	20	12	40	33	22	9.5	87	53	76	1.7E-04
5	a	41	23	9	36	38	25	6.4	100	40	51	1.3E-04
	b	41	23	12	42	38	25	9.5	100	60	76	1.9E-04
	c	41	23	15	48	38	25	12.7	100	81	102	2.5E-04
6	a	50	29	7	36	48	32	4.8	125	38	38	1.2E-04
	b	50	29	10	42	48	32	7.9	125	63	63	2.0E-04
	c	50	29	12	45	48	32	9.5	125	75	76	2.4E-04
	d	50	29	15	52	48	32	12.7	125	100	101	3.2E-04
7	a	60	36	7	38	57	38	4.8	150	45	38	1.4E-04
	b	60	36	12	48	57	38	9.5	150	91	76	2.9E-04
	c	60	36	15	54	57	38	12.7	150	121	102	3.8E-04
8	a	69	42	10	48	67	45	7.9	174	88	64	2.8E-04
	b	69	42	14	54	67	45	11.1	174	123	89	3.9E-04
	c	69	42	17	61	67	45	14.3	174	159	115	5.0E-04
9	a	79	49	14	58	76	51	11.1	200	141	89	4.4E-04
	b	79	49	17	64	76	51	14.3	200	182	114	5.7E-04
	c	79	49	15	61	76	51	12.7	200	161	101	5.1E-04
10	a	90	54	15	65	86	57	11.1	224	158	89	5.0E-04
	b	90	54	18	72	86	57	14.3	224	204	114	6.4E-04
	c	90	54	23	81	86	57	19.1	224	272	153	8.6E-04
11	a	112	66	16	78	108	70	12.7	279	242	109	7.6E-04
	b	112	66	19	85	108	70	15.9	279	303	136	9.5E-04
	c	112	66	23	91	108	70	19.1	279	364	164	1.1E-03
12	a	132	78	18	90	127	83	14.3	329	317	121	1.0E-03
	b	132	78	23	100	127	83	19.1	329	424	162	1.3E-03
	c	132	78	29	112	127	83	25.4	329	564	215	1.8E-03
13	a	164	110	18	90	159	114	14.3	429	322	94	1.0E-03
	b	164	110	23	100	159	114	19.1	429	430	126	1.4E-03
	c	164	110	29	112	159	114	25.4	429	572	167	1.8E-03
14	a	28	17	6	23	25	19	3.2	70	10	18	3.1E-05
	b	28	17	9	29	25	19	6.4	70	20	36	6.3E-05

Notes: -

1. All dimensions in mm or mm^2
2. A_L Assumes μ_R=100,000.
3. V_{MAX} for sinusoidal peak flux density of 0.5T, one turn at 1Hz

Copper wire data

Imperial standard wire gauge (SWG)

No.	16	18	20	22	24	26	28
Diameter (mm)	1.626	1.219	0.914	0.711	0.559	0.457	0.376
$m\Omega m^{-1}$	8.3	14.8	26.0	43.5	70.5	105	155

No.	30	32	34	36	38	40
Diameter (mm)	0.315	0.274	0.234	0.193	0.152	0.122
$m\Omega m^{-1}$	222	293	404	590	950	1480

Metric wire data

Diameter (μm)	50	63	80	100	125	160	200	250	315	400	500
Ωm^{-1}	8.78	5.53	3.43	2.20	1.41	0.86	0.55	0.35	0.22	0.137	0.088

Appendix 2: Detailed equation bashing (see section 4.3): -

The original transfer function is:

$$T(s)=\frac{s(R_2L_1+R_2L_2+R_1L_2)+s^2L_1L_2}{R_2(R_S+R_1)+s(R_2L_1+R_2L_2+R_1L_2)+s^2L_1L_2}$$

I shall make the substitution:

$$R_2 \rightarrow Z=\frac{R+R_2+sR_2RC}{1+sRC}$$

$$T(s)=\frac{\frac{R+R_2+sR_2RC}{1+sRC}sL_1+\frac{R+R_2+sR_2RC}{1+sRC}sL_2+R_1sL_2+s^2L_1L_2}{\frac{(R+R_2)(R_S+R_1)+sR_2RC(R_S+R_1)}{1+sRC}+\frac{R+R_2+sR_2RC}{1+sRC}sL_1+\frac{R+R_2+sR_2RC}{1+sRC}sL_2+R_1sL_2+s^2L_1L_2}$$

Multiply top and bottom by (*1+sRC):* -

$$T(s)=\frac{\frac{R+R_2+sR_2RC}{(\ )}sL_1+\frac{R+R_2+sR_2RC}{(\ )}sL_2+R_1sL_2(1+sRC)+sL_1sL_2(1+sRC)}{\frac{(R+R_2)(R_S+R_1)+sR_2RC(R_S+R_1)}{(\ )}+\frac{R+R_2+sR_2RC}{(\ )}sL_1+\frac{R+R_2+sR_2RC}{(\ )}sL_2+R_1sL_2(1+sRC)+sL_1sL_2(1+sRC)}$$

Sort into powers of *s* noting the extra term in the denominator (in square brackets): -

$$T(s)=\frac{((R+R_2)(L_1+L_2)+R_1L_2)s+(R_2RC(L_1+L_2)+R_1L_2RC+L_1L_2)s^2+L_1L_2RCs^3}{(R+R_2)(R_S+R_1)+[R_2RC(R_S+R_1)]s+((R+R_2)(L_1+L_2)+R_1L_2)s+(R_2RC(L_1+L_2)+R_1L_2RC+L_1L_2)s^2+L_1L_2RCs^3}$$

Divide top and bottom by $(R+R_2)(R_S+R_1)$ then it is clear that, in normalised form, in which I shall anticipate that δ can be made small compared to *a*: -

$$T(s)=\frac{as+bs^2+s^3}{1+(a+\delta)s+bs^2+s^3} \qquad \omega_N=\left(\frac{(R+R_2)(R_S+R_1)}{RCL_1L_2}\right)^{\frac{1}{3}} \qquad a=\frac{(R+R_2)(L_1+L_2)+R_1L_2}{(R+R_2)(R_S+R_1)}\omega_N \quad b=\frac{(R_2RC(L_1+L_2)+R_1L_2RC+L_1L_2)}{(R+R_2)(R_S+R_1)}\omega_N^2$$

$$\delta=\frac{R_2RC}{(R+R_2)}\omega_N$$

Appendix 3: the inverse of a 3 × 3 matrix

Given $V_i = Z_{ij} I_j$ then $I_j = \dfrac{|\mathbf{Z}[j]|}{|\mathbf{Z}|}$ where $\mathbf{Z}[j]$ = the **Z** matrix with the j^{th} column replaced by V_i.

The determinant is moved to the left hand side for simplicity: -

For $\boldsymbol{I_1}$ the V_i go in the first column: -

$$|\mathbf{Z}|I_1 = \begin{vmatrix} V_1 & Z_{12} & Z_{13} \\ V_2 & Z_{22} & Z_{23} \\ V_3 & Z_{32} & Z_{33} \end{vmatrix} = V_1(Z_{22}Z_{33} - Z_{23}Z_{32}) + Z_{12}(Z_{23}V_3 - V_2Z_{33}) + Z_{13}(V_2Z_{32} - Z_{22}V_3)$$
$$= (Z_{22}Z_{33} - Z_{23}Z_{32})V_1 + (Z_{13}Z_{32} - Z_{12}Z_{33})V_2 + (Z_{12}Z_{23} - Z_{13}Z_{22})V_3$$
$$= Y_{11}V_1 + Y_{12}V_2 + Y_{13}V_3$$

Second column: -

$$|\mathbf{Z}|I_2 = \begin{vmatrix} Z_{11} & V_1 & Z_{13} \\ Z_{21} & V_2 & Z_{23} \\ Z_{31} & V_3 & Z_{33} \end{vmatrix} = Z_{11}(V_2Z_{33} - Z_{23}V_3) + V_1(Z_{23}Z_{31} - Z_{21}Z_{33}) + Z_{13}(Z_{21}V_3 - V_2Z_{31})$$
$$= (Z_{23}Z_{31} - Z_{21}Z_{33})V_1 + (Z_{11}Z_{33} - Z_{13}Z_{31})V_2 + (Z_{13}Z_{21} - Z_{11}Z_{23})V_3$$
$$= Y_{21}V_1 + Y_{22}V_2 + Y_{23}V_3$$

Third column: -

$$|\mathbf{Z}|I_3 = \begin{vmatrix} Z_{11} & Z_{12} & V_1 \\ Z_{21} & Z_{22} & V_2 \\ Z_{31} & Z_{32} & V_3 \end{vmatrix} = Z_{11}(Z_{22}V_3 - V_2Z_{32}) + Z_{12}(V_2Z_{31} - Z_{21}V_3) + V_1(Z_{21}Z_{32} - Z_{22}Z_{31})$$
$$= (Z_{21}Z_{32} - Z_{22}Z_{31})V_1 + (Z_{12}Z_{31} - Z_{11}Z_{32})V_2 + (Z_{11}Z_{22} - Z_{12}Z_{21})V_3$$
$$= Y_{31}V_1 + Y_{32}V_2 + Y_{33}V_3$$

First row: -	Second row: -	Third row: -
$Y_{11} = (Z_{22}Z_{33} - Z_{23}Z_{32})$	$Y_{21} = (Z_{23}Z_{13} - Z_{21}Z_{33})$	$Y_{31} = (Z_{21}Z_{32} - Z_{22}Z_{31})$
$Y_{12} = (Z_{13}Z_{32} - Z_{12}Z_{33})$	$Y_{22} = (Z_{11}Z_{33} - Z_{13}Z_{31})$	$Y_{32} = (Z_{12}Z_{31} - Z_{11}Z_{32})$
$Y_{13} = (Z_{12}Z_{23} - Z_{13}Z_{22})$	$Y_{23} = (Z_{13}Z_{21} - Z_{11}Z_{23})$	$Y_{33} = (Z_{11}Z_{22} - Z_{12}Z_{21})$

The determinant is: -

$$|\mathbf{Z}| = \begin{vmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{vmatrix} = Z_{11}(Z_{22}Z_{33} - Z_{21}Z_{32}) + Z_{12}(Z_{23}Z_{31} - Z_{21}Z_{33}) + Z_{13}(Z_{21}Z_{32} - Z_{22}Z_{31})$$

Three-stage ratio transformers

1. Introduction

The in-phase accuracy of a two-stage ratio transformer is not quite sufficient for the first decade of a low frequency bridge. With the addition of a second energising core and winding, however, something wonderful happens. The input impedance of the ratio winding is boosted further so that it can be connected directly to a transfer standard resistor with negligible error. Also, the in-phase ratio accuracy (from primary to secondary) becomes fourth order with respect to frequency and high accuracy can be achieved at half power supply frequency (25 or 30Hz). The transformer requires fewer turns and the result is an elegant overall bridge design.

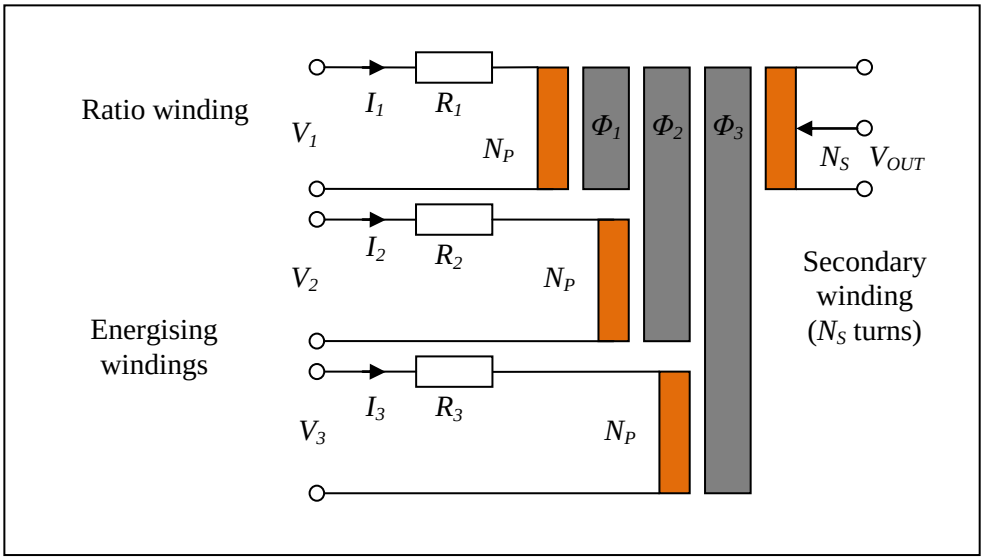


Fig. 1.1 A basic three-stage ratio transformer

Analysis of the three-stage transformer network reveals the same structure as the three-stage high-pass filter. For a 1:1 transformer the equivalent circuit is: -

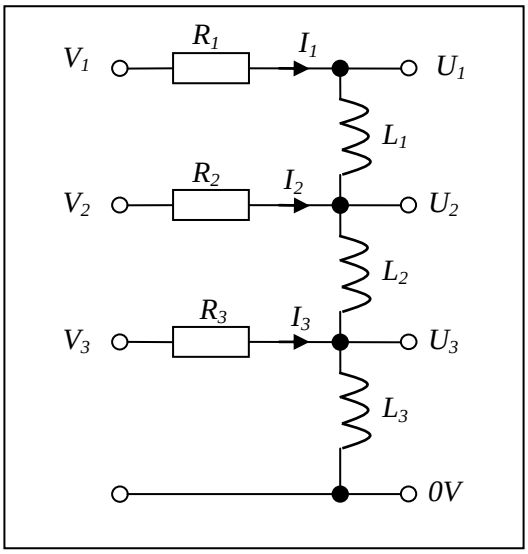


Fig. 1.2 The equivalent circuit for a three stage voltage transformer

One can re-use the matrix analysis from the monograph “Three-stage filters” [1] with symbol substitutions: -

$$R_1 \rightarrow sL_1 \quad \text{etc} \qquad\qquad X_1 \rightarrow R_1 \quad \text{etc}$$

1. Part 2, monograph 2: “Three-stage filters”. See appendices 1, 2 and 3.

The input impedance matrix equation becomes: -

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} sL_1 + sL_2 + sL_3 + R_1 & sL_2 + sL_3 & sL_3 \\ sL_2 + sL_3 & sL_2 + sL_3 + R_2 & sL_3 \\ sL_3 & sL_3 & sL_3 + R_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} \qquad \text{or} \qquad \mathbf{V} = \mathbf{ZI}$$

The inverse matrix allows us to calculate input currents and impedances: $\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$

Define **Y** such that: $\mathbf{Y} = |\mathbf{Z}|\mathbf{Z}^{-1} \quad \Rightarrow \quad \mathbf{I} = \dfrac{\mathbf{Y}}{|\mathbf{Z}|}\mathbf{V}$

The inverse matrix can be derived using Cramer's rule (see "Three-stage filters" Appendix 1 [1]) which, with the symbol substitutions, supplies the components: -

$$Y_{11} = s^2 L_2 L_3 + (sL_2 + sL_3)R_3 + sL_3 R_2 + R_2 R_3$$
$$Y_{12} = -s^2 L_2 L_3 - (sL_2 + sL_3)R_3 \qquad Y_{11} + Y_{12} + Y_{13} = R_2 R_3$$
$$Y_{13} = -sL_3 R_2$$
$$Y_{21} = -s^2 L_2 L_3 - (sL_2 + sL_3)R_3$$
$$Y_{22} = (sL_1 + sL_2)sL_3 + sL_3 R_1 + (sL_1 + sL_2 + sL_3)R_3 + R_1 R_3 \qquad Y_{21} + Y_{22} + Y_{23} = sL_1 R_3 + R_1 R_3$$
$$Y_{23} = -s^2 L_1 L_3 - sL_3 R_1$$
$$Y_{31} = -sL_3 R_2$$
$$Y_{32} = -s^2 L_1 L_3 - sL_3 R_1$$
$$Y_{33} = sL_1(sL_2 + sL_3) + (sL_2 + sL_3)R_1 + (sL_1 + sL_2 + sL_3)R_2 + R_1 R_2$$

$$Y_{31} + Y_{32} + Y_{33} = s^2 L_1 L_2 + (L_2 R_1 + L_1 R_2 + L_2 R_2)s + R_1 R_2$$

The determinant is (see "Three-stage filters" appendix 2): -

$$|\mathbf{Z}| = R_1 R_2 R_3 + s\{(L_1 + L_2 + L_3)R_2 R_3 + (L_2 + L_3)R_1 R_3 + L_3 R_1 R_2\} + s^2\{L_2 L_3 R_1 + L_1(L_2 + L_3)R_3 + (L_1 + L_2)L_3 R_2\} + s^3 L_1 L_2 L_3$$

To derive the transfer function one needs the output impedance matrix. Swapping symbols again one finds: -

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} sL_1 + sL_2 + sL_3 & sL_2 + sL_3 & sL_3 \\ sL_2 + sL_3 & sL_2 + sL_3 & sL_3 \\ sL_3 & sL_3 & sL_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} \qquad \text{or} \qquad \mathbf{U} = \mathbf{XI}$$

From above: $\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$ so that: $\mathbf{U} = \mathbf{XZ}^{-1}\mathbf{V} = \mathbf{TV}$

Where $\mathbf{T}$ is the transfer function matrix: $\mathbf{T} = \mathbf{XZ}^{-1}$

Define: $\boldsymbol{\Omega} = |\mathbf{Z}|\mathbf{XZ}^{-1}$ so that $\mathbf{U} = \dfrac{\boldsymbol{\Omega}}{|\mathbf{Z}|}\mathbf{V}$ i.e. $\mathbf{T} = \dfrac{\boldsymbol{\Omega}}{|\mathbf{Z}|}$ and $\boldsymbol{\Omega} = \mathbf{XY}$

To calculate the output, in terms of the input, one only needs the relevant components of $\boldsymbol{\Omega}$.

1. Part 2, monograph 2: "Three-stage filters".

1.1 The effects of follower errors, noise, loading and equalisation of windings

The effects of follower errors, noise and loading of the energising winding and ratio winding were analysed in some detail in the monograph "Two-stage IVDs and RTs" [1]. Equalisation of the most significant decade and subsequent decades was also analysed in the monograph "Single-stage IVDs and RTs". In the case of a three-stage structure detailed (mathematical) analysis would require the minimum of 4 × 4 matrices and the algebra becomes complicated.

Fortunately one can apply the same rules of thumb with a high degree of confidence – the first energising winding "bootstraps" the second and the two together bootstrap the ratio windings. Once a design is near complete, based on these principles, it could then be worth the effort to simulate it with suitable software. One can then vary parameters in order to optimise the design, especially if the device is to be manufactured in volume at the lowest possible cost.

With loading, for example, the flux in the first energising winding is typically a minimum of 99% of that required for the applied voltage. The missing flux is due to the voltage lost across the winding resistance and the small current flowing in the secondary. The same applies to the second energising core (99% of the deficit) so that the combined flux is 99.99%. The currents flowing in the ratio windings (primary and secondary) are correspondingly much lower, supplying 99.9% of the deficit so that the total flux in all three cores is 99.99999% and a ratio error of <100ppb is readily achieved.

Confirmation of the simulation results with tests on a prototype is also fairly simple. The effect of follower error and noise, for example, can be measured by injecting a small signal with a transformer. Similarly, the effect of loading can be observed by temporary removal.

The rules of thumb are repeated here for convenience: -

a). One must avoid loading the ratio secondary, especially with resistance. An inductive load is less of a problem (small quadrature error) and it is possible to include an extra strand for the purposes of equalising subsequent decade ropes with negligible error. The auxiliary core needs to be large enough and with a sufficient number of turns to present a sufficiently high inductive load.

b). One can load the first and second energising cores with little effect on accuracy – e.g. for energising a subsequent IVD or RT device for decades 3 and 4. See, for example, the monograph "An F18 type ratio transformer bridge" [2].

c). The contribution to noise and error due to the followers, via the energising windings, can be made negligible. The follower noise current, however, flows through the source resistance and the voltage generated is added to the Johnson noise of the source. If the followers are used to drive only the energising windings it is necessary, therefore, to set the noise resistance considerably higher than the source resistance (lower noise current but higher noise voltage).

d). If the requirement is for a "one-off" transformer one may as well spend the cash on larger, high permeability cores (in the required proportions). One can then use the thickest wire gauges (lowest resistance) that will comfortably fit in neat, complete and uniform layers. The result will be overkill, in terms of accuracy, and a little larger than necessary. Size 11b (SM100) for the energising cores and a 11c (SM200) for the ratio core are recommended. See example calculation section 3.3.1 (later) and [2] for more detail.

1. Part 3, monograph 3: "Two-stage IVDs and RTs".
2. Part 3, monograph 7: "An F18 type ratio transformer bridge".

2. Active drive for all three primary windings

As with two-stage IVDs and transformers one could drive all three primary windings with no stability problems**.** The accuracy, as well as noise, is then likely to be limited by the followers. The routing of currents to the energising windings is very important. No energising current should flow in a section of wire between the inverting input of the follower and the ratio winding. Depicted diagrammatically: -

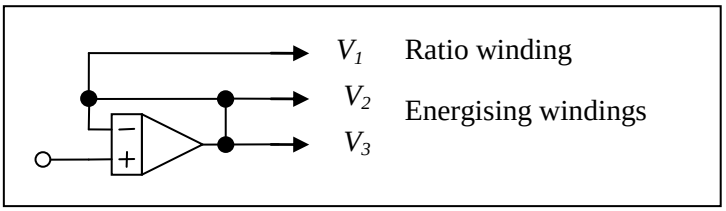


Fig. 2.1 Connection order for high accuracy

Resistance bridges based on this approach are versatile and relatively simple to construct. See the monograph "An F17 type ratio transformer bridge" by the same author [1]. If one wishes to achieve maximum performance, especially in terms of noise, however, it is better to drive only the energising windings although this again raises the issue of stability (see section 3).

2.1 Input currents and impedances

If the source impedance is very low, compared to the winding resistances, the analysis is fairly simple and reveals the main advantages of the three-stage "bootstrap" principle. If one assumes that all three inputs are connected to the same voltage source then: -

$$I_1 = \frac{(Y_{11}+Y_{12}+Y_{13})}{|\mathbf{Z}|}V_{IN} = \frac{R_2R_3}{|\mathbf{Z}|}V_{IN}$$

$$I_2 = \frac{(Y_{21}+Y_{22}+Y_{23})}{|\mathbf{Z}|}V_{IN} = \frac{sL_1R_3+R_1R_3}{|\mathbf{Z}|}V_{IN}$$

$$I_3 = \frac{(Y_{31}+Y_{32}+Y_{33})}{|\mathbf{Z}|}V_{IN} = \frac{s^2L_1L_2+s(L_2R_1+L_1R_2+L_2R_2)+R_1R_2}{|\mathbf{Z}|}V_{IN}$$

With the determinant: -

$$|\mathbf{Z}| = R_1R_2R_3 + s\{(L_1+L_2+L_3)R_2R_3+(L_2+L_3)R_1R_3+L_3R_1R_2\} + s^2\{L_2L_3R_1+L_1(L_2+L_3)R_3+(L_1+L_2)L_3R_2\} + s^3L_1L_2L_3$$

At high frequency, relative to the winding characteristic frequencies, to a good approximation: -

$$|s| >> \frac{R}{L} \;\Rightarrow\; |\mathbf{Z}| \approx s^3L_1L_2L_3 \quad I_1 \approx \frac{R_2R_3}{s^3L_1L_2L_3}V_{IN} \quad I_2 \approx \frac{R_3}{s^2L_2L_3}V_{IN} \quad \text{and} \quad I_3 \approx \frac{1}{sL_3}V_{IN}$$

The resulting input impedances are: -

$$Z_1 \approx sL_1 \times \frac{sL_2}{R_2} \times \frac{sL_3}{R_3} \qquad Z_2 \approx sL_2 \times \frac{sL_3}{R_3} \qquad \text{and} \qquad Z_3 \approx sL_3$$

The impedance, Z_3, of the first energising winding is, as expected, that of a simple inductor. The impedance of the second energising winding is boosted by a large factor and real (a large negative resistance), exactly as in the case of a two stage inductor. The ratio winding impedance is boosted even more and, usefully, is mainly imaginary. If the source impedance is resistive then the loading effect is mainly a quadrature error.

1. Part 3, monograph 6: "An F17 type ratio transformer bridge".

The next most significant component is the real part. The full expression is: -

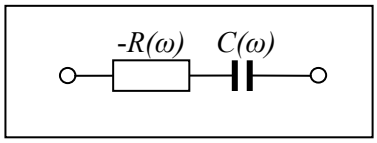
$$Z_1 = \frac{R_1R_2R_3 + s\{(L_1+L_2+L_3)R_2R_3 + (L_2+L_3)R_1R_3 + L_3R_1R_2\} + s^2\{L_2L_3R_1 + L_1(L_2+L_3)R_3 + (L_1+L_2)L_3R_2\} + s^3L_1L_2L_3}{R_2R_3}$$

At high frequency, to a good approximation, retaining both real and imaginary components: -

$$|s| >> \frac{R}{L} \Rightarrow \qquad Z_1 \approx \frac{s^3L_1L_2L_3 + s^2(L_2L_3R_1 + L_1(L_2+L_3)R_3 + (L_1+L_2)L_3R_2)}{R_2R_3}$$

The largest term is negative imaginary and the next largest is negative real. The input impedance, apart from the frequency dependence, looks like a small capacitor in series with a large negative resistor: -

$$Z_1 \approx -R(\omega) - j\frac{1}{C(\omega)}$$



2.2 The transfer function

The transformer output, assuming a 1:1 ratio of turns is: $V_{OUT} = U_1 = \frac{\Omega_{11}V_1 + \Omega_{12}V_2 + \Omega_{13}V_3}{|\mathbf{Z}|}$

If all three inputs are connected to the same low resistance source**:** $V_1 = V_2 = V_3 = V_{IN}$

The main output is then: -

$$V_{OUT} = U_1 = \frac{\Omega_{11} + \Omega_{12} + \Omega_{13}}{|\mathbf{Z}|}V_{IN}$$

One can keep the algebra to the minimum by noting that, in subscript notation: $\Omega_{ij} = \sum_k X_{ik}Y_{kj}$

The first subscript numbers the row and the second the column so that: $\sum_j \Omega_{1j} = \sum_j \sum_k X_{1k}Y_{kj}$

It is possible to swap the order of addition: $\sum_j \Omega_{1j} = \sum_k X_{1k}\sum_j Y_{kj} = \sum_k X_{1k}(Y_{k1} + Y_{k2} + Y_{k3})$

$$\Rightarrow \quad T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{X_{11}(Y_{11}+Y_{12}+Y_{13}) + X_{12}(Y_{21}+Y_{22}+Y_{23}) + X_{13}(Y_{31}+Y_{32}+Y_{33})}{|\mathbf{Z}|}$$

From above: -

$Y_{11} + Y_{12} + Y_{13} = R_2R_3$

$Y_{21} + Y_{22} + Y_{23} = sL_1R_3 + R_1R_3$

$Y_{31} + Y_{32} + Y_{33} = s^2L_1L_2 + (L_2R_1 + L_1R_2 + L_2R_2)s + R_1R_2$

$X_{11} = sL_1 + sL_2 + sL_3$

$X_{12} = sL_2 + sL_3$

$X_{13} = sL_3$

$$T(s) = \frac{((L_1+L_2+L_3)R_2R_3 + R_1R_3(L_2+L_3) + L_3R_1R_2)s + ((L_2R_3 + L_3R_3)L_1 + (L_2R_1 + L_1R_2 + L_2R_2)L_3)s^2 + L_1L_2L_3s^3}{|\mathbf{Z}|}$$

As if by magic the numerator replicates the determinant, apart from the first term. From above: -

$$|\mathbf{Z}| = R_1R_2R_3 + s\{(L_1+L_2+L_3)R_2R_3 + (L_2+L_3)R_1R_3 + L_3R_1R_2\} + s^2\{L_2L_3R_1 + L_1(L_2+L_3)R_3 + (L_1+L_2)L_3R_2\} + s^3L_1L_2L_3$$

Divide top and bottom by $R_1R_2R_3$ and multiply the coefficients by the natural frequency raised to the appropriate power to give the transfer function in normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{s^3 + bs^2 + as}{s^3 + bs^2 + as + 1} \quad \text{with} \quad \omega_N = \left(\frac{R_1R_2R_3}{L_1L_2L_3}\right)^{\frac{1}{3}}$$

$$a = \frac{\left((L_1+L_2+L_3)R_2R_3 + (L_2+L_3)R_1R_3 + L_3R_1R_2\right)}{R_1R_2R_3}\omega_N$$

$$b = \frac{L_2L_3R_1 + L_1(L_2+L_3)R_3 + (L_1+L_2)L_3R_2}{R_1R_2R_3}\omega_N^2$$

It is not surprising that the transfer function has the same form as a three-stage high-pass filter. Simulation shows that acceptable values for *a* and *b* are between 2 and 10. As with three-stage filters the lower the value the greater is the resonant peaking. Error analysis shows that, at high frequency, the in-phase error term is proportional to *b*. The parameter *a* contributes very little to the imaginary error (fifth order): -

To analyse for high frequency errors divide top and bottom by $s^3 + bs^2 + as$: -

$$T(s) = \left(1 + \frac{1}{s^3 + bs^2 + as}\right)^{-1}$$

Take out a factor of s^3 and apply an approximation, neglecting terms of order s^{-5} and higher [1]: -

$$|s| << 1 \;\Rightarrow\; T(s) \approx 1 - \frac{1}{s^3}\left(1 - \frac{b}{s}\right) = 1 - \frac{1}{s^3} + \frac{b}{s^4}$$

The in-phase error is now fourth order, in terms of frequency, compared to second order for a two-stage transformer. In more convenient form with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 - j\left(\frac{f_N}{f}\right)^3 + b\left(\frac{f_N}{f}\right)^4 \quad \text{with} \quad f_N = \frac{1}{2\pi}\left(\frac{R_1R_2R_3}{L_1L_2L_3}\right)^{\frac{1}{3}}$$

If the inductances and resistances are the same one obtains the values: $a = 6$ and $b = 5$. With these values peaking is minimal and the stability margin more than adequate. It is possible, in practice, to vary the resistances and inductances to reduce these values with savings, in terms of smaller and lower permeability cores, fewer turns and still retain accuracy at low operating frequency.

1. Method of approximation: $|z| << 1 \;\Rightarrow\; \frac{1}{1+z} = \frac{1-z}{1-z^2} \approx 1 - z$

2.3 Example calculation

With the extra bootstrapping effect one should be able to reduce the number of turns and use thicker wire, making it easier to construct. I shall use the same (medium) size cores as the example in the monograph "two-stage IVDs and RTs" [1] but this time with 200 turns, instead of 400. This should be fairly simple to construct by employing the BNNL winding technique [2].

All three cores type 7c mumetal ($\mu_R \approx 50{,}000$; see tables of data at the end of this monograph): -
OD = *60mm* ID = *36mm* axial = *15mm*.
The length of wire per turn on the first energising winding is approximately: *15 × 2 + 60 – 36 = 54mm*

Inner circumference ≈ 36mm × π ≈ 113mm can accommodate 200 turns in two layers of 1mm OD wire.
Outer circumference ≈ 60mm × π ≈ 188mm can accommodate 200 turns of 0.94mm OD in a single layer.
The outer circumference is the limiting dimension: the largest wire diameter that is a comfortable fit is 20SWG (enamelled) at 0.914mm diameter with a resistance of 26mΩm^{-1}.

The first energising winding requires: $200 \times 54mm \approx 11m$ of wire.
Resistance R_3 would be approx: $R_3 \approx 11m \times 26m\Omega m^{-1} \approx 0.29\Omega$
Permittance of all three cores: $A_L = 51\mu H/turn^2$
Inductance of all three windings: $L_1 = L_2 = L_3 = N^2 A_L = 200 \times 200 \times 51 \times 10^{-6} \approx 2H$

The second energising winding is similar but takes more wire. The first core is now slightly bigger due to the first energising winding and so the length of wire per turn is approximately: *15 × 2 +17×2+ 62 – 34 = 92mm*

With 200 turns the total length is $200 \times 92mm \approx 19m$
Resistance R_2 would be approx: $R_2 \approx 19m \times 26m\Omega m^{-1} \approx 0.49\Omega$ (i.e. very approximately twice R_3.)

The ratio winding consists of a rope with typically 20 strands wound 20 times (also BNNL) of the same wire (20SWG). Ten strands are connected in series for the ratio primary and the other ten for the secondary. With 20 strands the rope is approximately 5mm in diameter. The inner diameter is reduced slightly: -
Inner circumference ≈ 30mm × π ≈ 94mm can accommodate 20 turns in a single layer (4.7mm/turn).
Outer circumference ≈ 60mm × π ≈ 188mm can easily accommodate 20 turns also in a single layer.

As a ball park estimate the length of wire, for the ratio primary winding (200 turns), is roughly three times that of the first energising winding, partly because it is wound around all three cores and partly because the strands are twisted into a rope.

Resistance of 200 turns of ratio primary: $R_1 \approx 3R_3 \approx 0.87\Omega$

At 25Hz the bootstrapped input impedance of the ratio winding is: -

$$|Z_1| \approx \left| sL_1 \times \frac{sL_2}{R_2} \times \frac{sL_3}{R_3} \right| \approx (2\pi \times 25)^3 \times 2 \times \frac{2}{0.5} \times \frac{2}{0.3} \approx 207M\Omega$$

The source resistance (cable from the output of the followers: typically <0.1Ω) is, therefore, negligible.

N.B. This does not include the inter-winding capacitance which then becomes dominant.

It is interesting to compare the stability and accuracy with a two-stage actively driven transformer…

1. Part 3, monograph 3: "Two-stage IVDs and RTs". See section 1.1.
2. Part 3, monograph 1: "IVDs and RTs – the basics". For balanced no-net-loop (BNNL) see section 1.3.

The calculation need only be approximate: $L_1 = L_2 = L_3 \approx 2H \quad R_1 \approx 1\Omega \quad R_2 \approx 0.5\Omega$ and $R_3 \approx 0.3\Omega$

The transfer function (from the previous section) is, in the normalised form $(s = j\omega/\omega_N)$.

$$T(s) = \frac{s^3 + bs^2 + as}{s^3 + bs^2 + as + 1} \quad \text{with} \quad f_N = \frac{\omega_N}{2\pi} = \frac{1}{2\pi}\left(\frac{R_1R_2R_3}{L_1L_2L_3}\right)^{\frac{1}{3}} = \frac{1}{2\pi}\left(\frac{1\times 0.5\times 0.3}{2\times 2\times 2}\right)^{\frac{1}{3}} \approx 0.042Hz$$

$$a = \frac{\left((L_1 + L_2 + L_3)R_2R_3 + (L_2 + L_3)R_1R_3 + L_3R_1R_2\right)}{R_1R_2R_3}\omega_N$$
$$= \frac{\left((6)\times 0.5\times 0.3 + (4)\times 1\times 0.3 + 2\times 1\times 0.5\right)}{1\times 0.5\times 0.3} \times 2\pi \times 0.042 \approx 5.5$$

$$b = \frac{L_2L_3R_1 + L_1(L_2 + L_3)R_3 + (L_1 + L_2)L_3R_2}{R_1R_2R_3}\omega_N^2$$
$$= \frac{2\times 2\times 1 + 2\times (4)\times 0.3 + (4)\times 2\times 0.5}{1\times 0.5\times 0.3} \times (2\pi \times 0.042)^2 \approx 4.9$$

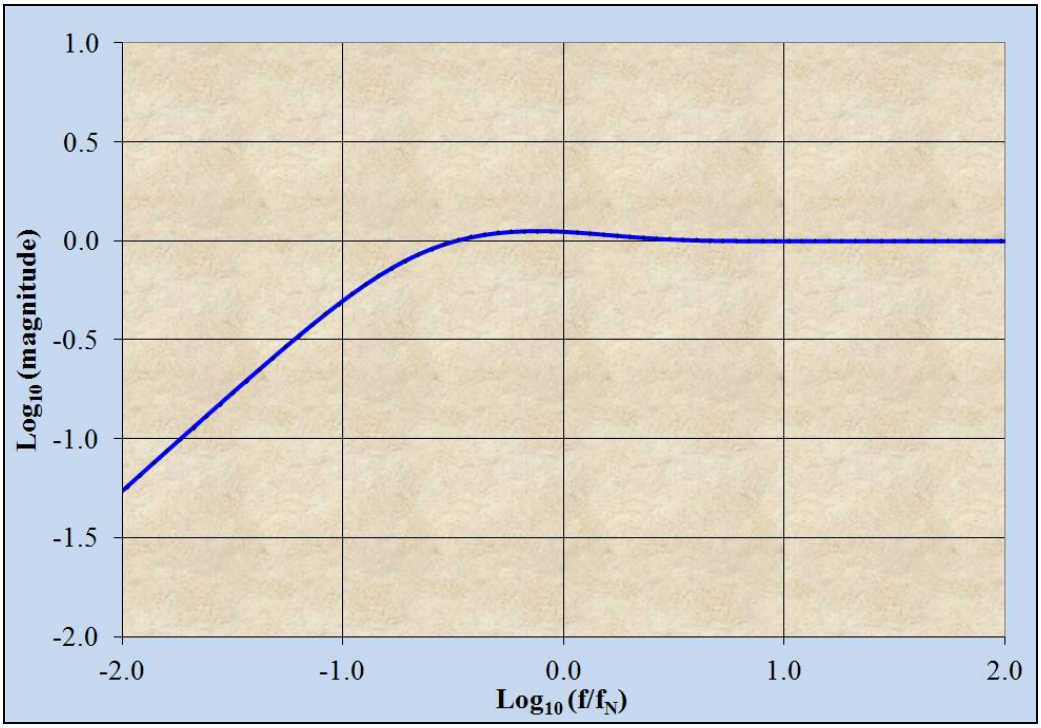


Fig. 2.3.1 A slight peak close to the natural frequency [1]

Transformer ratio accuracy at 25Hz is: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 - j\left(\frac{f_N}{f}\right)^3 + b\left(\frac{f_N}{f}\right)^4 \approx 1 - j4.7\times 10^{-9} + 3.9\times 10^{-11}$$

Both in-phase and quadrature error terms are truly negligible. The limit, in practice, is follower errors and flux leakage from the first energising core.

1. Spreadsheet: “Basic three-stage”.

3. Active drive for energising windings only

3.1 The importance of source resistance

Once again one can employ the filter circuit model. Note also the order of connections to the energising windings. The larger current, I_3, should not flow through any part of the connection to the second energising winding: -

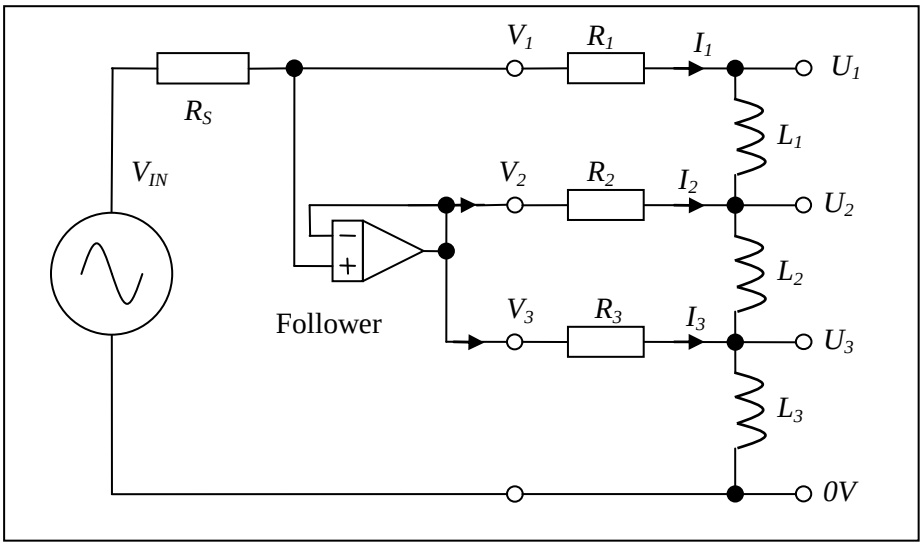


Fig. 3.1.1 Equivalent circuit for active drive for ratio windings only

The source resistance, R_S, results in some positive feedback. The feedback network consists of the energising windings, which has a high pass characteristic (two-stage), followed by the ratio winding, which has a low pass characteristic. The result is a band pass characteristic whose peak magnitude can approach unity depending on the relative cut-off frequencies. For adequate stability margin the high pass cut-off frequency needs to be higher than the low pass cut-off. Apart from the interaction between the high pass and low pass sections the cut-off frequencies are approximately: -

$$\omega_{HP} = \sqrt{\frac{R_2}{L_2}\frac{R_3}{L_3}} \qquad \omega_{LP} = \frac{R_S + R_1}{L_1} \qquad \text{required:} \quad \omega_{HP} > \omega_{LP} \quad \Rightarrow \quad \sqrt{\frac{R_2}{L_2}\frac{R_3}{L_3}} > \frac{R_S + R_1}{L_1}$$

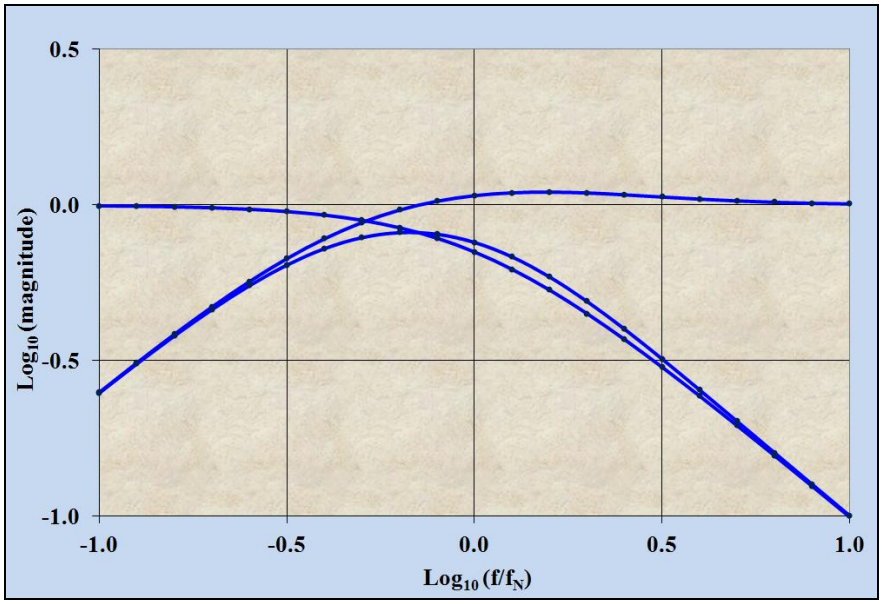


Fig. 3.1.2 Low-pass, high-pass and combined band-pass characteristic

Clearly the source resistance and ratio winding inductance determine the lowest natural frequency and, therefore, limit low frequency accuracy as the detailed analysis demonstrates.

For the circuit fig. 3.1.1, assuming an ideal voltage follower: -

$$V_1 = V_2 = V_3 \quad \text{and} \quad \mathbf{I} = \mathbf{Z}^{-1}\mathbf{V} \quad \Rightarrow \quad I_1 = \frac{Y_{11}V_1 + Y_{12}V_2 + Y_{13}V_3}{|\mathbf{Z}|} = \frac{R_2R_3}{|\mathbf{Z}|}V_1$$

$$\text{But } V_{IN} = V_1 + I_1R_S = V_1\left(1 + \frac{R_2R_3R_S}{|\mathbf{Z}|}\right) \text{ so that } V_1 = V_2 = V_3 = \frac{|\mathbf{Z}|}{|\mathbf{Z}| + R_2R_3R_S}V_{IN}$$

As above: $$V_{OUT} = U_1 = \frac{(\Omega_{11}V_1 + \Omega_{12}V_2 + \Omega_{13}V_3)}{|\mathbf{Z}|}$$

$$\Rightarrow \quad V_{OUT} = \frac{(\Omega_{11} + \Omega_{12} + \Omega_{13})}{|\mathbf{Z}|} \times \frac{|\mathbf{Z}|}{|\mathbf{Z}| + R_2R_3R_S}V_{IN}$$

Resulting in: $$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{\Omega_{11} + \Omega_{12} + \Omega_{13}}{|\mathbf{Z}| + R_2R_3R_S}$$

The numerator is, from section 2.2: -

$$T(s) = \frac{s\{(L_1 + L_2 + L_3)R_2R_3 + R_1R_3(L_2 + L_3) + L_3R_1R_2\} + s^2\{(L_2R_3 + L_3R_3)L_1 + (L_2R_1 + L_1R_2 + L_2R_2)L_3\} + s^3L_1L_2L_3}{|\mathbf{Z}| + R_2R_3R_S}$$

As above the determinant is: -

$$|\mathbf{Z}| = R_1R_2R_3 + s\{(L_1 + L_2 + L_3)R_2R_3 + (L_2 + L_3)R_1R_3 + L_3R_1R_2\} + s^2\{L_2L_3R_1 + L_1(L_2 + L_3)R_3 + (L_1 + L_2)L_3R_2\} + s^3L_1L_2L_3$$

The whole transfer function is, therefore: -

$$T(s) = \frac{s\{(L_1 + L_2 + L_3)R_2R_3 + R_1R_3(L_2 + L_3) + L_3R_1R_2\} + s^2\{(L_2 + L_3)R_3L_1 + (L_2R_1 + L_1R_2 + L_2R_2)L_3\} + s^3L_1L_2L_3}{(R_S + R_1)R_2R_3 + s\{(L_1 + L_2 + L_3)R_2R_3 + (L_2 + L_3)R_1R_3 + L_3R_1R_2\} + s^2\{L_2L_3R_1 + L_1(L_2 + L_3)R_3 + (L_1 + L_2)L_3R_2\} + s^3L_1L_2L_3}$$

Divide top and bottom by $(R_S + R_1)R_2R_3$ and multiply the coefficients by the natural frequency raised to the appropriate power to give the transfer function in normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{as + bs^2 + s^3}{1 + as + bs^2 + s^3} \quad \text{with} \quad \omega_N = \left(\frac{(R_S + R_1)R_2R_3}{L_1L_2L_3}\right)^{\frac{1}{3}}$$

$$a = \frac{(L_1 + L_2 + L_3)R_2R_3 + (L_2 + L_3)R_1R_3 + L_3R_1R_2}{(R_S + R_1)R_2R_3}\omega_N$$

$$b = \frac{L_2L_3R_1 + L_1(L_2 + L_3)R_3 + (L_1 + L_2)L_3R_2}{(R_S + R_1)R_2R_3}\omega_N^2$$

The result is very nearly the same as the previous case (see page 6) with two important differences. The resistance R_1 is replaced by $R_1 + R_S$ in the formula for the natural frequency and in the denominator for parameters *a* and *b*. As the source resistance increases, therefore, the natural frequency increases (lower accuracy) and the values of *a* and *b* decrease (less stability margin).

The reduced stability margin can be understood in terms of the increased cut-off frequency of the low-pass part of the feedback network causing increasing overlap with the high-pass characteristic of the energising windings.

This confirms the same trade-offs as with two-stage IVDs and transformers. It is useful here to recall and extend the rules of thumb for the active drive of two-stage transformers as the same principles apply. See the monograph "Two-stage IVDs and RTs" [1].

1. To improve stability add extra resistance in series with the energising winding(s) with a large capacitor in parallel (active drive "compensator"). In practice it is usually sufficient to add a compensator to only the first energising winding (R_S < 100Ω). If the source resistance is higher it may be necessary to add a compensator to both, though the analysis becomes more complicated.

2. To improve stability further employ a thicker (axially) and higher permeability top core and add extra turns to the ratio winding (around the top core only). This increases L_1 substantially.

3. The follower noise (via the energising windings) is much reduced. The follower current noise, however, flows through the source resistance and the voltage generated appears across the ratio primary. Best noise performance is achieved, therefore, when the noise resistance of the followers is significantly greater than the source resistance. This is usually done by operating the follower front end BJTs at a low operating current. See the monograph "Low noise BJT pre-amps" for more detail [2].

3.1.1 Example calculation

If one repeats the calculation of the previous example with the values:

$$L_1 = L_2 = L_3 \approx 2H \quad R_1 \approx 1\Omega \quad R_2 \approx 0.5\Omega \quad R_3 \approx 0.3\Omega \text{ and } \quad R_S \approx 100\Omega$$

$$f_N = \frac{\omega_N}{2\pi} = \frac{1}{2\pi}\left(\frac{(R_S + R_1)R_2R_3}{L_1L_2L_3}\right)^{\frac{1}{3}} = \frac{1}{2\pi}\left(\frac{101\times 0.5\times 0.3}{2\times 2\times 2}\right)^{\frac{1}{3}} \approx 0.2Hz$$

The natural frequency is increased by almost a factor of five (cube root of 101) compared to the previous example, entirely due to the increased source resistance. This is not too bad, however, as the in-phase error, at the lowest operating frequency of 25Hz, is of the order $4b$ ppb.

$$a = \frac{\left((L_1 + L_2 + L_3)R_2R_3 + (L_2 + L_3)R_1R_3 + L_3R_1R_2\right)}{(R_S + R_1)R_2R_3}\omega_N$$
$$= \frac{\left((6)\times 0.5\times 0.3 + (4)\times 1\times 0.3 + 2\times 1\times 0.5\right)}{101\times 0.5\times 0.3}\times 2\pi\times 0.195 \approx 0.26$$

$$b = \frac{L_2L_3R_1 + L_1(L_2 + L_3)R_3 + (L_1 + L_2)L_3R_2}{R_1R_2R_3}\omega_N^2$$
$$= \frac{2\times 2\times 1 + 2\times (4)\times 0.3 + (4)\times 2\times 0.5}{101\times 0.5\times 0.3}\times (2\pi\times 0.195)^2 \approx 1.1$$

Computer modelling shows that these values are on the low side, resulting in a large resonant peak close to the resonant frequency – stability is the main problem.

Fortunately there is a simple way to maintain a low natural frequency and improve stability margin: rule of thumb 1(above) with an active drive compensator [1].

1. Part 3, monograph 3: "Two-stage IVDs and RTs". See section 4.2.
2. Part 5, monograph 2: "Low noise BJT pre-amps".

3.2 Active drive with compensator

The basic idea is that at low frequency (below the natural frequency) the compensator resistance dominates, ensuring stability but, as frequency increases the impedance of the capacitor drops so that, at the operating frequency, the bootstrapping effect and high accuracy is restored. As a ball park estimate (very approximately) the impedance of the capacitor is of the same order as the compensator resistor at the natural frequency of the transformer: -

For stability: $R \approx R_S$ and, at the natural frequency: $X_C = \dfrac{1}{\omega_N C} \approx R$

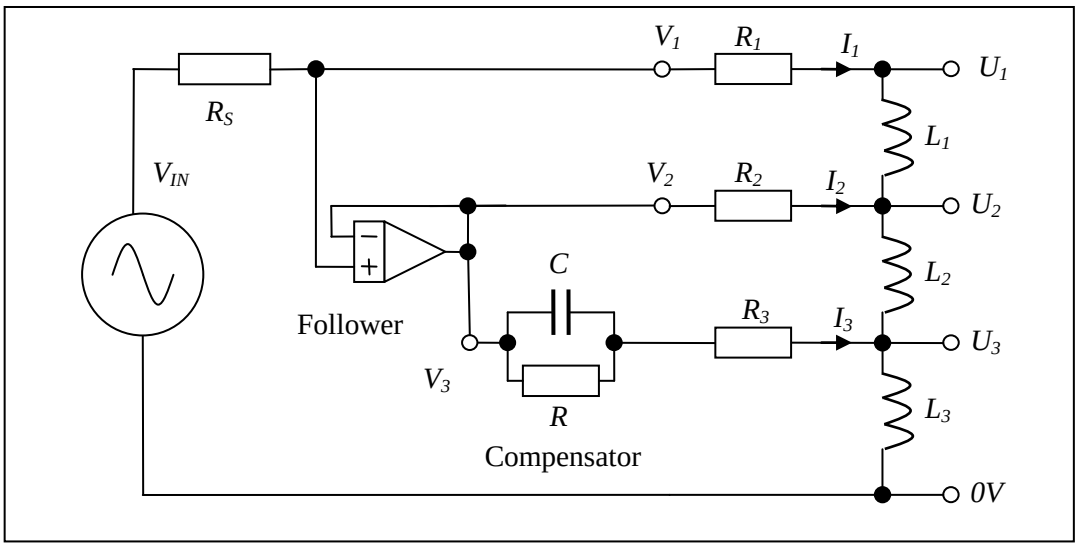


Fig. 3.2.1 Active drive with compensator (first stage only)

At the operating frequency the reduced current in the first energising winding (as a result of the extra series impedance) results in a slightly lower voltage induced in the second energising winding. The small error is thus largely corrected by a slight increase in current in the second energising winding, as long as one keeps R_2 small. The high input impedance and accuracy of the ratio winding is hardly affected by the compensator R and C. The transfer function can be derived by the substitution: -

$$R_3 \rightarrow Z_3 = R_3 + \frac{R}{1+sRC} \quad \text{or, in more convenient form:} \quad R_3 \rightarrow Z_3 = \frac{R + R_3 + sRR_3C}{1+sRC}$$

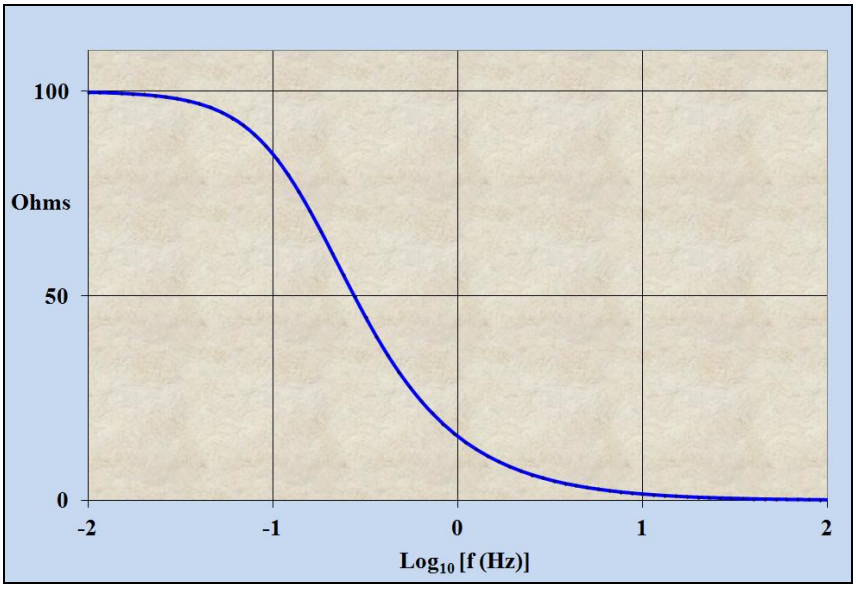


Fig. 3.2.2 Impedance magnitude of a typical compensator vs frequency (R_S = 100Ω) [1]

1. Spreadsheet: “Compensator impedance”

The algebra is a bit messy and relegated to appendix 1. The main results are, in normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{as + bs^2 + cs^3 + s^4}{1 + (a+\delta)s + bs^2 + cs^3 + s^4} \qquad \text{with} \qquad \omega_N = \left(\frac{(R_S + R_1)R_2(R + R_3)}{L_1 L_2 L_3 RC} \right)^{\frac{1}{4}}$$

$$a = \frac{[(L_1 + L_2 + L_3)R_2 + R_1(L_2 + L_3)](R + R_3) + L_3 R_1 R_2}{(R_S + R_1)R_2(R + R_3)} \omega_N$$

$$b = \frac{(L_2 + L_3)(R + R_3)L_1 + (L_1 + L_2 + L_3)R_2 RR_3 C + R_1 RR_3 C(L_2 + L_3) + (L_2 R_1 + L_1 R_2 + L_2 R_2)L_3 + (L_3 R_1 R_2 RC)}{(R_S + R_1)R_2(R + R_3)} \omega_N^2$$

$$c = \frac{(L_2 + L_3)RR_3 CL_1 + L_1 L_2 L_3 + (L_2 R_1 + L_1 R_2 + L_2 R_2)L_3 RC}{(R_S + R_1)R_2(R + R_3)} \omega_N^3 \qquad \delta = \frac{RR_3 C}{R + R_3} \omega_N$$

3.2.1 Error analysis

Divide top and bottom by $as + bs^2 + cs^3 + s^4$: $\quad T(s) \approx \left(1 + \frac{1 + \delta s}{as + bs^2 + cs^3 + s^4} \right)^{-1}$

A very good approximation is [1]: -

$$|s| << 1 \;\Rightarrow\; T(s) \approx \left(1 - \frac{1 + \delta s}{as + bs^2 + cs^3 + s^4} \right)$$

Take out a factor of s^4 and repeat the approximation for the small error term: -

$$T(s) \approx 1 - \frac{1 + \delta s}{s^4} \left(1 - \frac{c}{s} - \frac{b}{s^2} - \frac{a}{s^3} \right)$$

Terms of s^{-6} and smaller are negligible so that, to a very good approximation: -

$$\delta c << 1 \Rightarrow \qquad T(s) \approx 1 - \frac{1 + \delta s}{s^4} \left(1 - \frac{c}{s} \right) = 1 - \frac{\delta}{s^3} - \frac{1}{s^4} + \frac{c}{s^5}$$

This is sufficient to retain both in-phase (real) and quadrature (imaginary) components. Note that the term cs^{-5} is retained because it may be comparable to δs^{-3}**.** In more convenient form, with frequency in Hz: -

$$f >> f_N \Rightarrow \quad T(f) \approx 1 - j\left\{ \delta \left(\frac{f_N}{f} \right)^3 + c \left(\frac{f_N}{f} \right)^5 \right\} - \left(\frac{f_N}{f} \right)^4 \qquad \text{with} \qquad f_N = \frac{1}{2\pi} \left(\frac{(R_S + R_1)R_2(R + R_3)}{L_1 L_2 L_3 RC} \right)^{\frac{1}{4}}$$

In practice the compensator resistance is much greater than the winding resistance so that a convenient approximation is: -

$$R >> R_3 \;\Rightarrow\; \delta = \frac{RR_3 C}{R + R_3} \omega_N \approx R_3 C \omega_N$$

1. Method of approximation: $|z| << 1 \;\Rightarrow\; \frac{1}{1+z} = \frac{1-z}{1-z^2} \approx 1 - z$

3.3 Extra turns around the top core

Adding extra turns to the top core can be represented according to the usual convention: windings horizontally adjacent to cores are wound around those cores, at least in part. In this case N_P turns of the ratio winding rope are wound around the top core (only) and then N_P turns around all three cores. The bootstrapping voltage induced in the ratio winding is the same as before. Ratio accuracy is also maintained as the secondary winding is part of the same rope as the primary and shares the same flux.

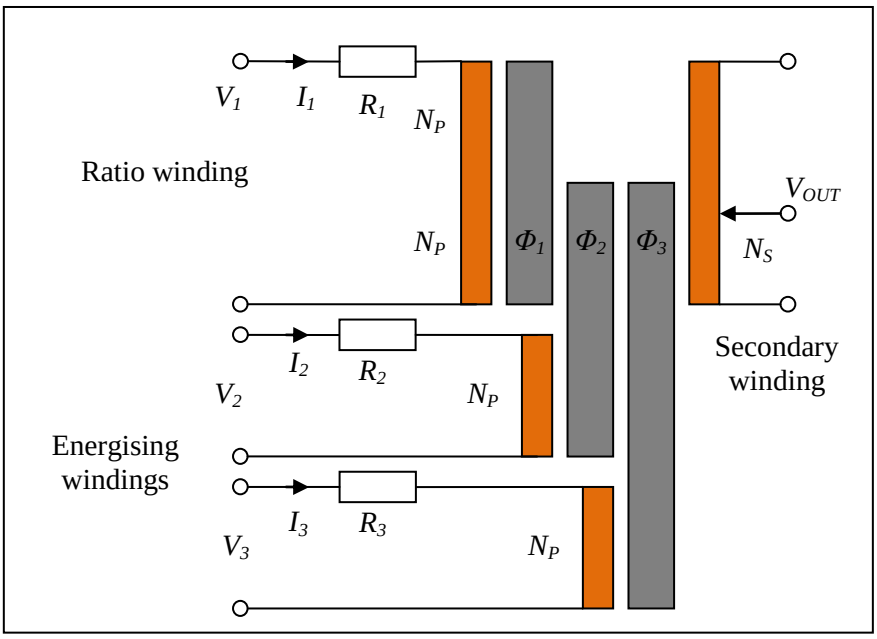


Fig. 3.3.1 Extra turns around the top (ratio) core

The overall effect is to increase the inductance L_1 by a factor of four, as the following demonstrates: -

According to Ohm's, Kirchhoff's and Faraday's laws in the complex representation $(s = j\omega \equiv d/dt)$: -

$$V_1 = I_1R_1 + s\left(2N_P\Phi_1 + N_P\Phi_2 + N_P\Phi_3\right) \qquad V_2 = I_2R_2 + s\left(N_P\Phi_2 + N_P\Phi_3\right) \text{ etc.}$$

Note the extra factor of two is because of the extra turns around the top core. The expressions for V_2 and V_3 remain unaltered. Similarly, for ideal magnetic circuits, the magneto-motive force in the top core is also doubled so that the fluxes are: -

$$\Phi_1 = 2N_PI_1A_{L1} \qquad \Phi_2 = \left(N_PI_1 + N_PI_2\right)A_{L2} \qquad \Phi_3 = \left(N_PI_1 + N_PI_2 + N_PI_3\right)A_{L2}$$

A_{L1} = permittance of core 1. A_{L2} = permittance of core 2. A_{L3} = permittance of core 3.

From the above: $V_1 = I_1R_1 + s\left(4N_P^2I_1A_{L1} + \left(N_P^2I_1A_{L2} + N_P^2I_2A_{L2}\right) + \left(N_P^2I_1A_{L3} + N_P^2I_2A_{L3} + N_P^2I_3A_{L3}\right)\right)$

The winding inductances are: $L_1 = N_P^2A_{L1}$ and $L_2 = N_P^2A_{L2}$ etc

Then: $V_1 = I_1\left(R_1 + 4sL_1 + sL_2 + sL_3\right) + I_2\left(sL_2 + sL_3\right) + I_3\left(sL_3\right)$ etc

The rest of the analysis is exactly the same as before (sections 1, 2 and 3) but with the substitution: $L_1 \rightarrow 4L_1$.One may as well retain the formulae for the transfer functions and input impedances and calculate the inductance L_1 with twice the number of turns (i.e. the actual number of turns around the top core): $L_1 = \left(2N_P\right)^2A_{L1}$.

3.3.1 Example calculation

The analysis and previous example (section 3.1) point the way to go – with a higher source resistance (up to 100Ω) the only option is to significantly increase all inductances (especially L_1) and apply at least one compensator. The top (ratio) core is increased to a size 11c SM200 with both energising cores also size 11 but only thickness b (SM100) and, to make winding practicable, one may as well stick to 200 turns. The larger cores also allow one to use thicker wire (1.5mm is about the thickest that can be easily wound) and accommodate the extra turns around the top core. When calculating total length of the energising windings one adds a bit extra (about 5%) because of its thickness and bend radius. Using the thicker wire on the first energising winding is required, not to reduce the resistance but to fill the larger core – creating, as near as possible, a uniform sheet of current around the surface of the core.

First energising core type 11b (see appendix 2): -
OD = 112*mm* ID = *66mm* axial = *19mm*.
The length of wire per turn on the first energising winding is approximately: *19 × 2 + 112 – 66 = 84mm*

Inner circumference ≈ 66mm × π ≈ 207mm can accommodate 200 turns in two layers of 2mm OD wire.
Outer circumference ≈ 112mm × π ≈ 352mm can accommodate 200 turns of 1.75mm OD in a single layer.
The outer circumference is the limiting dimension: the largest wire diameter that is a comfortable fit is 1.5mm (enamelled) with a resistance of 9.8mΩm^{-1} (extrapolated from 16SWG of 1.626mm at 8.3mΩm^{-1} because the metric size is more readily available from stock).

The first energising winding requires (+5%): $200 \times 88mm \approx 18m$ of wire.
Resistance R_3 would be approx: $R_3 \approx 18m \times 9.8m\Omega m^{-1} \approx 180m\Omega$

The second energising winding is around both 11b cores with a slightly reduced ID and increased OD

Inner circumference ≈ 60mm × π ≈ 188mm can accommodate 200 turns in two layers of 1.8mm OD wire.
Outer circumference ≈ 114mm × π ≈ 358mm can accommodate 200 turns of 1.8mm OD in a single layer.
One may as well use the same wire size (i.e. 1.5mm - any larger is tricky because of the radius of bends).

OD = 114*mm* ID = *60mm* combined axial = *19 + 3 + 19 = 41mm*.
The length of wire per turn on the second energising winding is approximately: *41 × 2 + 114 – 60 = 136mm*
The second energising winding requires (+5%): $200 \times 141mm \approx 28m$ of wire.
Resistance R_2 would be approx: $R_2 \approx 28.5m \times 9.8m\Omega m^{-1} \approx 0.3\Omega$

Permittance of both energising cores: $A_L = 136\mu H / turn^2$
Inductance of both energising windings: $L_2 = L_3 = N^2 A_L = 200 \times 200 \times 136 \times 10^{-6} \approx 5.4H$

The ratio winding consists of a rope with typically 20 strands wound 20 times around the top core only and another 20 times around all three cores – a total of 40 turns around the top core (see section 4.3.3). Ten of the strands are connected in series and used for the primary, the other 10 for the secondary. This would usually require thinner wire (about 0.5mm diameter is comfortable). With 20 strands the rope is approximately 4mm in diameter. The inner diameter is again reduced slightly: -
Inner circumference ≈ 57mm × π ≈ 179mm can accommodate 40 turns in a single layer (4.5mm/turn available).
The outer circumference is not a problem and can easily accommodate 40 turns.

Ratio core type 11c (see appendix 2): -
OD = 112*mm* ID = *66mm* axial = *23mm*.
The length of wire per turn on the top core only is approximately: *23 × 2 + 112 – 66 = 92mm*

The average inner and outer diameters remain about the same but the axial length increases: -
OD = 114*mm* ID = *60mm* combined axial = *19 + 3 + 19 +3 +23= 67mm*.
The length of rope per turn of the ratio winding around all three cores is approximately: -
67 × 2 + 114 – 60 = 188mm

The ratio primary winding (ten strands or 200 turns) requires: $200\times(188+92)\approx 56m$

Resistance R_1 is approximately: $R_1\approx 56m\times 88m\Omega m^{-1}\approx 5\Omega$

Permittance of ratio core: $A_L=328\mu H/turn^2$ (SM200 grade)

Inductance of ratio windings: $L_1=N^2A_L=400\times 400\times 328\times 10^{-6}\approx 52H$

I shall assume the following values: -

a). Transformer characteristics: $R_1\approx 5\Omega \quad R_2\approx 0.3\Omega \quad R_3\approx 0.18\Omega$

b). The compensator resistance is set at the maximum expected source resistance: $R\approx R_S\approx 100\Omega >> R_3, R_1$

c). The two energising windings have the same inductance: $L_2\approx L_3\approx 5.4H$

d).Boosted inductance of the ratio winding primary: $L_1\approx 52H$

e). The target natural frequency is approximately: $\omega_N\approx 1$rads^{-1}.

f). The capacitor has (very approximately) the same impedance as *R* at the natural frequency.

The compensator capacitor is, therefore: $X_C\approx\dfrac{1}{\omega_N C}\approx R \quad\Rightarrow\quad C\approx\dfrac{1}{\omega_N R}\approx 10mF$

These calculations are “ball park” (the inductances, in particular, are highly unpredictable). One can simplify the manual calculation with approximations: -

$$L_1\approx 10L \text{ with } L\approx 5.4H \text{ and } R_S+R_1\approx R_S \text{ and } R+R_3\approx R$$

$$\omega_N=\left(\frac{(R_S+R_1)R_2(R+R_3)}{L_1L_2L_3RC}\right)^{\frac{1}{4}} \quad\Rightarrow\quad \omega_N\approx\left(\frac{R_SR_2}{10L^3C}\right)^{\frac{1}{4}}\approx 1.17 \text{ rads}^{-1}$$

The natural frequency is near enough

$$\delta=\frac{RR_3C}{R+R_3}\omega_N \quad\Rightarrow\quad \delta\approx R_3C\omega_N\approx 2.1\times 10^{-3} \text{ and small enough.}$$

$$a=\frac{[(L_1+L_2+L_3)R_2+R_1(L_2+L_3)](R+R_3)+L_3R_1R_2}{(R_S+R_1)R_2(R+R_3)}\omega_N \quad\Rightarrow\quad a\approx\left(\frac{12L}{R_S}+\frac{2LR_1}{R_SR_2}+\frac{LR_1}{R_SR}\right)\omega_N\approx 3$$

The full expression for *b* is: -

$$b=\frac{(L_2+L_3)(R+R_3)L_1+(L_1+L_2+L_3)R_2RR_3C+R_1RR_3C(L_2+L_3)+(L_2R_1+L_1R_2+L_2R_2)L_3+(L_3R_1R_2RC)}{(R_S+R_1)R_2(R+R_3)}\omega_N^2$$

Only the first term is significant so that: $b\approx\left(\dfrac{20L^2}{R_SR_2}\right)\omega_N^2\approx(19)(1.17)^2\approx 26$

The full expression for *c* is: $c=\dfrac{(L_2+L_3)RR_3CL_1+L_1L_2L_3+(L_2R_1+L_1R_2+L_2R_2)L_3RC}{(R_S+R_1)R_2(R+R_3)}\omega_N^3$

$$\Rightarrow\qquad c\approx\left(\frac{20L^2R_3C}{R_SR_2}+\frac{10L^3}{R_SR_2R}+\frac{L^2(R_1+11R_2)C}{R_SR_2}\right)\omega_N^3\approx(0.035+0.52+0.08)(1.17)^3\approx 1$$

The result is a peak in the frequency response but well below the operating frequency. There is further scope for increasing the compensator capacitor though practical considerations come into play. See the monograph "A simulated large capacitor circuit" by the same author [1].

The precise values from a spreadsheet calculator are: -

$$L_1 = 52H \quad L_2 = L_3 = 5.4H \quad R = R_S = 100\Omega \quad R_1 = 5\Omega \quad R_2 = 0.3\Omega \quad R_3 = 0.18\Omega$$

$$\omega_N \approx 1.2 \ \left(f_N \approx 0.19Hz\right) \quad a = 2.78 \quad b = 25.8 \quad c = 1.02 \quad \text{and} \quad \delta = 2.16\times10^{-3}$$

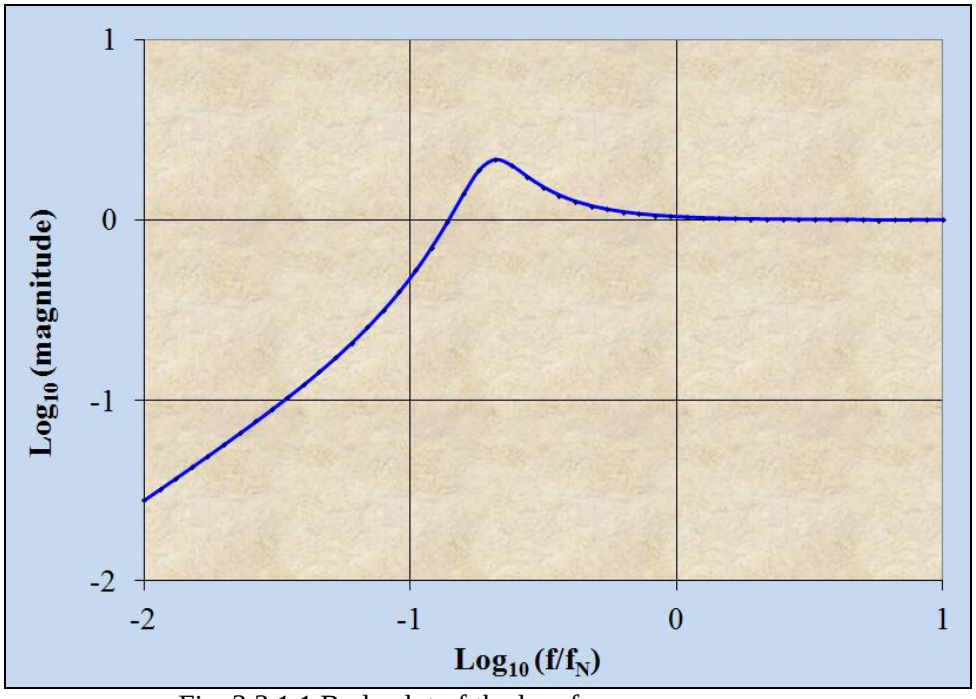


Fig. 3.3.1.1 Bode plot of the low frequency response

The error analysis, from the previous section, is:-

$$f >> f_N \ \Rightarrow \ T(f) \approx 1 - j\left\{\delta\left(\frac{f_N}{f}\right)^3 + c\left(\frac{f_N}{f}\right)^5\right\} - \left(\frac{f_N}{f}\right)^4 \quad \text{with} \quad f_N = \frac{\omega_N}{2\pi} \approx 0.19Hz$$

$$\delta \approx R_3 C\omega_N \approx 2.16\times10^{-3} \quad \text{and, at 25Hz} \quad \frac{f_N}{f} \approx 7.6\times10^{-3} \ \Rightarrow \ T(f) \approx 1 - j0.8\times10^{-9} - 2.7\times10^{-9}$$

Both quadrature and in-phase errors are very small. The contribution from the 5th order term is negligible.

1. Part 6, monograph 2: "A simulated large capacitor circuit".

4. Energising from the secondary

Whereas most applications are best served by driving both energising windings from the primary side a potentially useful variant employs a single low noise high accuracy voltage follower (HAVF) on the secondary side to provide the energising current. The primary winding input is fully differential and the input impedance remains boosted but the output is relative to local 0V. Loading on the secondary winding is negligible and the very low output resistance of the HAVF could be useful.

The main advantage is the saving of a voltage follower and its floating power supply (lower cost and size). Despite this there are no obvious (high accuracy) applications – the output relative to 0V is a problem.

One might expect that a possible application is for noise matching. This is not so – a two-stage approach provides a sufficient boost to the input impedance. The main problem is the phase error and the lower magnitude error of a three-stage transformer is of no benefit. For more detail see the monograph "Noise matching transformers" by the same author [1].

High accuracy with a high source resistance could be one possibility. Previous analysis has shown that the best way to ensure stability with moderately high source resistance (up to 100Ω) is to boost the top core inductance and add a compensator in series with at least one of the first energising windings. With higher source resistance (> 100Ω), however, it would make sense to add compensators to both energising windings. For example: -

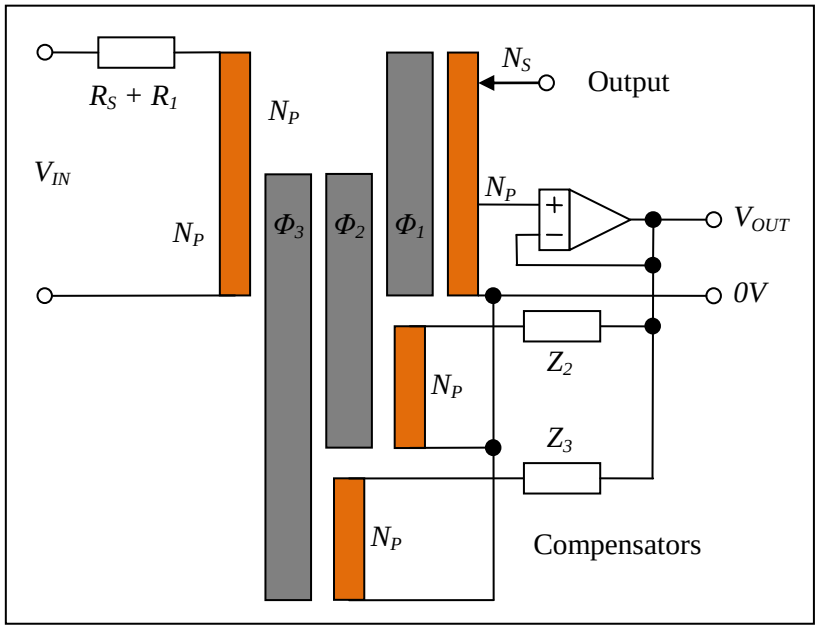


Fig. 4.1 Outline schematic of an impedance matching transformer

The follower noise injected via the energising windings is negligible (see section 5) but the noise current is reflected back to the primary and flows through the source resistance. Best performance is achieved, therefore, when the noise resistance of the follower is significantly greater than the source resistance. With compensators in series with both energising windings the DC performance of the follower becomes less important and a JFET input stage becomes an attractive option, especially for the higher range of source resistance.

To the best of my knowledge there is nothing in the literature, based on this idea, and no commercial kit is currently available. If you can think of an application, dear reader, please contact the author.

1. Part 3, monograph 5: "Noise matching transformers".

Noise matching transformers

1. Introduction

Best noise performance is achieved when the noise resistance of the pre-amplifier matches the signal source resistance [1]. A reasonably good match is within a factor of three. A larger mismatch can be corrected with a transformer. Optimal noise matching is when the transformer ratio is the square root of the resistance ratio: -

For optimal noise matching: $$\frac{N_S}{N_P} = \sqrt{\frac{R_N}{R_S}}$$

Where: -

N_P = Number of primary turns

N_S = Number of secondary turns

R_N = Noise resistance of pre-amp (voltage noise/current noise)

R_S = Source resistance

Excellent matching from about 1kΩ to over 100kΩ is easily achieved by direct connection to a pre-amp with a bipolar junction transistor (BJT) front end [2]. The noise resistance of BJTs can be controlled over a range 1kΩ to over 100kΩ by selecting the required collector current. Lower source resistance can be accommodated by connecting a number of BJTs in parallel, though the law of diminishing returns sets in rather quickly: the reduction is proportional to the square root of the number of pairs [1]. Reducing the noise resistance by a factor of ten, for example, would require 100 matched pairs at considerable expense. At the upper end of the range it may be possible to connect directly to a large area (low noise resistance) type junction field effect transistor (JFET) pair even if matching is not ideal – noise power spectral density (and noise temperature equivalent) is at least two orders of magnitude lower for the best JFETs (typically 10^{-21}W/Hz or 20K for a BJT versus 4×10^{-24}W/Hz or 0.1K for a JFET).

JFETs with noise resistance from 100kΩ to over 10MΩ are readily available. Matching can be adjusted, but only to a very limited extent, by selecting the required drain-source current. JFETs also work well at very low temperature though that would be expensive and beyond the scope of these monographs.

One application for a noise matching transformer is from low source resistance (e.g. platinum resistance thermometers (typically 0.3 - 400Ω)) to a BJT pre-amp (noise resistance typically ≈ 1kΩ). This is readily achieved with a basic (single-stage) transformer. There is no point in improving on this (e.g. matching to a JFET) as the main source of noise is Johnson noise from the source (T > 20K).

The main problem with single-stage transformers is the phase shift due to source resistance and primary inductance. The result is a design with a large, high permeability core with a large number of turns for the primary (hundreds) requiring a very large number (thousands) of turns for the secondary. Whereas this is not particularly difficult (technically) it is laborious and time consuming.

There may be other applications, however, where the source is very cold (< 20K) but the resistance is too low for a direct match to JFETs. This raises the possibility of a two-stage approach [3]. To the best of my knowledge no commercial kit is available and a comprehensive search of the literature has drawn a blank. Section 3 remains in the realm of theory but if you are interested in making and testing a prototype please contact the author.

1. Part 5, monograph 1: “Null detectors – the basics”. See section 2.
2. Part 5, monograph 2: “Low noise BJT pre-amps”
3. Part 3, monograph 3: “Two-stage IVDs and RTs”.

2. A single stage transformer

Near ideal impedance matching (from 0.3Ω – 300Ω to a BJT matched pair) can be achieved with a single-stage transformer. The main problem is the requirement for a large (typically >10c) high permeability (SM100) core and the very large number of turns required for the secondary in order to achieve the necessary input impedance. The following has three settings (nominally 1, 10 and 100Ω source resistance). The BJT amplifier has collector currents set for a noise resistance of 1kΩ (typically 0.1mA) [1].

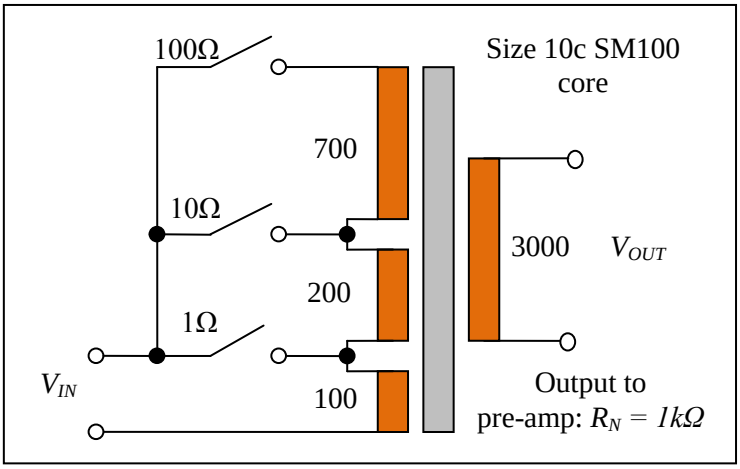


Fig. 2.1 A noise matching transformer for low source resistance

The input impedance at the lowest operating frequency (typically 25Hz) is: -

$$Z = 2\pi fL \quad \text{with} \quad L = N_P^2 A_L \quad \text{and} \qquad A_L = 153\,\mu H/turn^2$$

The resulting input impedances are: for the 1Ω winding (100 turns): $L = 1.53H, |Z| \approx 240\Omega$; 10Ω (300 turns): $L = 13.8H, |Z| \approx 2k\Omega$ and for the 100Ω winding (1000 turns): $L = 153H, |Z| \approx 24k\Omega$. The winding resistances are negligible compared to the inductive impedances. According to basic transformer theory the transfer function has a first order high pass characteristic [2]: -

$$T(s) = \frac{\tau s}{1+\tau s} \quad \text{with} \quad \tau = \frac{L}{R_S}$$

At the lowest operating frequency the effect on magnitude is small and not particularly important. The phase shift, however, could affect the null detector operation. In polar form: -

$$T(s) = \cos(\theta)\exp(j\theta) \quad \text{with} \quad \theta = \arctan\left(\frac{1}{\tau\omega}\right)$$

$$\tau\omega >> 1 \;\Rightarrow\; \cos(\theta) \approx 1 - \frac{\theta^2}{2} \quad \text{and} \quad \theta \approx \frac{1}{\tau\omega} = \frac{R_S}{\omega L}$$

The result is less than one degree phase shift for source resistance up to 300Ω. See fig. 2.4

1. Part 5, monograph 2: “Low noise BJT pre-amps”. See section 2.4.
2. Part 3, monograph 2: “Single stage inductors and transformers”. See section 4.

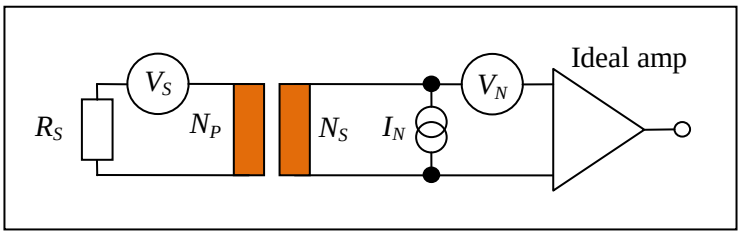


Fig. 2.2 Noise model with matching transformer

Define the transformer ratio and noise resistance: $\alpha = \frac{N_S}{N_P}$ and $R_N = \frac{V_N}{I_N}$

According to basic (single-stage) transformer theory the current and voltage reflected to the primary side are [1]: -

$$I_P = \alpha I_N \quad \text{and} \quad V_P = \frac{V_N}{\alpha}$$

The source noise, V_S, noise current and pre-amp noise voltage are statistically independent and the total (RMS) noise voltage, referred to the primary side is, therefore [2]: -

$$V_T^2 = V_S^2 + \left(\alpha I_N R_S\right)^2 + \left(\frac{V_N}{\alpha}\right)^2$$

This is a minimum when the gradient is zero with respect to the transformer ratio: -

$$\frac{\partial\left(V_T^2\right)}{\partial\alpha} = 2\alpha I_N^2 R_S^2 - \frac{2V_N^2}{\alpha^3} = 0$$

$$\Rightarrow \quad \alpha^4 = \frac{V_N^2}{I_N^2 R_S^2} \quad \Rightarrow \quad \alpha = \sqrt{\frac{R_N}{R_S}} \quad \text{QED.}$$

If one assumes a noise-free source $\left(V_S = 0\right)$ the noise power referred to the primary side is: -

$$P_S(\alpha) = \frac{V_T^2}{R_N} = \left(\alpha I_N R_S\right)^2 + \left(\frac{V_N}{\alpha}\right)^2 = \frac{I_N^2}{R_N}\left(\left(\alpha R_S\right)^2 + \left(\frac{R_N}{\alpha}\right)^2\right)$$

The minimum power is, therefore: $\alpha = \sqrt{\frac{R_N}{R_S}} \quad \Rightarrow \quad P_{MIN} = \frac{I_N^2}{R_N} \times 2R_N R_S$

Expressed as a ratio the power match efficiency is, therefore: -

$$\frac{P_{MIN}}{P_S(\alpha)} = \frac{2R_N R_S}{\left(\alpha R_S\right)^2 + \left(R_N/\alpha\right)^2}$$

1. Part 3, monograph 2: “Single stage inductors and transformers”. See section 4.
2. Part 5, monograph 1: “Null detectors – the basics”. See appendix 1.
3. Spreadsheet: “Noise matching transformer”.

For the ideal transformer (see fig. 2) with three settings 1Ω, 10Ω and 100Ω the degree of matching is satisfactory from 0.3Ω to about 300Ω [1]: -

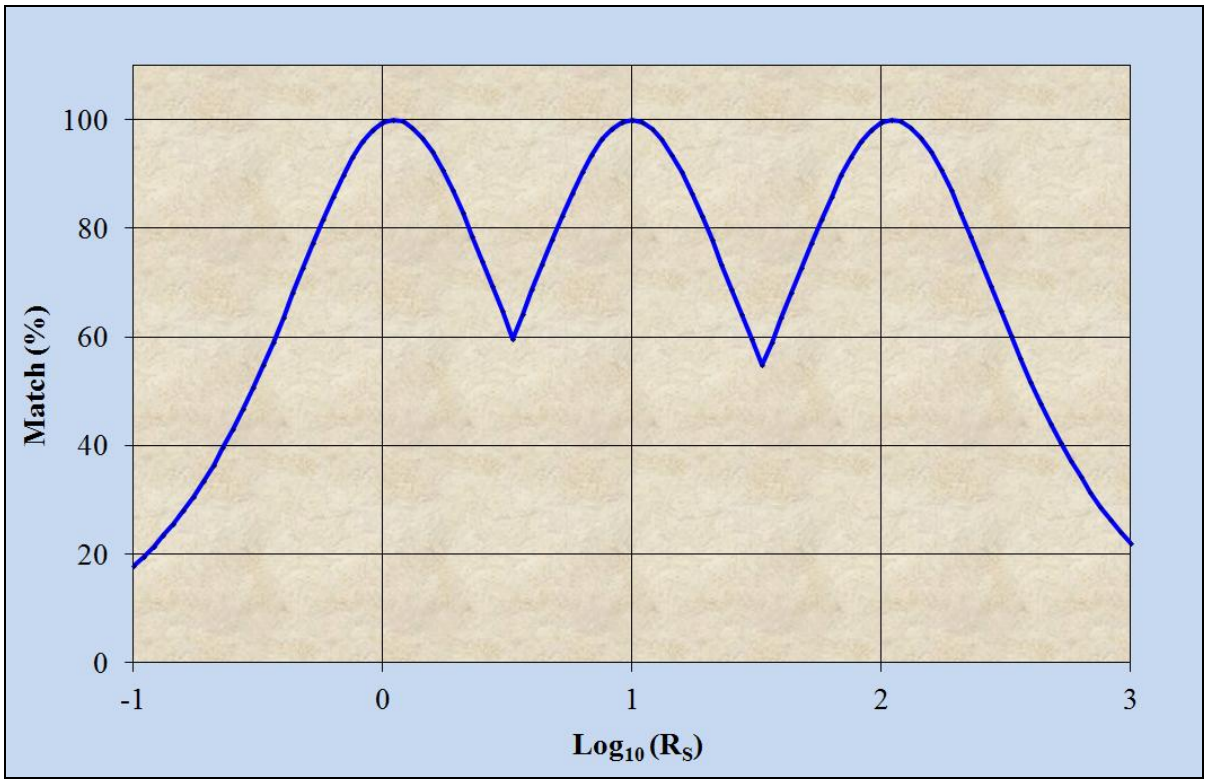


Fig. 2.3 Power matching efficiency for a single-stage transformer

The phase shift is sufficiently low for a null detector [1]: -

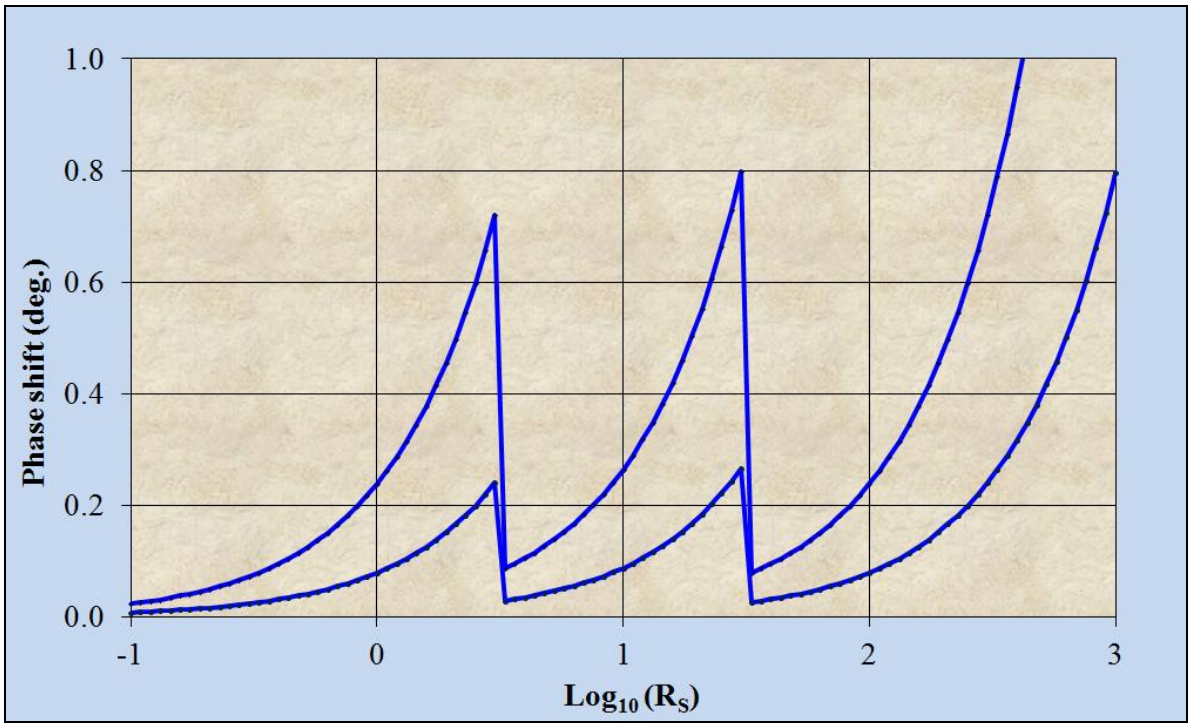


Fig. 2.4 Phase shift (25Hz and 75Hz)

(The lower frequency results in greater phase shift.)

1. Spreadsheet: “Noise matching transformer 1”.

3. A possible two-stage transformer

3.1 Outline design

A two-stage transformer could offer advantages compared to the single-stage approach: fewer turns and smaller cores. The actively driven energising winding provides an enormous boost to the input impedance which, at the operating frequency, becomes real (resistive). The result is a significant reduction in phase shift. It is then possible to reduce the number of turns on the primary and secondary for a given source resistance.

A higher primary to secondary ratio is also practicable (e.g. to match a low to medium source resistance to the much higher noise resistance of a JFET pre-amplifier). The proposed design, however, is aimed at matching the output of a single-stage transformer (1kΩ) to the lowest noise JFET pre-amplifier: a step-up ratio of about 1:30.

Previous analysis [1] has shown that the best way to ensure stability with moderately high source resistance (up to 1kΩ) is to boost the top core inductance with extra windings and a higher permeability core plus a compensator (a large capacitor and parallel resistor) in series with the energising winding. The approach chosen is active drive, from the secondary side [1], with a (JFET input) compensator [2].

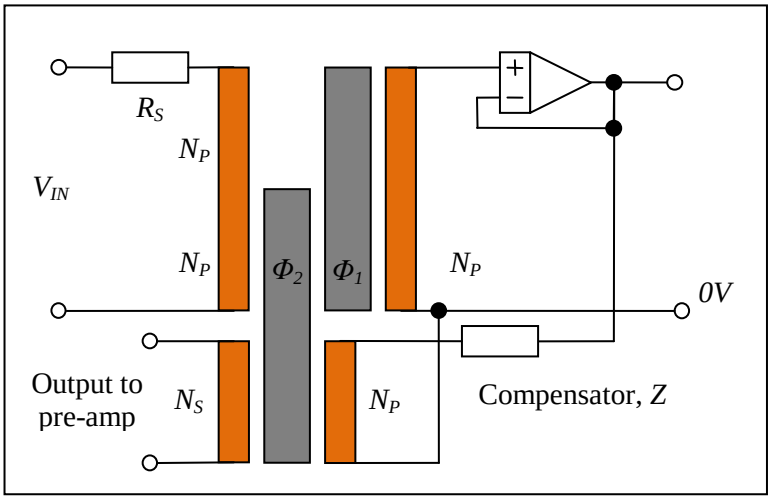


Fig. 3.1.1 Outline schematic of an impedance matching transformer

The lower accuracy requirement also means that the primary and secondary windings do not need to be on the same rope which can be an advantage (low capacitive coupling between primary and secondary). Also, most of the flux, corresponding to the input voltage, is in the energising core. The main secondary (output to the pre-amp) could be wound around that core only, rather than both. This would save a great deal on wire though it may not be the most practicable approach. It should be possible to employ a relatively simple compensator (no need for a high accuracy voltage follower), albeit of composite type with a low noise JFET front end [3].

1. Part 3, monograph 3: “Two-stage IVDs and RTs”. See section 7 (active drive from the secondary).
2. Part 6, monograph 2: “A simulated large capacitor circuit”.
3. Part 5, monograph 3: “Low noise JFET pre-amps”

3.2 Circuit analysis

It is shown elsewhere [1] that the equivalent circuit (assuming a ratio of 1:1) is a high pass filter with active drive from the output. The follower and series impedance are hypothetical – the actual circuit is a simulated large capacitor which performs the same function [2]: -

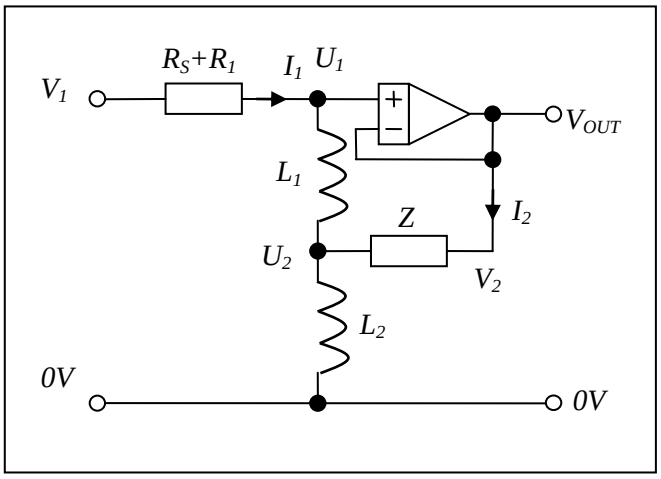


Fig. 3.2.1 The equivalent circuit

With no compensator the impedance Z is simply the energising winding resistance R_2. With a compensator make the substitution $R_2 \rightarrow Z$. This is the compensator in series with the winding resistance: -

$$Z = R_2 + \frac{R}{1+sRC}$$

Also, the resistance R_1 consists of both source and winding resistance. The transfer function (from [1]) becomes: -

$$T(s) = \frac{V_{OUT}}{V_1} = \frac{s(L_1+L_2)Z + s^2 L_1 L_2}{(R_S+R_1)Z + s(L_1+L_2)Z + s^2 L_1 L_2}$$

A useful parameter is the determinant: $|\mathbf{Z}| = (R_S+R_1)Z + s\{L_2(R_S+R_1) + (L_1+L_2)Z\} + s^2 L_1 L_2$

One can check this with the matrix method [3]. The output impedance matrix is: -

$$\mathbf{U} = \mathbf{XI} \quad \text{or} \quad \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} sL_1 + sL_2 & sL_2 \\ sL_2 & sL_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$$

The input impedance matrix is: -

$$\mathbf{V} = \mathbf{ZI} \quad \text{or} \quad \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} R_S + R_1 + sL_1 + sL_2 & sL_2 \\ sL_2 & Z + sL_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$$

The inverse is: $\mathbf{I} = \mathbf{Z}^{-1}\mathbf{V}$ or $\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \frac{1}{|\mathbf{Z}|} \begin{pmatrix} Z + sL_2 & -sL_2 \\ -sL_2 & R_S + R_1 + sL_1 + sL_2 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$

1. Part 3, monograph 3: “Two-stage IVDs and RTs”. See section 3.
2. Part 6, monograph 1: “A simulated large capacitor circuit”.
3. Part 2, monograph 1: “Two-stage filters”. For an introduction to the matrix method see section 2.

From which the transfer function matrix equation is: -

$$\mathbf{U} = \frac{\boldsymbol{\Omega}}{|\mathbf{Z}|}\mathbf{V} \quad \text{or} \quad \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \frac{1}{|\mathbf{Z}|}\begin{pmatrix} s(L_1+L_2)Z + s^2 L_1 L_2 & sL_2(R_S+R_1) \\ sL_2 Z & sL_2(R_S+R_1)+s^2 L_1 L_2 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

In this case the energising drive is from the output side. If one assumes an ideal voltage follower/compensator: -

$$V_1 = V_{IN} \quad \text{and} \quad V_2 = U_1 = V_{OUT}$$

$$V_{OUT} = \frac{\Omega_{11}V_1 + \Omega_{12}V_2}{|\mathbf{Z}|} \quad \Rightarrow \quad T_1(s) = \frac{V_{OUT}}{V_{IN}} = \frac{\Omega_{11}}{|\mathbf{Z}| - \Omega_{12}}$$

A little algebra produces the same transfer function as above and confirms the equivalent circuit method.

Ratio accuracy is not particularly important and it is worth checking if the voltage developed across the energising core is sufficiently accurate.

$$U_2 = \frac{\Omega_{21}V_1 + \Omega_{22}V_2}{|\mathbf{Z}|} \quad \text{and} \quad V_2 = V_{OUT} = T_1(s)V_{IN} \quad \Rightarrow \quad T_2(s) = \frac{U_2}{V_{IN}} = \frac{\Omega_{21} + \Omega_{22}T_1(s)}{|\mathbf{Z}|}$$

One could expand the transfer function to the usual polynomial form but that method reveals little. It is much simpler to simulate the equations with a spreadsheet and bite-size calculations (first calculate the matrix elements, determinant and transfer function $T_1(s)$). The results confirm that the phase error is sufficiently small.

3.3 Example calculation

In this case the main issue is stability – the inductance of the top core needs to be substantially greater than the energising core. This can be achieved with extra turns around the top core only and the highest possible permeability. Effective bootstrapping requires the energising winding to be uniformly distributed around the core to minimise flux leakage. A number of windings can be connected in parallel and this also ensures low resistance. The reduced number of turns for the primary and energising windings also allows the use of thicker wire and medium sized cores. A useful starting point is the example based on 5c cores (mumetal and super mumetal: $\mu_R \approx 50{,}000$ and 200,000 respectively) [1]. The primary windings are 40 turns BNNL of 0.457mm enamelled copper wire (26SWG).

Both cores type 5c: -
OD = *41mm* ID = *23mm* axial = *15mm*.
The length of wire per turn on the energising winding is approximately: *15 × 2 + 41 – 23 = 48mm*

Inner circumference $= 23 \times \pi = 72mm$ mm can accommodate 160 turns (four windings of 40 turns each) in a single layer of 0.45mm OD wire. It may be a bit of a squeeze but worth a try with 26SWG (enamelled) at 0.457mm diameter as one is likely to have some in stock. A couple of extra turns of double layer at the centre of each BNNL as the winding is finished off should leave plenty of room for the secondary winding. Otherwise 0.4mm diameter wire would be a comfortable fit. The outer circumference is not critical.

Each energising winding requires: $40 \times 48mm + 200mm \approx 2.1m$ of wire, including 100mm tails.
Resistance of each energising winding would be approx: $R \approx 2.1m \times 105m\Omega m^{-1} \approx 0.22\Omega$
Four windings connected in parallel produces a net resistance: $R_2 \approx 0.22/4 \approx 0.056\Omega$
Permittance of energising core (mumetal): $A_L = 51\mu H/turn^2$
Inductance of energising windings: $L_2 = N^2 A_L = 40 \times 40 \times 51 \times 10^{-6} \approx 0.082H$

1. Part 3, monograph 3: “Two-stage IVDs and RTs”. See section 7.1.

The secondary (output to the pre-amp) consists of 126 turns (BNNL) of a ten strand rope of fine wire. The winding resistance is not important but the wire should be strong enough to be manageable (without breaking). Enamelled copper of 0.2mm diameter should be satisfactory. The primary to secondary ratio is then up to 40:1260 (1:31.5 and a maximum noise resistance match of approximately 1:1000).

The inner diameter is reduced by 1mm so that the inner circumference is reduced to $22 \times \pi = 69mm$

The rope is approximately 1mm in diameter requiring two layers on the inner circumference: -
First layer: possibly $22mm \times \pi \div 1mm \approx 69$ (35 per side looks good). The internal diameter reduces to 20mm.
Second layer: possibly $20mm \times \pi \div 1mm \approx 63$ turns (28 per side to make up the total required). The internal diameter reduces to 18mm and more than sufficient for the ratio windings.

The primary winding consists of 40 turns around the top core only (BNNL) plus another 40 turns around both cores (BNNL) occupying half the core. With twice as many turns and four times the permeability the inductance is increased by a factor of 16: -

Permittance of top core (SM200): $A_L = 204\mu H / turn^2$
Inductance of energising windings: $L_2 = N^2 A_L = 80 \times 80 \times 204 \times 10^{-6} \approx 1.3H$
The primary winding resistance is negligible compared to the source resistance (up to 1kΩ).

The ratio secondary also consists of 40 turns around the top core only plus 40 turns around both cores (BNNL) on the other half of the core. This should keep interwinding capacitance sufficiently small without the need for screening. Winding resistance is again negligible as the compensator has very high input impedance.

Simulation with a spreadsheet indicates that a suitable compensator is 10Ω in parallel with 0.1F. The resulting phase shift is: -

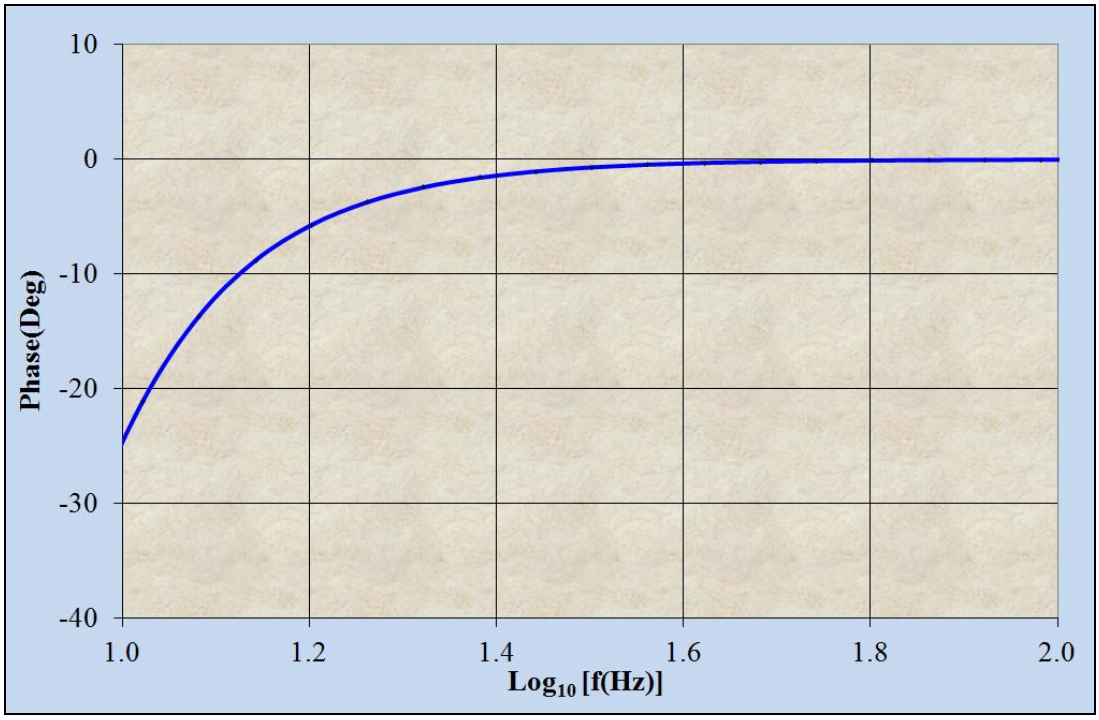


Fig. 3.3.1 Phase shift for 1kΩ source resistance

There is also a slight peak in the magnitude response but too small to be of concern.

1. Spreadsheet "A two-stage noise matching transformer"

An F17 type ratio transformer bridge

1. Introduction

In the summer of 1978 I graduated and joined ASL as a Development Engineer, thanks to Peter Wolfendale, based on my knowledge and experience of microprocessors. My first job was to design the printed circuit board for the first of what became the "F series" of resistance thermometry bridges. JDY had come up with some new ideas which, with the introduction of a microcontroller, made it possible to significantly reduce size and cost, compared to the market leader – the ASL A7 double Kelvin automatic bridge. When I asked John "what are we going to call it?" he replied "Fred", short for "fairly random electronic design". The prefix "F" survived for quite a few years in the form of models F16, F17, F18, F25 and F26.

Fig. 1.1 The F17 (Picture courtesy ASL Ltd)

One of JDY's main innovations was the high accuracy "inside-out" voltage follower (HAVF) [1]. With very high input impedance and very low output impedance it resulted in a major simplification in the required characteristics of the ratio transformer.

The resulting design was very simple to construct: a three stage transformer with mid–range sized toroids, only 100 turns for the energising stages and sufficient room for an extra winding which allowed a range of: -

0.000000 to 3.999999

This range is well suited to resistance thermometry as the maximum value of resistance of a typical commercially available SPRT [2] is just less than four times its resistance at the triple point of water.

The F17 operates at 1.5 × supply frequency (75 or 90Hz) for maximum rejection of supply harmonic interference. This is low enough for most thermometers and standard resistors so that AC/DC differences are negligible.

The only disadvantage is the noise generated by the followers, reducing resolution for a given bandwidth, compared to what is possible theoretically. For many industrial customers and academics, however, this was not a problem and the F17 sold like hot cakes.

1. Part 4, monographs 1 and 2: "High gain blocks" and "High accuracy voltage followers"
2. For example the Tinsley model 5187SA Standard Platinum Resistance Thermometer.

1.1 The basic principle of operation

The followers drive both energising windings and ratio windings in parallel. Note the order of connections – the current flowing through the first energising winding does not flow in any part of the connection to the second energising winding or the ratio winding. Similarly, the (smaller) current flowing through the second energising winding does not flow in any part of the connection to the ratio winding. The connections to the ratio winding are thus closest to the inverting input which, by the action of negative feedback, is at the same voltage as the non-inverting input to a very high degree of accuracy. This order of connections is essential for accurate operation – and not a little disconcerting for the production department.

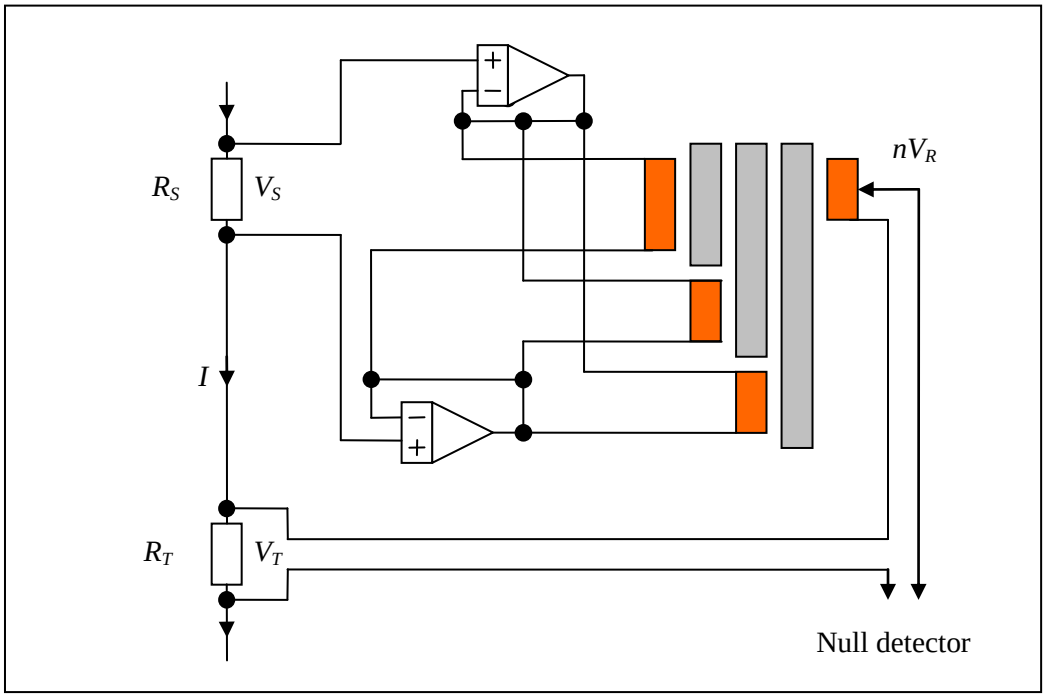


Fig. 1.1.1 The F17 basic bridge configuration (simplified)

The ratio transformer is very easy to construct. The followers had sufficient accuracy (sub-ppm) and low noise to achieve the target specification – accuracy to better than 1mK with a 25Ω SPRT.

The internal reference resistor is contained in a very simple temperature controlled oven, which most users found perfectly adequately stable over time.

N.B. The alternative to high accuracy followers was to use two extra reference resistors, in series with R_S, to provide the voltages for the two energising windings. The main disadvantages of this approach are twofold: -
1. The reference resistors are fixed internal devices, requiring a temperature controlled environment. It is preferable to have a single external reference resistor e.g. a Wilkins type.
2. The high source resistance requires the toroidal cores to be larger and/or of higher grade (higher permeability) or with more turns.

Historical note: This approach was taken by ASL's main competitor, Tinsley Instruments, resulting in the far inferior model 5840 originally developed by R. B. D. Knight of NPL [1].

1. Knight R. B. D., "A precision bridge for resistance thermometry using a single inductive current divider." Euromeas. 77, IEE conference. Pub. 152.

1.2 Active guard circuit

Another innovative design feature of the F17 is the use of a high gain block (HGB, two-stage, type 1 [1]) to create a "virtual earth" point for the bridge. The HGB, standard resistor and thermometer are configured as an inverting amplifier. The action of feedback ensures that the inverting input of the HGB is maintained at 0V very accurately, setting the bridge potentials relative to local earth. This is much more practicable than a "Wagner balance" [2].

The F17 employs coaxial cables for both external standard resistor and thermometer. The outer conductors (current and voltage) provide adequate screening of the more sensitive (to interference) inner conductors. Any interference coupled to the (high impedance) current source side outers flows through both resistors and has no affect on the balance. The HGB side has very low impedance and any interference is absorbed by the HGB output.

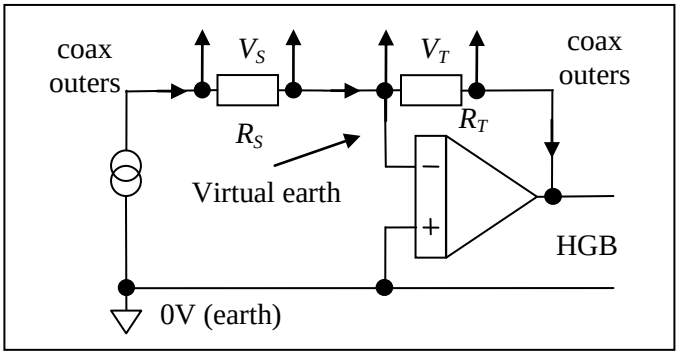


Fig. 1.2.1 The F series active guard

2. The main ratio transformer [3]

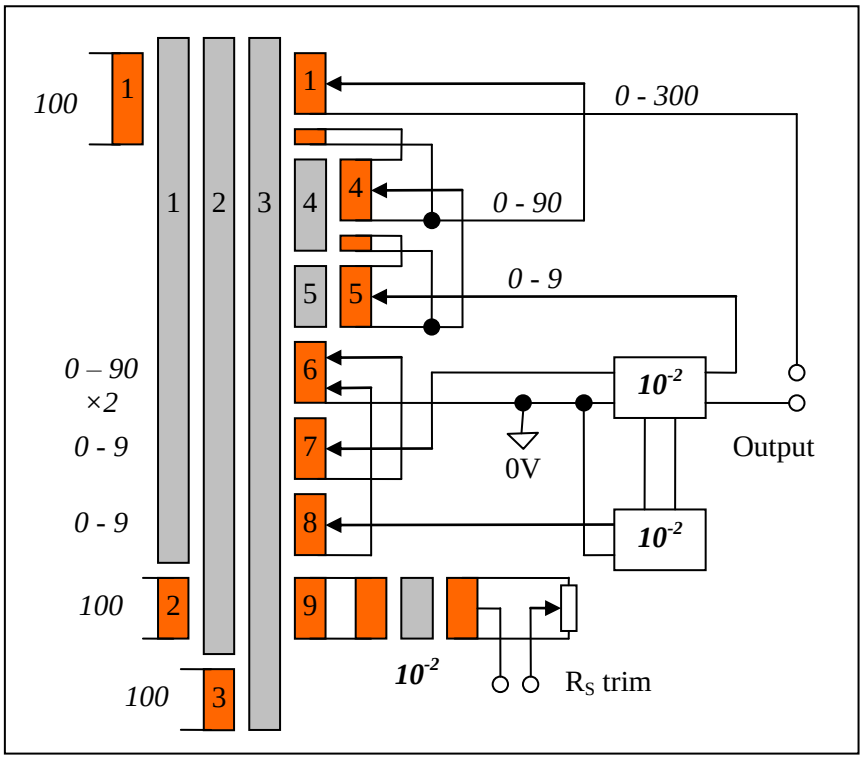


Fig. 2.1 The F17 main ratio transformer

For more detail (theory) see: -
1. Part 4, monograph 1: "High gain blocks".
2. Part 3, monograph 9: "Coaxial AC chokes"
3. Part 3, monograph 4: "Three-stage RTs".

One of the main advantages of direct drive with followers is the possibility of employing a three-stage ratio transformer with thinner wire (0.5mm) on medium sized cores (7a) and a much reduced number of turns. The result: all seven decades can be incorporated in a single transformer unit which is easy to manufacture.

Winding 1: 5 strand rope with 100 turns around all three main cores (BNNL). One of the strands is used as the ratio primary, three for a variable secondary (decade 1) and one for equalisation of the next decade. Being part of the same rope the voltage induced in each strand is the same to a high degree of accuracy.

Windings 2 & 3: Energising windings - 100 turns (BNNL) in a single layer to fill the inner circumference.

Winding 4: 11 strand rope with 10 turns (BNNL) around all three main cores plus a few turns through small auxiliary core 4. 10 of the strands are connected in series and provide a variable secondary (decade 2) which is wired in parallel with the decade 1 equalisation strand. The extra strand is used for equalisation of the next decade.

Winding 5: 10 strand rope with 1 turn around all three main cores and a few turns through small auxiliary core 5. The 10 strands are connected in series and provide a variable secondary (decade 3). It is also connected in parallel with the decade 2 equalisation strand.

N.B. Decades 1, 2 and 3 ratio settings (“taps”) are selected by low resistance glass encapsulated reed relays.

Winding 6: 10 strand rope with10 turns around all three main cores. Two sets of 10 analogue switches are used to select the required taps which are then connected to the bottom of windings 7 and 8.

Winding 7: 10 strand rope with 1 turn. The required output is again selected by analogue switch. When added to one of the outputs selected from winding 6 the result is reduced by a factor of 100 and then connected to the bottom of winding 5. Winding 6 selection 1 and winding 7 thereby furnish decades 4 and 5 respectively.

Winding 8: Same as winding 7 but the result is reduced by a factor of 10000 before being added to the bottom of winding 5. Winding 6 selection 2 and winding 8 thereby furnish decades 6 and 7 respectively.

Windings 6, 7 and 8 voltages are combined using two small transformers, supplied by a pair of high pass filters/voltage followers (see fig. 2.3). Both followers require very low DC offset adjustments due to the low resistance of the primary windings: -

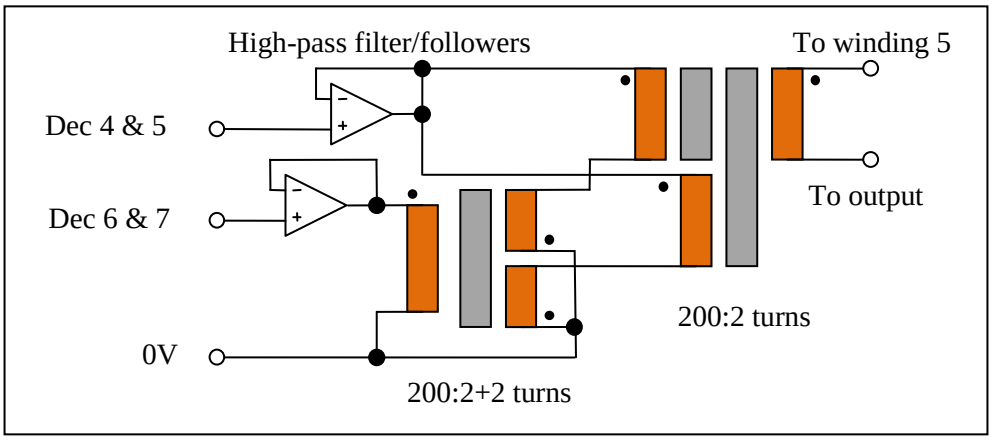


Fig. 2.2 Decades 4 to 7 are furnished by small step-down transformers

For the last two decades (6 & 7) a single-stage transformer (ratio 100:1) is sufficiently accurate. The outputs of this transformer (two windings of 2 turns each) are added to the inputs of a two-stage transformer (also 100:1). The output resistance is very low, but not negligible, consisting mainly of the resistance of the interconnecting wires. The energising current would cause a small voltage drop and so a separate output winding is used for the purpose. The second output winding is then added to the ratio winding which has much higher (bootstrapped) input impedance. The small error in the energising voltage has negligible effect on the two-stage transformer ratio accuracy. Note the relative polarity of the windings for correct adding of the voltages.

The other end of the two-stage transformer primary windings are driven by the voltage follower for decades 4 & 5. Note the order of connection to the energising winding – the energising current does not flow through any part of the connection to the ratio winding. A two stage transformer is used to reduce the phase error to a negligible level (required for the more critical decades 4 & 5). The combined output is then decades 4 to 7 isolated from 0V and correctly scaled relative to the main ratio transformer. A twisted pair is used to convey this voltage to the bottom of winding 5 where it is combined with the twisted pair output from decades 1, 2 & 3.

The reason for the two-stage high pass filters remains a mystery as there should be no DC from the analogue switches. The filter has a very high input impedance and negligible gain and phase error at 75Hz [1].

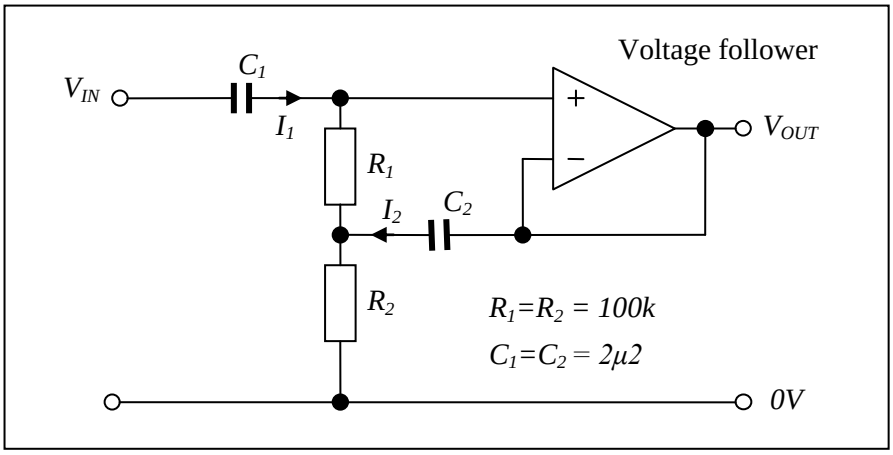


Fig. 2.3 Actively driven high pass filter (two-stage)

3. Quadrature servo and null detector

The null detector pre-amp [2] and high accuracy voltage followers [3] employ matched bipolar junction transistor pairs (LM394) at the front end with collector current set at 1mA: noise matching to approximately 350Ω. The contribution from current noise is, therefore, negligible (< 100Ω source resistance). The voltage noise contribution from each is about 1nV (RMS) in 1Hz of bandwidth and is comparable to the Johnson noise generated in a 100Ω resistor at room temperature.

After the low noise pre-amp is a band pass filter for the 75/90Hz carrier frequency and a notch filter for the first supply harmonic (50/60Hz) followed by a switchable gain voltage amplifier. Both in-phase and quadrature synchronous rectifiers employ analogue switches in a double-balanced mode [4] to produce a DC output. One of three low-pass filter bandwidths (10, 1 and 0.1Hz) can be selected for the in-phase signal.

The bridge can be balanced manually (using the front panel lever switches and meter) or automatically with a micro-controller and efficient algorithm.

The output of the quadrature synchronous rectifier is integrated and then multiplied by the in-phase reference sine wave using a precision analogue multiplier (AD533). The accurate 90 degree phase shift is implemented with a low phase error ferrite (gapped) pot core transformer [5]. The output of the pot core is added to the main bridge output for a continuous quadrature null balance.

The active guard is a high gain block (two-stage, type 1). [3] See section 1.2.

For a bit more detail on the active guard, quad servo and null detector circuits see the monograph "An F18 type ratio transformer bridge" [6].

1. Part 2, monograph 1: "Two-stage filters". See, especially, section 5.
2. Part 5, monograph 2: "Low noise BJT pre-amps".
3. Part 4, monographs 1 and 2: "High gain blocks" and "High accuracy voltage followers".
4. Part 5, monograph 1: "Null detectors – the basics".
5. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators". See section 5.
6. Part 3, monograph 7: "An F18 type ratio transformer bridge".

An F18 type ratio transformer bridge

1. Introduction

In 1983 JDY designed and wound a prototype ratio transformer that took the three stage technique to new heights of sophistication and ingenuity. The result formed the basis of the seven decade (< 0.1ppm accuracy) F18 resistance thermometry bridge which went on to become the de facto standard in all of the major temperature calibration labs around the world. It was later superceded by the WIKA model F900 with an extra decade which is now called the CTR9000 (< 20ppb accuracy) [1]. For some reason the ASL badge remains.

Fig. 1.1 The F18 (Picture courtesy WIKA/ASL Ltd)

Various standards institutions around the world were trying, at the time, to extend the range of platinum resistance thermometry to (at least) the freezing point of silver (962ºC) [2]. At such high temperatures the slight leakage of even the best insulating materials required a much lower resistance of the thermometer (2.5Ω or even 0.25Ω at 0ºC). This placed even more stringent requirements on the measurement system - in terms of noise performance and the ability to detect frequency dependent effects. A major concern was for mutual inductance between the current carrying pair and the voltage sensing pair – an effect which is more pronounced for low value resistors.

The challenge, then, was for an instrument that could be optimally matched for noise at low resistance and still maintain sub 0.1ppm accuracy with conventional 25Ω and 100Ω platinum resistance thermometers (with the relevant reference resistors), operating at the lowest possible frequency.

The resistance of a 2.5Ω SPRT increases to almost 1.3Ω at the gold point and so the range of the F18 was chosen to be: -

0.0000000 to 1.2999999

This was based on the ready availability of high grade transfer standard resistors with nominal values of 0.1Ω, 1Ω, 10Ω, 25Ω and 100Ω.

The frequency of operation was chosen to be selectable: either half or 1.5 × supply frequency (25/30Hz and 75/90Hz) so that any AC effects could be detected and, possibly, accounted for.

1. www.wika.co.uk
2. Sostmann, H. E: “Fundamentals of thermometry part III. The Standard Platinum Resistance Thermometer”. Easily found with a google search for ITS90.

2. The basic principle of operation

As with the basic three stage design [1] the two energising cores provide the great majority of the magnetic flux [2]. The current flowing in the ratio winding is substantially reduced and the accuracy is increased. The major advance of the F18 design, compared to the F17, was to eliminate the noise contribution of the followers by connecting the primary ratio winding directly across the reference resistor. This requires even higher input impedance as the reference resistance could be up to 100Ω.

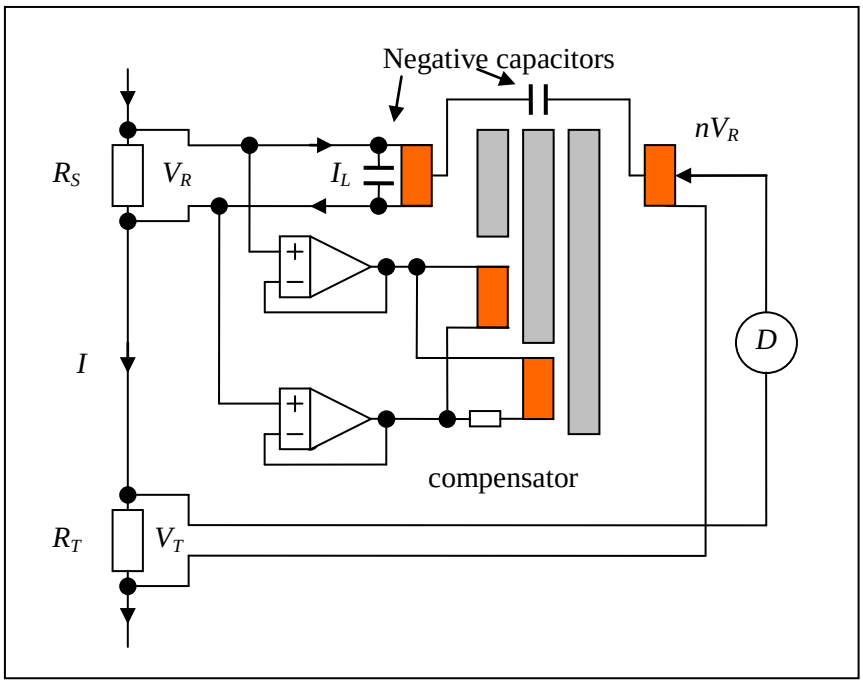


Fig. 2.1 The F18 basic bridge configuration (simplified)

This approach results in positive feedback and possible instability depending on the source resistance [2]. Stability can be restored by adding extra resistance in series with the first energising winding. Unfortunately this would reduce accuracy significantly (or increase the minimum operating frequency). JDY found an elegant and ingenious solution – a large capacitor in parallel with the extra resistance [2, 3]. At low frequency (where the instability occurs) the capacitor has little effect and the resistance dominates, maintaining stability. At the operating frequency the impedance of the capacitor is much reduced and the bootstrapping effect is largely restored – restoring the high input impedance of the primary and high accuracy.

Stability is also improved by significantly increasing the inductance of the ratio primary windings compared to the energising windings. This was achieved with another innovation – extra turns around the top core only [2].

In some applications the source resistance could be as high as 100Ω and thus the target input impedance was $>10^{10}$ Ω (the real part). The main limitation, it was found, was the inter-winding capacitances. It was found necessary to use PTFE insulated wire and a pair of negative capacitors [4] to neutralise the inter-winding capacitance plus a final manual adjustment - the "Wolfendale tweek". This was "not elegant" according to JDY but we were in a hurry to get it into production! He was also unhappy with the "ratio tweek" – a final 1:1 ratio adjustment. See section 6.

The follower current noise flows through the source resistance and the voltage generated appears across the ratio primary. Best performance is achieved, therefore, when the noise resistance of the followers is significantly greater than the source resistance. A noise resistance of 1kΩ was found to be a good compromise.

1. Part 3, monograph 6: "An F17 type ratio transformer bridge".
2. Part 3, monograph 4: "Three-stage ratio transformers".
3. Part 6, monograph 1: "A simulated large capacitor circuit".
4. Part 6, monograph 1: "A simulated large capacitor circuit".

3. Active guard circuit

The F18 uses the same active guard as the F17 - a high gain block (HGB, two-stage, type 1 [1]) to create a "virtual earth" point for the bridge. The HGB, standard resistor and thermometer are configured as an inverting amplifier. The action of feedback ensures that the inverting input of the HGB is maintained at 0V very accurately, setting the bridge potentials relative to local earth. This is an improvement over a simple direct connection to earth by making the bridge immune to "leaky" cables and connectors, which can happen in some applications. In the following diagram, for example, C_1 to C_3 represent leaky capacitances. If a direct earth connection is made to the same point then any in-phase component of the impedance of C_1 to ground would look like a large resistor in parallel with R_S, affecting the measurement. Similarly C_3 acts in parallel with R_T. No current flows in C_2 as there is no potential difference. With a virtual earth at the same point the leakage current through C_3 is provided by the HGB and C_1 simply appears in parallel with the current source – problem solved.

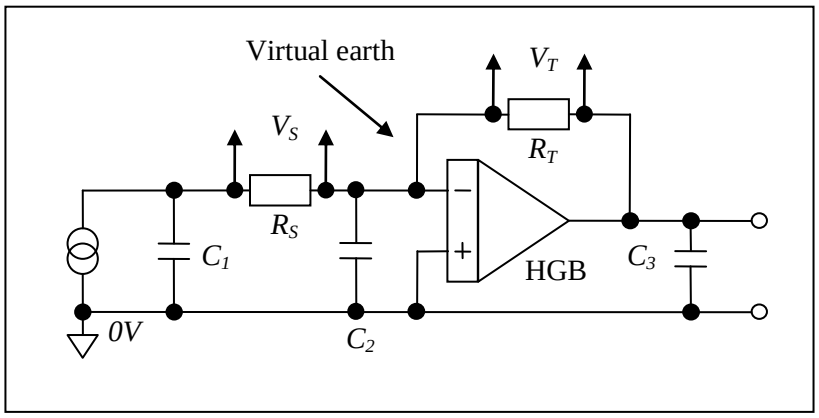


Fig. 3.1 The F series active guard

4. The divider

4.1 The ratio transformer 1 (decades 1 and 2)

Fig. 4.1.1 attempts to depict the structure of the main ratio transformer with the usual convention – A winding that is horizontally adjacent to toroidal core has windings around that core. For clarity the number of turns is specified alongside. In this case the tricky one is winding 7 – it is partly wound around the top core and then around all three.

Windings 1 and 3: Energising windings of 200 turns of 1.5mm OD wire. It is essential that these are highly uniform balanced, no-net loop (BNNL) type windings [2].

Windings 2 and 4: Energising secondary windings - for energising decades 3 and 4.
2 turns also of 1.5mm OD wire.

Winding 6: Ratio winding - 24 strands of 0.6mm OD PTFE insulated wire twisted to form a rope and wound 20 times around the top core only and then 20 times again around all three main cores (also uniform BNNL). When calculating the inductance of this winding it has, in effect, a total of 40 turns. Ten strands are connected in series for the ratio primary winding and 12 are connected in series for the variable secondary. The two remaining strands are connected in parallel to provide equalisation for decade 2. Two are used in this way to reduce the source resistance. Being part of the same rope the ratio between the primary and secondary windings is very accurate.

Winding 7: 12 strands of 0.6mm OD PTFE insulated wire wound 2 turns around TC_1 and TC_2 plus 4 turns around TC_3 plus 30 turns around auxiliary TC_4. In this way it accurately shares the same flux combination as the main ratio rope [3]. Winding 7 is tapped to act as an inductive voltage divider (IVD) for decade 2 [2].

1. Part 4, monograph 1: "High gain blocks".
2. Part 3, monograph 1: "Inductive voltage dividers and ratio transformers – the basics".
3. Part 3, monograph 2: "Single stage inductors and transformer". See section 6 on "equalisation".

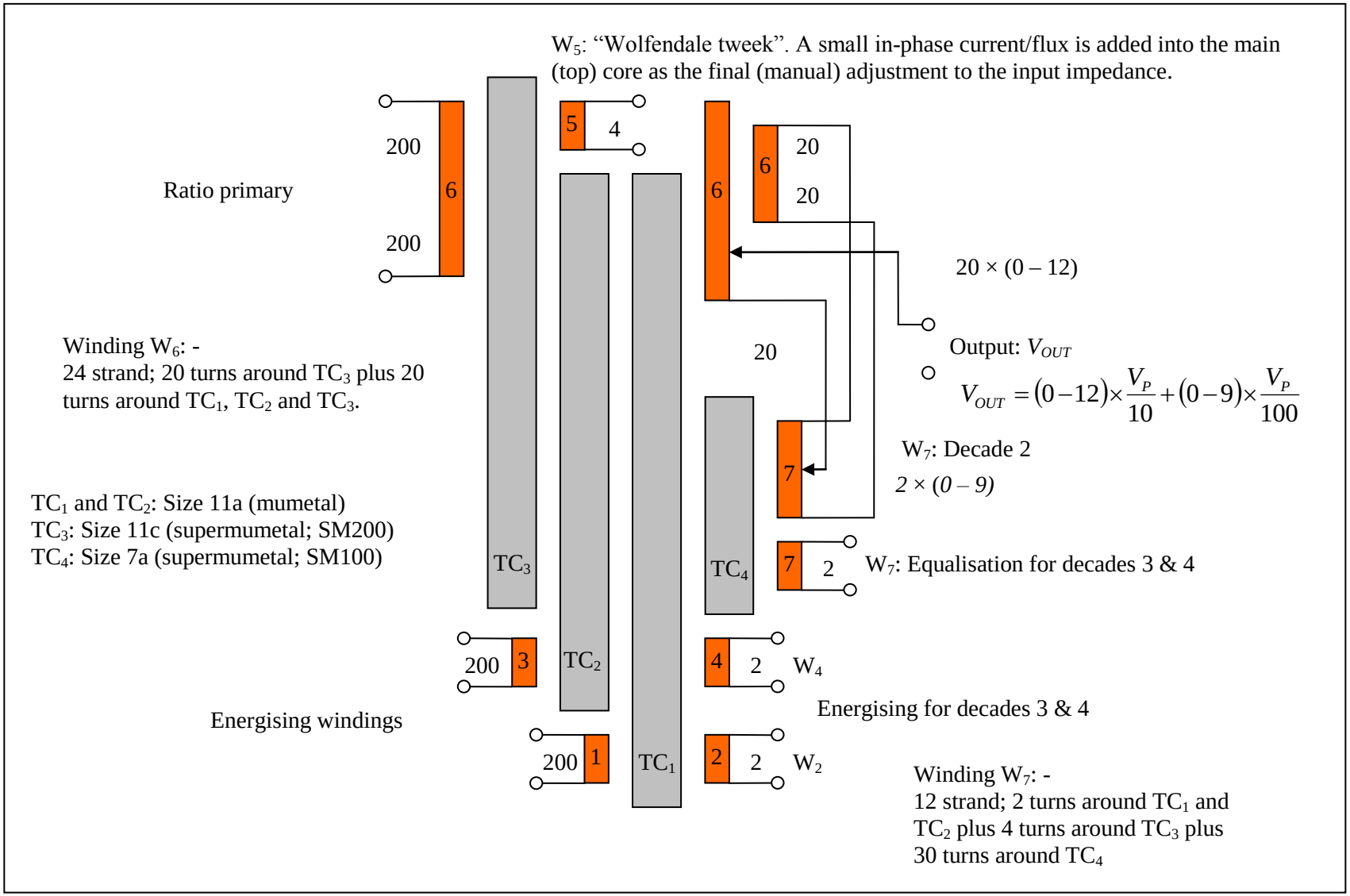


Fig. 4.1.1 F18 ratio transformer decades 1 and 2 (outline design)

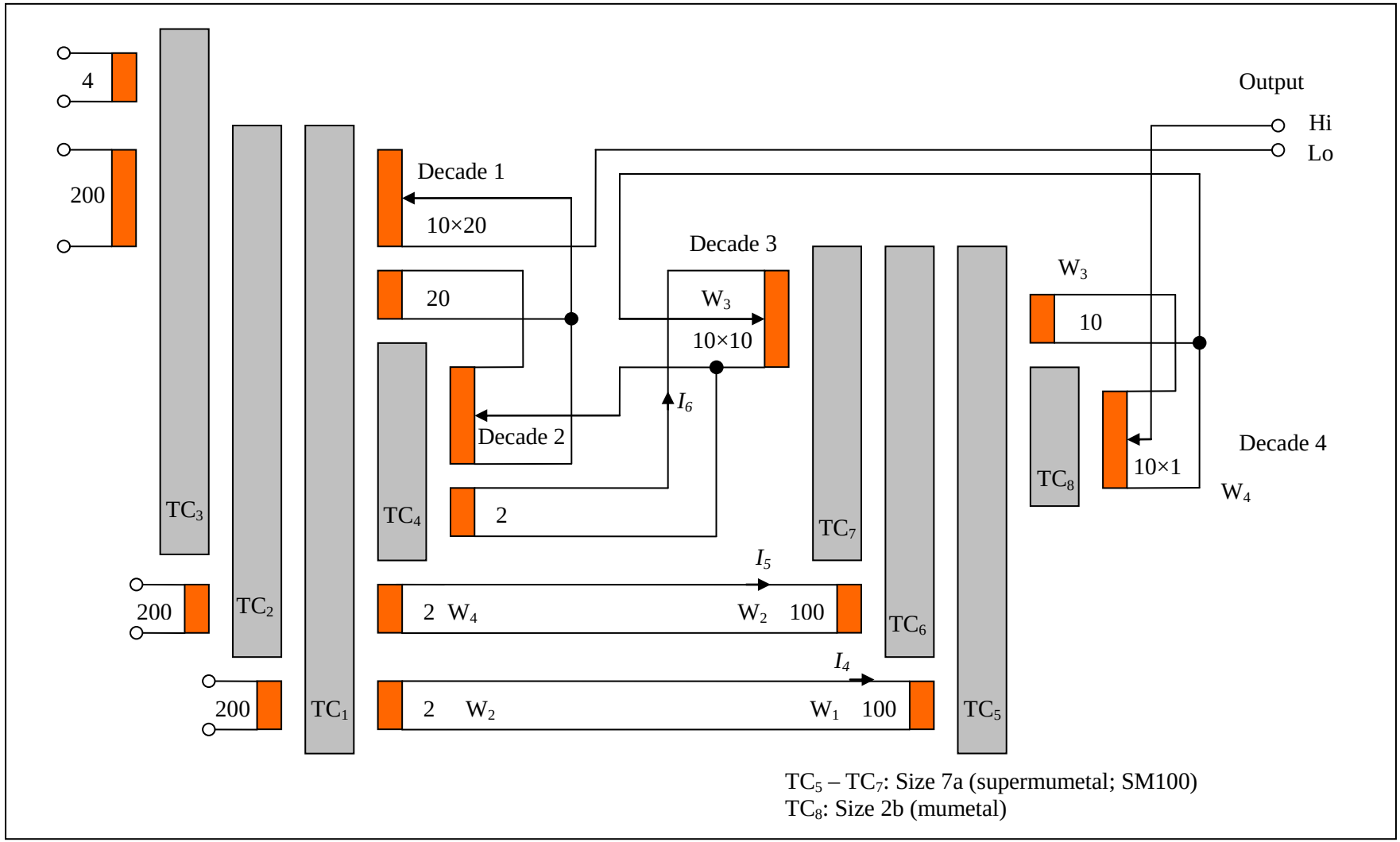


Fig. 4.1.2 Decades 1 to 4 (outline design)

4.2 The ratio transformer 2 (decades 3 and 4)

This is also a three-stage design with energising derived from ratio transformer 1. The ratio primary has such high input impedance that it is easily driven from an equalised decade 2 extra secondary (2 turns). The primary also acts as an IVD for decade 3. The equalised winding 4 secondary is also tapped as an IVD for decade 4.

Windings 1 and 2: 100 turns of 0.8mm enamelled copper wire (BNNL).

Winding 3: 10 turns of a 12 strand rope of 0.8mm enamelled copper wire. 10 strands are used for the primary/decade 3 IVD and two are connected in parallel (half the resistance) and used to equalise winding 4.

Winding 4: 1 turn of a 10 strand rope of 0.8mm enamelled copper wire plus 3 extra turns through the 2b auxiliary core.

For core sizes and other data see appendix 1.

4.3 The resistive multiplying R-2R DAC

Decades 5, 6 and 7, being less critical are implemented using a 12 bit binary multiplying digital to analogue converter. The input is taken from the outputs of the followers via a two-stage transformer with active drive from the output side [1]. The main windings consist of 200 turns (BNNL). An extra 2 turns provides the reference voltage for the ratio tweek. See section 6.

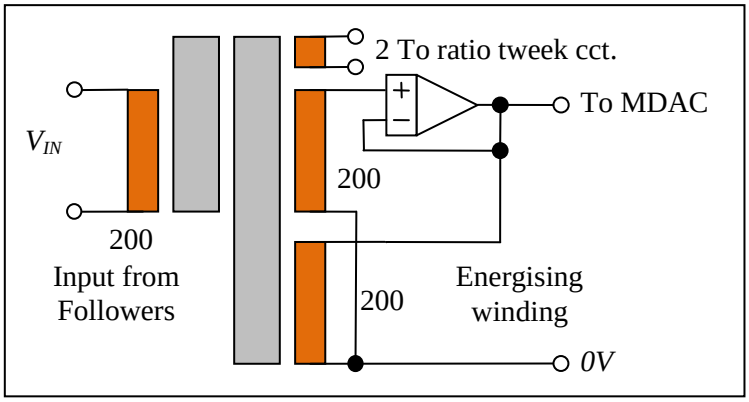


Fig. 4.3.1 MDAC input stage

The output of the MDAC is then passed through a two-stage high-pass filter (JDY special [2]) to remove the DC offset. It is then stepped down with two more two-stage transformers, each reducing the output by precisely 100:1.

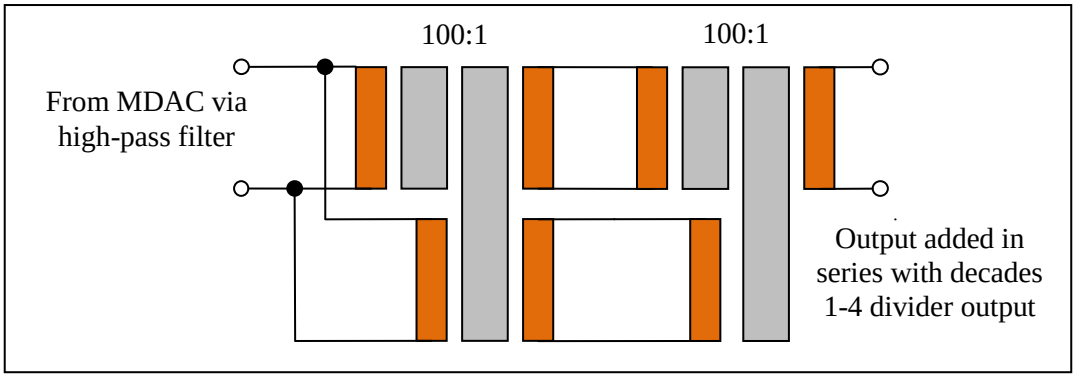


Fig. 4.3.2 Step-down stage

The output is then suitably scaled and added in series with the output of the main divider (decades 1 – 4) output.

1. Part 3, monograph 3: “Two-stage IVDs and RTs”. See section 7.
2. Part 2, monograph 1: “Two-stage filters”. See section 2 and, especially, 2.5.1 (example calculation).

5. Quadrature servo and null detector

The first stage of the null detector is a single-stage impedance (noise) matching transformer with 1, 10 and 100Ω settings providing better than 75% power transfer over the range 0.3 - 300Ω [1]. The low noise (bipolar junction transistor) pre-amp [2] is followed by notch filters for power supply first and third harmonics plus band-pass filters for half and three times local supply frequency and pair of analogue switch type synchronous detectors (in-phase and quadrature) and second order low-pass filters [3]. The in-phase signal is brought to a null either manually (using the front panel switches and meter) or automatically with a micro-controller. The quadrature null balance operates continuously with a "quadrature servo" [4].

The quadrature servo consists of an analogue multiplier which takes its in-phase reference AC input from the main carrier generator circuit. The other input is the integrated output of the quadrature synchronous rectifier. The output of the multiplier is converted to quadrature and transformed to the required level by a low phase error ferrite pot core transformer. The primary is the feedback element of an inverter circuit (input voltage converted to a current) so that it acts as a mutual inductor. The resistor/capacitor series combination ("snubber") in parallel with the primary contributes negligible phase error at the operating frequency but limits the gain of the circuit at very high frequency [5].

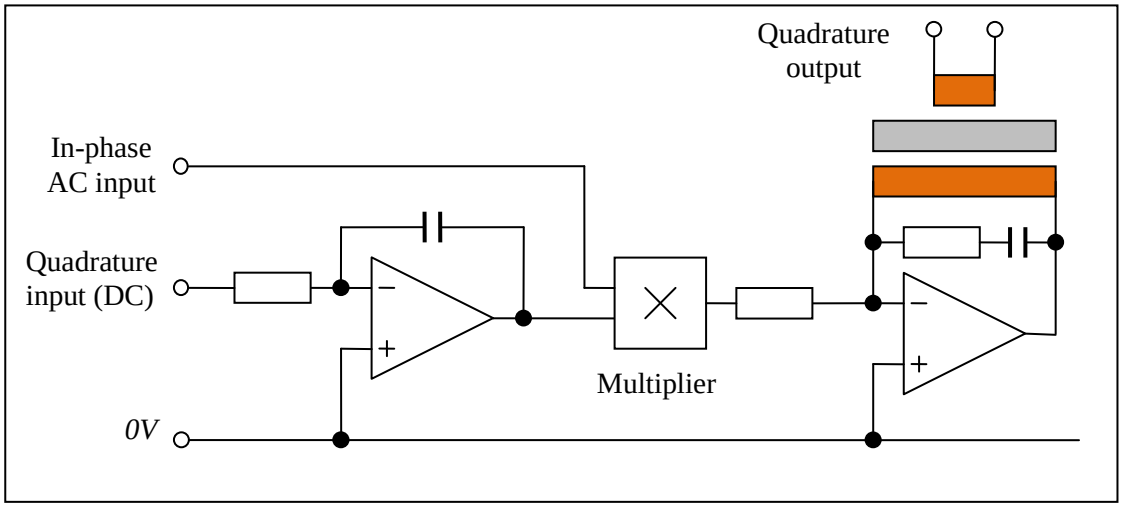


Fig. 5.1 Final stages of the quadrature servo

The output of the quad servo is then added in series with the output of the main divider (decades 1 – 4) output.

1. Part 3, monograph 5: "Noise matching transformers". See section 2.
2. Part 5, monograph 2: "Low noise BJT pre-amplifiers".
3. Part 5, monograph 1: "Null detectors – the basics".
4. In the original A7 automatic bridge the mutual inductance of the ferrite pot core was varied by rotating one element with a servo motor – hence the term: "quadrature servo".
5. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators". See section 5

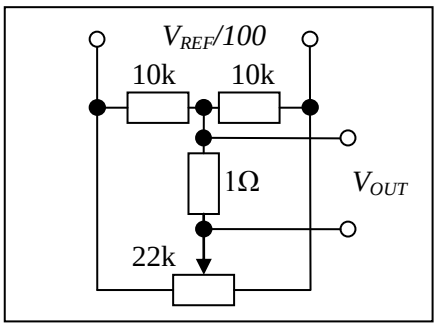
6. Manual "tweeks"

Two final (manual) adjustments were incorporated: A "ratio tweek" and an input impedance or "Wolfendale tweek". Both were justified by the fact that they were simple (low cost) and substantially contributed to the achievement of the target specification (<100ppb).

6.1 The ratio tweek

The main limitation on accuracy is due to stray flux "leakage", especially from the ends of the thin strip of high permeability metallic glass that makes up the energising core. If a part of the first decade rope happens to coincide with one or both ends the result can be a small non-linearity and a slope error.

Whereas it is possible to design the layout of an instrument [1] to minimise offsets (the instrument "zero") it is less practicable to avoid the non-linearity and slope errors. Fortunately both are stable and a simple (low cost) manual trim is possible for the latter.



Figs. 6.1 The ratio tweek circuit

The ratio tweek takes its (isolated) input from a secondary of the two-stage transformer used to provide the reference input to the MDAC (see fig. 4.3.1). It is accurately one hundredth (10^{-2}) of the reference voltage (the difference at the outputs of the followers). The output is added in series with the ratio winding primary with a further attenuation in the range $\pm 10^{-4}$. The total range of slope adjustment is, therefore, ±1ppm. The 1:1 ratio is set to within ±0.01ppm.

6.2 The input impedance tweek

The three-stage bootstrap results in very high input impedance but not quite high enough for some applications.

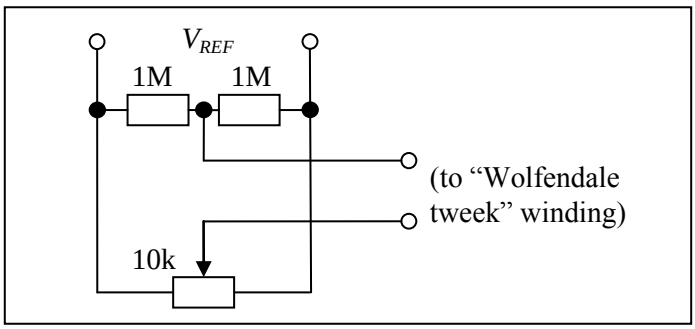


Fig.6.2 The impedance tweek circuit

The reference input is taken directly from the outputs of the voltage followers. The high source resistance presented to the tweek winding results in a very small in-phase current (and flux), adjusting the real part of the input impedance. The adjustment is made (1:1 ratio ±0.01ppm) with 100Ω in series with the transformer primary/follower inputs.

1. A game we called "hunt the nanovolt": Using the very sensitive null detector to locate and eliminate transmitters and receivers of stray flux - with twisted pairs or coax connections, correct orientation of components and routing of connections. The only component that required magnetic screening was the main ratio transformer and the noise matching transformer.

Appendix: Core data

Core size	Minimum relative permeability μ_R	Effective magnetic length L_E (mm)	Effective magnetic area A_E (mm^2)	Permittance A_L ($\mu H/turn^2$)	Case OD (mm)	Case ID (mm)	Case axial (mm)
11a	50,000	279	241	54	112	66	16
11c	200,000	279	363	328	112	66	23
7b	100,000	150	91	76	60	36	12
2b	50,000	57	25	27	24	12	8

A 16 bit binary differential capacitance bridge

1. Introduction

The following design is based on an original, by JDY, as part of a high accuracy rotary servo-mechanism. The basic idea is to drive the centre of a single-stage centre-tapped ratio transformer with a control signal, V_C. The output of the charge amplifier/synchronous rectifier, after suitable loop compensation and a power amplifier, drives a torque motor/capacitive transducer to achieve a null balance. The resulting angular displacement is proportional to the control signal. For more detail see the monograph "Rotary capacitive displacement transducers" by the same author [1].

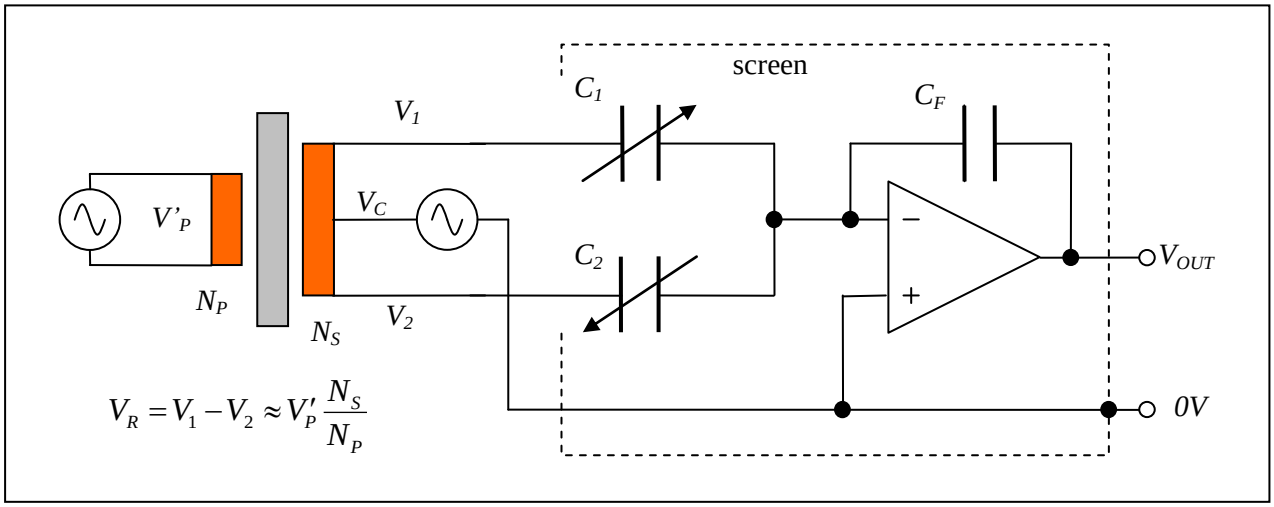


Fig. 1.1 A differential capacitance bridge with centre-tap drive

The control voltage is derived from a reference voltage with a hybrid ratio transformer/inductive voltage divider (RT/IVD) for the most significant 8 bits and a resistive (R-2R type) multiplying DAC for the lower 8 bits. The input to the servo is then a simple 16 bit binary code which is accurately related (proportional) to the angular displacement required.

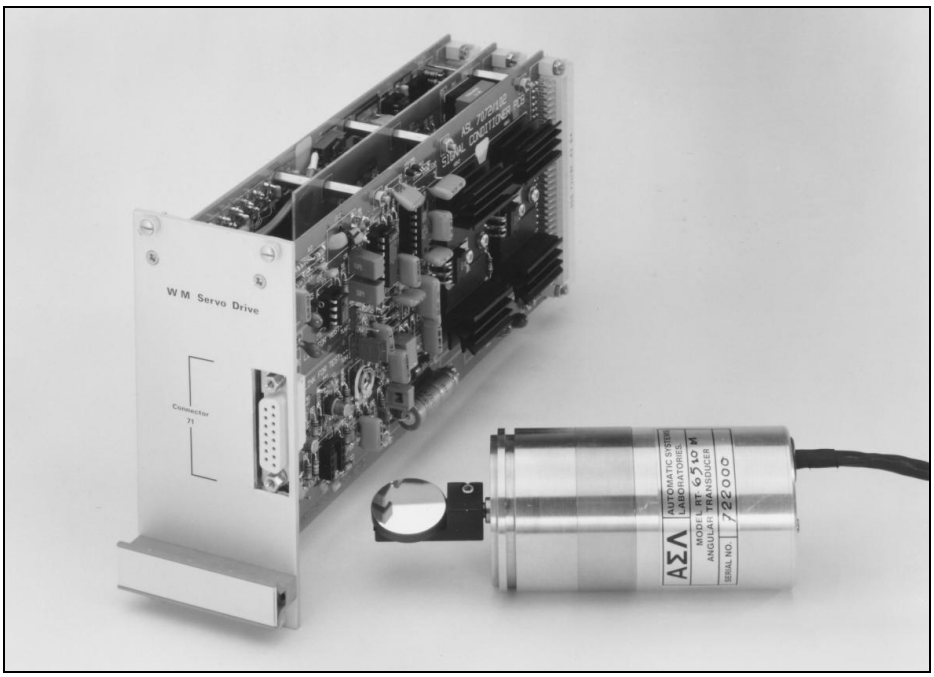


Fig. 1.2 A Combined actuator/transducer and interface
(picture courtesy ASL Ltd)

1. Part 1, monograph 3: "Rotary capacitive displacement transducers"

2. Outline design

Even more ingenious is the use of a high gain block (HGB) in an "inside-out" voltage follower mode [1 and 2]. The very high input resistance allows the use of miniature, low cost FET switches for selecting the required RT/IVD taps. Both inputs of the HGB are maintained at local 0V ensuring high accuracy (no common mode) and avoiding the need for a floating power supply [2].

The "inside-out" principle is to place the input signal in series with the feedback connection. The action of feedback is to ensure the two inputs of the HGB are at the same voltage – in this case 0V. The output of the HGB is, therefore, accurately the same as the control signal but with much lower source impedance. This is exactly the same as a high accuracy follower apart from the power supply centre is now connected to local 0V and common mode is not a problem – a direct result of the fact that the control is isolated from local 0V made possible with a transformer/IVD.

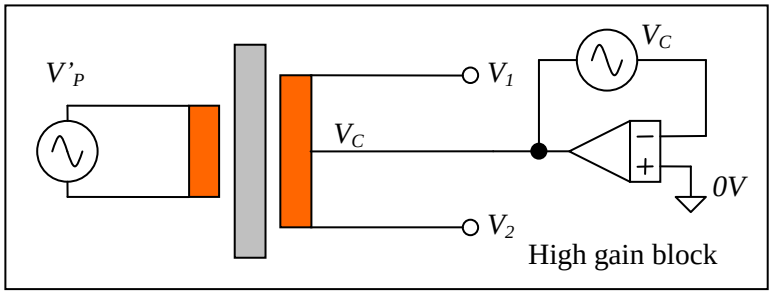


Fig. 2.1 The "inside-out" principle.

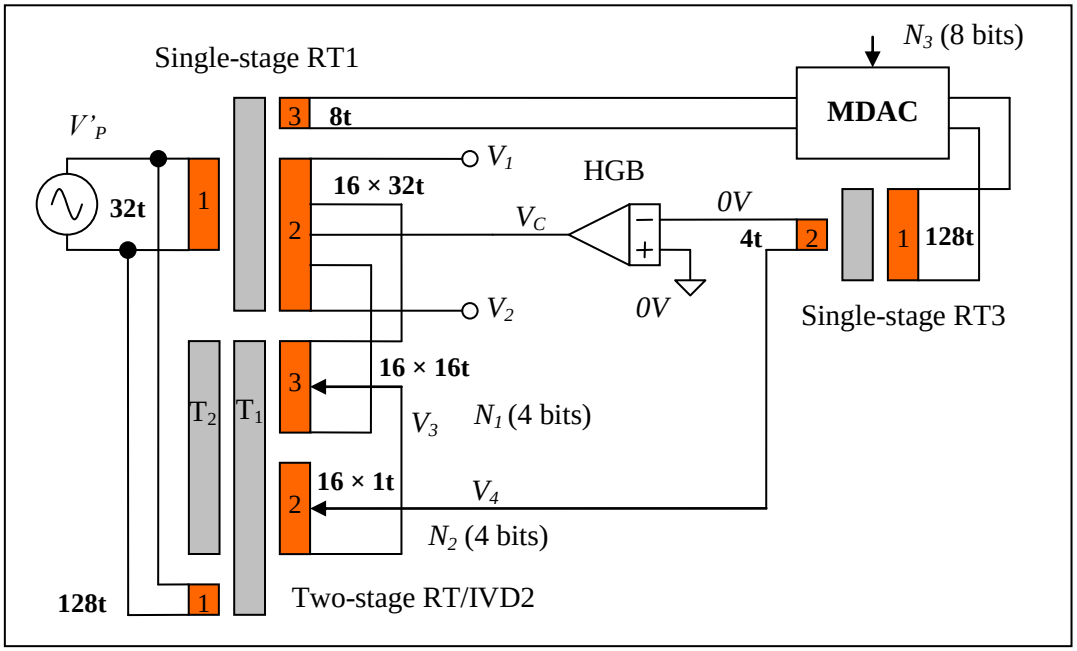


Fig. 2.2 A differential capacitance bridge in more detail

According to Kirchhoff's law the outputs of (centre tapped) transformer RT1 are (fig. 2.2): -

$$V_1 = V_C + \frac{V_R}{2} \quad \text{and} \quad V_2 = V_C - \frac{V_R}{2} \quad \text{with} \quad V_R = V_1 - V_2$$

1. Part 4, monograph 1: "High gain blocks"
2. Part 4, monograph 2: "High accuracy voltage followers". See section 2.2

At balance, assuming an ideal op-amp (fig. 1.1), the output of the charge amplifier is: -

$$V_{OUT} = \frac{V_1 C_1 + V_2 C_2}{C_F} = 0V$$

$$\Rightarrow \quad \left(V_C + \frac{V_R}{2}\right)C_1 + \left(V_C - \frac{V_R}{2}\right)C_2 = 0$$

$$\Rightarrow \quad \frac{V_C}{V_R} = \frac{1}{2} \times \frac{C_2 - C_1}{C_2 + C_1}$$

It is shown elsewhere [1] that the capacitance ratio of a rotary differential capacitive transducer is related to angle (in degrees) by the formula: -

$$\frac{C_2 - C_1}{C_2 + C_1}(\theta) = \frac{\theta}{45}$$

The voltage ratio is, therefore, proportional to angle: $\frac{V_C}{V_R} = \frac{\theta}{90}$

Transformer RT1 has 32 turns for the primary and 32 for each strand of the main secondary winding. The outputs (see fig 2.2) are, therefore: -

$$V_1 = V_C + N_0 V_P \quad \text{and} \quad V_2 = V_C - N_0 V_P$$

Where V_P refers to the voltage across each strand of the secondary side. With a single-stage transformer V_P is only approximately the same as the primary drive voltage $(V_P \approx V_P')$. It is, however, accurately related to the reference voltage: -

$$V_R = V_1 - V_2 = 2N_0 V_P \Rightarrow \quad V_P = \frac{V_R}{2N_0}$$

The integer N_0 refers to the number of strands (32 turns each) either side of the centre tap (connected to V_C). A practical design (fig. 2.2) employs a 16 strand rope so that the range of choice is: -

$$N_0 = 1 \rightarrow 8$$

The control voltage is accurately related to V_P : -

$$V_C = \alpha V_P \Rightarrow \alpha = N_0 \frac{\theta}{90} \quad \text{or} \quad \theta = \alpha \frac{90}{N_0}$$

Where α is the voltage ratio (V_C / V_P) determined by the transformer/IVD (RT/IVD2) and MDAC settings.

The range of the servo is determined, therefore, by the range of α and the choice of N_0 .

1. Part 1, monograph 3: "Rotary capacitive displacement transducers". See section 5.1

RT/IVD2 is a two-stage hybrid ratio transformer/inductive voltage divider. Winding 1 is the energising winding providing most of the flux in toroidal core T1. Winding 3 is the ratio primary (wound around both cores) and is also used as an inductive voltage divider. The voltage across winding 3 is "equalised" to the reference voltage by direct connection to suitable taps on RT1. The very small current flowing through winding 3 results in extra flux in the top core (T_2). The resistance of each strand is approximately the same so that ratio accuracy is maintained – the main principle of an IVD. Winding 2 is a ratio transformer secondary - it is also around both cores so that the voltage induced is also accurately related to the reference voltage. For more details on IVDs and RTs see the relevant monographs [1 and 2].

The voltage at the selected tap of IVD winding 3 is, with the ratio of turns shown explicitly (refer to fig. 2.2): -

$$V_3 = V_C - V_P + N_1 \frac{16t}{128t} V_P \quad \Rightarrow \quad V_C = V_P - N_1 \frac{16t}{128t} V_P + V_3$$

The integer N_1 represents the IVD setting: $N_1 = 0 \rightarrow 15$

Similarly, the voltage at the selected tap of winding 2 is: -

$$V_4 = V_3 + N_2 \frac{1t}{128t} V_P \quad \Rightarrow \quad V_3 = V_4 - N_2 \frac{1t}{128t} V_P$$

The integer N_2 represents the IVD setting: $N_2 = 0 \rightarrow 15$

The control voltage is, therefore: $V_C = V_P - N_1 \frac{16}{128} V_P + V_4 - N_2 \frac{1}{128}$

The output of RT3 depends on the MDAC setting and the ratio of turns. If one assumes an ideal HGB (inverting input at precisely 0V): -

$$V_4 = \frac{8t}{32t} V_P \times \frac{N_3}{256} \times \frac{4t}{128t}$$

The integer N_3 represents the MDAC setting: $N_1 = 0 \rightarrow 255$

The polarity of RT3 is chosen so that the control voltage is: -

$$V_C = V_P - N_1 \frac{16}{128} V_P - N_2 \frac{1}{128} V_P - \frac{N_3}{256} \frac{1}{128} V_P$$

The ratio α is, therefore: $\alpha = 1 - \frac{1}{128}\left(16N_1 + N_2 + \frac{N_3}{256}\right)$

It is now clear that the control input is a simple binary code: -

State	Code (Hex)	Ratio
$N_1 = N_2 = N_3 = 0$	0000	$\alpha = 1$
$N_1 = 8$ and $N_2 = N_3 = 0$	8000	$\alpha = 0$
$N_1 = N_2 = 15$ and $N_3 = 255$	FFFF	$\alpha \approx -1$ (actually: $\alpha = -0.99997$)

1. Part 3, monograph 1: "IVDs and RTs – the basics"
2. Part 3, monograph 2: "Single-stage inductors and transformers"
3. Part 3, monograph 3: "Two-stage IVDs and RTs"

3. Winding suggestions

The design of inductors and transformers for operation with capacitive transducers is much easier than for resistance bridges. The frequency of operation (typically in the range 1 – 10kHz) is sufficiently high for winding resistance to be negligible but not so high that interwinding capacitance becomes a problem. Much finer wire can be used for the windings and, as a result, the toroidal cores can be quite small. The main thing to remember is that the number of turns must be exact and energising windings should be uniformly distributed around the core (balanced no-net-loop or BNNL) [1]. Uniform distribution helps to ensure that most of the flux is contained within the toroid casing. Rope strands are connected in series, paying careful attention to polarity (start/finish).

The following are suggestions and not critical. All windings are with enamelled copper wire [2].

3.1 The single-stage ratio transformer (RT1)

Core size: 4a mumetal.

Winding 1 (energising primary): 32 turns balanced no-net loop (BNNL) of 0.5mm OD wire.

Winding 2 (main ratio winding): 32 turns of a 16 strand rope (BNNL) of 0.25mm OD wire.

Winding 3: 8 turns of 0.25mm OD wire.

The equalising taps (above and below the centre tap) are from one strand each (32 turns) so that the voltage difference is precisely $2V_P$.

3.2 The two-stage inductive voltage divider (RT/IVD2) [3]

Both toroidal cores are size 3a mumetal. All windings are with 0.25mm OD enamelled copper wire.

Winding 1 (energising primary): 128 turns (BNNL) around the bottom core only.

Winding 2 (ratio secondary): 1 turn of a 16 strand rope around both cores.

Winding 3(ratio primary and IVD): 16 turns (BNNL) of a 16 strand rope around both cores.

The energising winding bootstraps winding 3 – the induced voltage is approximately $2V_P$ (128:256 turns). The ratio is "equalised" by direct connection to RT1 (precisely $2V_P$) with a low resistance twisted pair. The small voltage difference results in a very small current flowing and a little extra flux in the second core. The current flowing through the ratio winding does not introduce significant errors (<1ppm) because it is an IVD – the winding resistances are very nearly the same.

3.3 The single stage ratio transformer (RT3)

The toroidal core is size 1a mumetal. Both windings are with 0.25mm OD enamelled copper wire.

Winding 1: 128 turns (BNNL).

Winding 2: 4 turns.

Accuracy could be improved to 20 bits by replacing the 8 bit MDAC with a 12 bit device and upgrading RT3 to a two-stage transformer. It may also be necessary to use slightly larger cores and thicker wire.

1. Part 3, monograph 1: "IVDs and RTs – the basics".
2. Data on toroidal cores and copper wire is appended to the monograph: "Two-stage IVDs and RTs".
3. Part 3, monograph 3: "Two-stage IVDs and RTs".

High gain blocks

1. Introduction

A high gain block (HGB) is a differential amplifier with dynamic characteristics designed to ensure stable operation with negative feedback [1]. The important properties of an ideal HGB include very high gain (at low frequency), low noise, very high input impedance, low output impedance and insensitivity to power supply variations. A basic example is the ubiquitous integrated circuit (IC) operational amplifier [2] on which is based a wide variety of circuit configurations. This monograph includes a few of them as they are also useful building blocks for many high accuracy applications, including multi-stage HGBs. The multi-stage HGB can be thought of as an extension of the op-amp concept. An op-amp can be thought of as a single-stage HGB.

Despite the high (open loop) gain of some op-amps the closed loop accuracy is not quite sufficient for some high accuracy applications.

Stability can also be an issue. Some manufacturers, for example, push gain-bandwidth product to the limit, for marketing reasons, at the expense of stability. The output resistance combines with any load capacitance (e.g. cable capacitance) to produce extra phase shift in the feedback loop, reducing stability margin or, even worse, causing sustained oscillation at high frequency.

Multi-stage HGBs combine two or more op-amp circuits, in series, to achieve much higher open loop gain and, therefore, improved closed loop accuracy. There are three main types [3]: -
1. One or more "one-plus-integrators" (OPIs) followed by an integrator, each based on an op-amp IC with external feedback components (resistors and capacitors) to tailor the frequency response. This type is robustly stable even with significant capacitive load (>1nF).
2. Two or more op-amps (open loop) in series, with "feed-forward".
3. A hybrid of the first two with the order reversed (integrator first). The first stage is an op-amp, with a fully differential input, followed by an OPI with RC feedback.

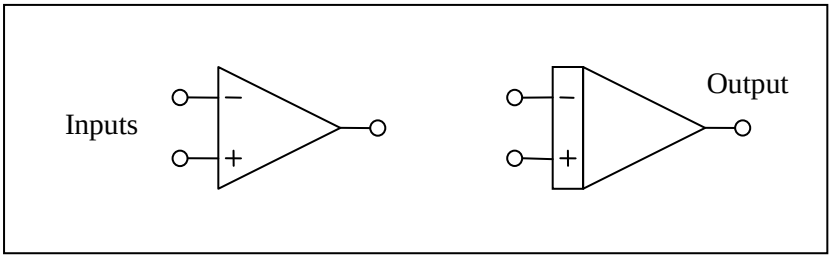


Fig.1 Circuit symbols for a single-stage (op-amp) and a multi-stage HGB

One could consider a composite op-amp (e.g. a "long tail pair" with tailored frequency response followed by an op-amp) as another type of two-stage HGB but the main application is for low noise pre-amplifiers where high accuracy is not particularly important [4].

Applications of multi-stage HGBs include high accuracy voltage followers [5] inverting and non-inverting amplifiers, integrators and differentiators [6].

1. Block as in "building block". Multi-stage HGBs have a sufficiently well defined function and common characteristics to justify a single concept and circuit symbol.
2. Thomas H. Lee: "IC Op-Amps Through the Ages" rev. Nov. 2002: "The operational amplifier concept emerged from extensive development of electronic analogue computers in the 1940s."
3. All three types were developed by JDY in the mid 1980s. Type 2 was published ("High precision composite op-amps", Electronics and Wireless world, 1987). It has also appeared in a Comlinear Corporation data sheet (no author specified) dated April 1994, as a "five-decade integrator".
4. Monographs: "Low noise BJT pre-amps" and "Low noise JFET pre-amps".
5. Monograph: "High accuracy voltage followers".
6. Monograph: "High accuracy amplifiers, integrators and differentiators"

The following block diagram embodies the principle of an ideal multi-stage HGB: -

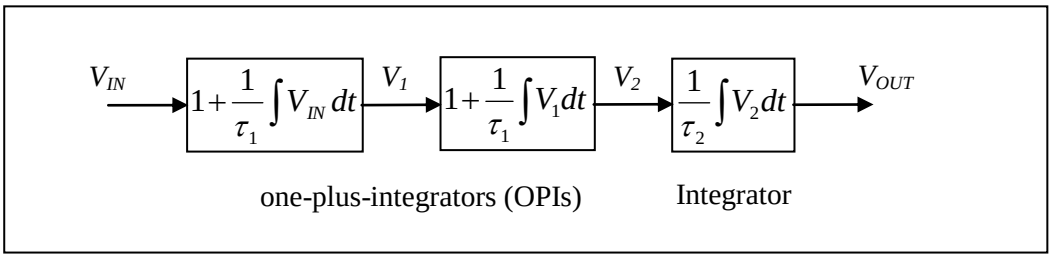


Fig. 1.2 Model for a three-stage HGB (type 1)

In principle the order is not important. The integrator is usually the final stage (types 1 and 2) for practical reasons – an integrator is better able to drive a capacitive load.

A very good mathematical model for an *N*-stage HGB, over a wide range of frequency (typically 1Hz to 1MHz), is the transfer function in the usual complex representation $(s = j\omega)$: -

$$H_N(s) = \frac{V_{OUT}}{V_{IN}} = \left(1 + \frac{1}{\tau_1 s}\right)^{N-1} \frac{1}{\tau_2 s}$$

At very low frequency the "DC gain" of the op-amps combine to provide enormous open loop gain. In the following example ($N = 4$) this is a minimum of 10^{20} (100dB or 10^5 per op-amp). As the frequency of the test signal increases the gain starts to fall at a rapid rate (*N*th order) but then reverts to first order before it passes through 0dB. This ensures closed loop stability with 100% negative feedback. At higher frequency (typically 5MHz) the op-amps run out of steam and slope increases again.

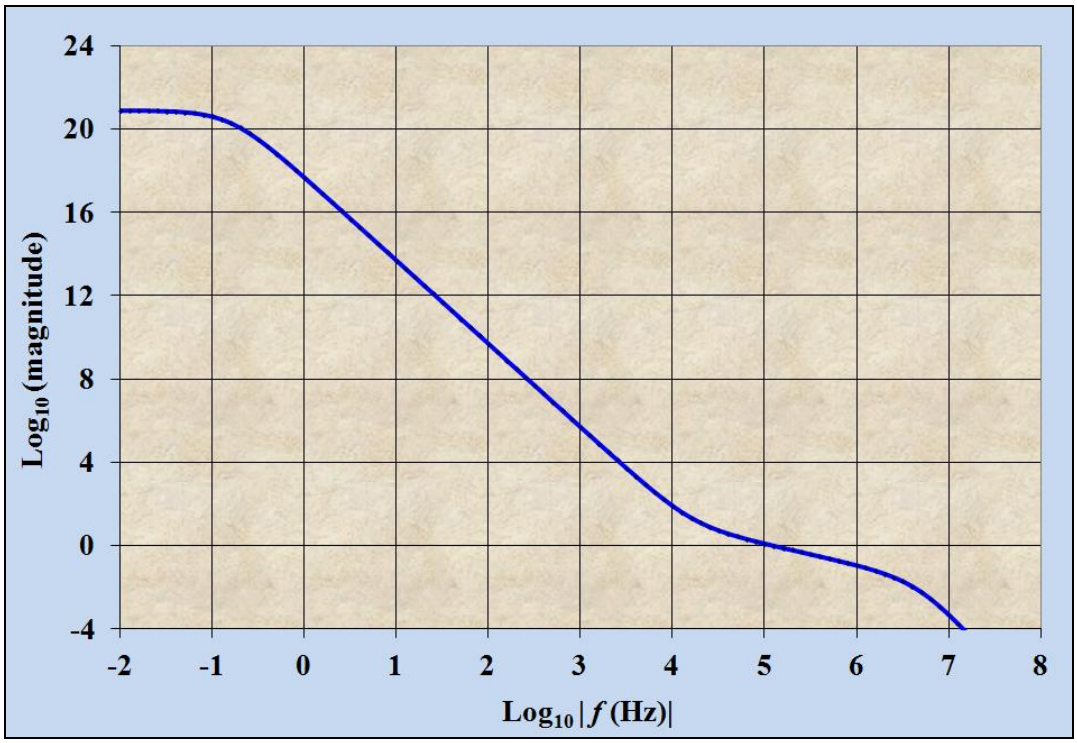


Fig. 1.3 Bode plot of the open loop gain of a four-stage HGB

In practice there is not much advantage in going beyond a three-stage HGBs though this author has built and tested a five-stage HGB (just for the fun of it) in (very carefully constructed) prototype form.

2. Feedback, stability and accuracy (general theory)

2.1 The closed loop transfer function

The most basic and frequently employed circuits employ negative feedback with two resistors or a resistor and capacitor. These are analysed with the following assumptions: -
a). Infinite input impedance and zero output impedance.
b). Perfect common mode and power supply rejection.
c). Linear behaviour (e.g. no slew rate limiting).

The circuit is inverting or non-inverting depending on which input is connected to the source 0V (V_2 = 0V for the inverting mode and V_1 = 0V for the non-inverting mode).

Whereas the input and output are usually thought of as single-ended (i.e. relative to local 0V) they are best thought of as quasi-differential signals: the input and output are a pair of connections (twisted or co-axial) which are routed to prevent magnetic interference and earth loops [1].

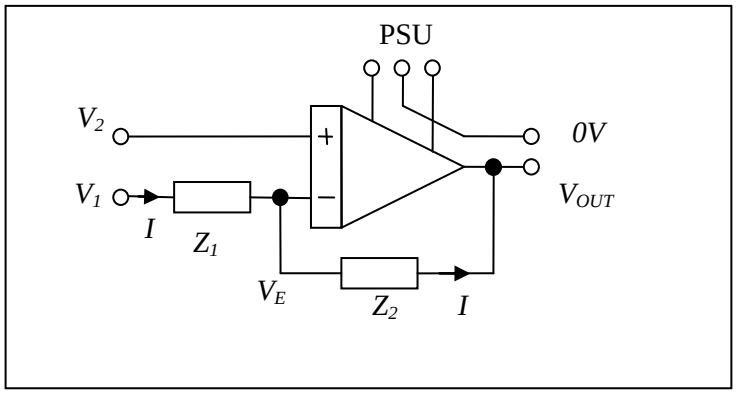


Fig. 2.1.1 An HGB with negative feedback

According to Kirchhoff's and Ohm's laws: $$I = \frac{V_1 - V_E}{Z_1} = \frac{V_E - V_{OUT}}{Z_2}$$

By definition the open loop transfer function is: $$H(s) = \frac{V_{OUT}}{V_2 - V_E}$$

From the first equation: $$V_{OUT} = V_E\left(1 + \frac{Z_2}{Z_1}\right) - V_1\frac{Z_2}{Z_1}$$

From the second: $$V_E = V_2 - \frac{V_{OUT}}{H(s)}$$

Combine the two to eliminate V_E: $$V_{OUT} = \left(V_2 - \frac{V_{OUT}}{H(s)}\right)\left(1 + \frac{Z_2}{Z_1}\right) - V_1\frac{Z_2}{Z_1}$$

From which: $$V_{OUT}\left(1 + \left(1 + \frac{Z_2}{Z_1}\right)\frac{1}{H(s)}\right) = V_2\left(1 + \frac{Z_2}{Z_1}\right) - V_1\frac{Z_2}{Z_1}$$

1. Monograph: "High accuracy voltage followers". For examples of signal routing see sections 3 and 6.

I shall define a parameter $\beta(s)$ which is the reciprocal of the feedback factor and may be a function of frequency. This will simplify the algebra: -

$$F(s)=\frac{Z_1}{Z_1+Z_2} \quad \text{and} \quad \beta(s)=\frac{1}{F(s)}=1+\frac{Z_2}{Z_1} \quad \Rightarrow \quad V_{OUT}\left(1+\frac{\beta(s)}{H(s)}\right)=V_2\beta(s)-V_1\frac{Z_2}{Z_1}$$

The output is, therefore: $$V_{OUT}=\left(V_2\beta(s)-V_1\frac{Z_2}{Z_1}\right)\left(\frac{H(s)}{H(s)+\beta(s)}\right)$$

Inverting mode $(V_2=0V)$: $$\frac{V_{OUT}}{V_1}=-\frac{Z_2}{Z_1}\left(\frac{H(s)}{H(s)+\beta(s)}\right)$$

Non-inverting mode $(V_1=0V)$: $$\frac{V_{OUT}}{V_2}=\left(1+\frac{Z_2}{Z_1}\right)\left(\frac{H(s)}{H(s)+\beta(s)}\right)$$

The term in the second bracket is of particular interest as it determines the closed loop stability and accuracy in both cases.

$$\frac{V_{OUT}}{V_1}=-\frac{Z_2}{Z_1}D(s) \quad \text{and} \qquad \frac{V_{OUT}}{V_2}=\left(1+\frac{Z_2}{Z_1}\right)D(s)$$

The "D" factor is, therefore: $$D(s)=\frac{H(s)}{H(s)+\beta(s)}=\frac{L(s)}{L(s)+1} \quad \text{where} \quad L(s)=H(s)F(s)=\frac{H(s)}{\beta(s)}$$

$L(s)$ is the "loop gain" – the transfer function of the HGB and feedback network combined.

The "*D*" factor expresses the difference between the actual and the ideal closed loop transfer functions. If the magnitude of the loop gain is infinite then $D(s)=1$ and the ideal transfer functions are as expected: -

$$|L(s)|\to\infty \quad \Rightarrow \quad \frac{V_{OUT}}{V_1}=-\frac{Z_2}{Z_1} \quad \text{and} \quad \frac{V_{OUT}}{V_2}=1+\frac{Z_2}{Z_1}$$

One can extract the error component with a little algebra: -

$$D(s)=1-\frac{\beta(s)}{H(s)+\beta(s)} \quad \text{or} \quad D(s)=1-\frac{1}{L(s)+1}$$

At low frequency the loop gain is very large and $D(s)\approx 1$ but at some high frequency one would expect the closed loop transfer function to have a resonant peak, depending on the stability margin, before falling below 0dB.

Finally, note the requirement which must relate to stability: -

$$D(s)=\frac{L(s)}{L(s)+1} \quad \Rightarrow \quad \text{what if} \quad L(s)=-1?$$

The accuracy and stability of each type of HGB are investigated in general in the relevant sections in this monograph. Further detail can be found in other monographs [1 and 2]

1. Monograph: "High accuracy voltage followers".
2. Monograph: "High accuracy amplifiers, integrators and differentiators".

2.2 The Bardayquist stability criterion.

The condition $(L(s)=-1)$ is the Barkhausen criterion for sustained oscillation (i.e. 100% positive feedback). The general Nyquist criterion for stability, on the other hand, is quite complicated as is JDY's preferred method: the Routh-Hurwitz test. For our purposes the condition for adequate stability can be summarised in a short statement (the "BarDayQuist" stability criterion) and illustrated with the aid of a new type of plot of frequency response in the complex plane (the "BoDayQuist" plot).

The loop consisting part of the positive real axis and the locus representing the loop gain, $L(s)$, in the complex plane, must not enclose, nor get too near to, the -1 point.

The new plot combines the best features of the Nyquist plot and the Bode plot. The magnitude is on a logarithmic scale (radial distance) so that the huge gain remains visible. The phase shift is retained on the same plot so that the phase margin can also be seen. The following is a rather extreme example chosen to illustrate its utility.

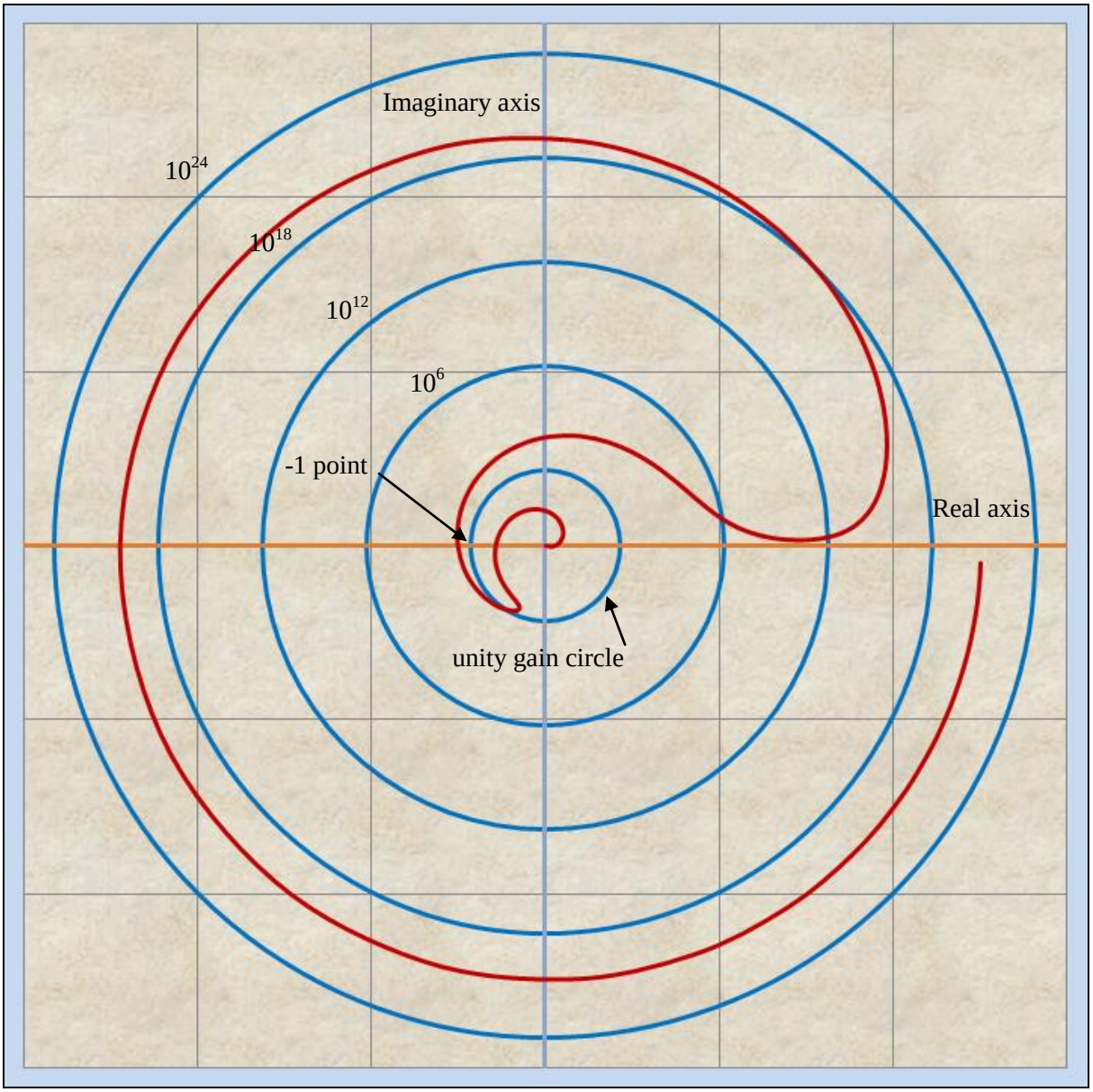


Fig. 2.2.1 The "BoDayQuist" plot illustrates how the stability criterion is satisfied (4-stage HGB)

The radial distance is: $$R = \log_{10}\left|\frac{H(f)}{H_{MIN}}\right|$$

The example shown is based on the LF356 (GBWP = 5MHz with low frequency gain of 100dB) with an HGB OPI frequency of 16kHz.

The gain, at very low frequency, is due to four op-amps in series (10^{20}). As frequency increases (from DC) the locus advances clockwise with phase (lag) increasing rapidly to very nearly 360 degrees (90 degrees per integrator stage). The gain drops rapidly and, as it approaches the unity gain circle, the phase lag reduces to just over 90 degrees. The locus passes through the unity gain circle (0dB) with a phase margin of about 60 degrees. As the gain continues to fall the phase starts to increase again, due to the limited GBWP of the op-amps and the locus spirals into the centre. With the model employed each of the four op-amp stages contributes up to 90 degrees plus another 90 degrees from the integrator stage. The ultimate phase shift is 450 degrees (5 × 90). The locus eventually approaches the centre point along the negative imaginary axis (not visible unless the range of frequency is extended upward).

If any part of the HGB overloads the gain drops and the locus could collapse towards the centre and pass through the -1 point. This can be avoided by making the locus to first unwind. This is achieved with back-back diodes across the OPI capacitors in type 1 HGBs [1]. When the diodes conduct each OPI becomes a follower stage with a gain of 1 and negligible phase shift. Type 3 HGBs (two-stage) also employ back-back zener diodes to avoid non-linear oscillation or latch-up. See section 4.3.

Type 2 HGB circuits (two-stage) appear to be inherently stable, though very fast op-amps should be avoided.

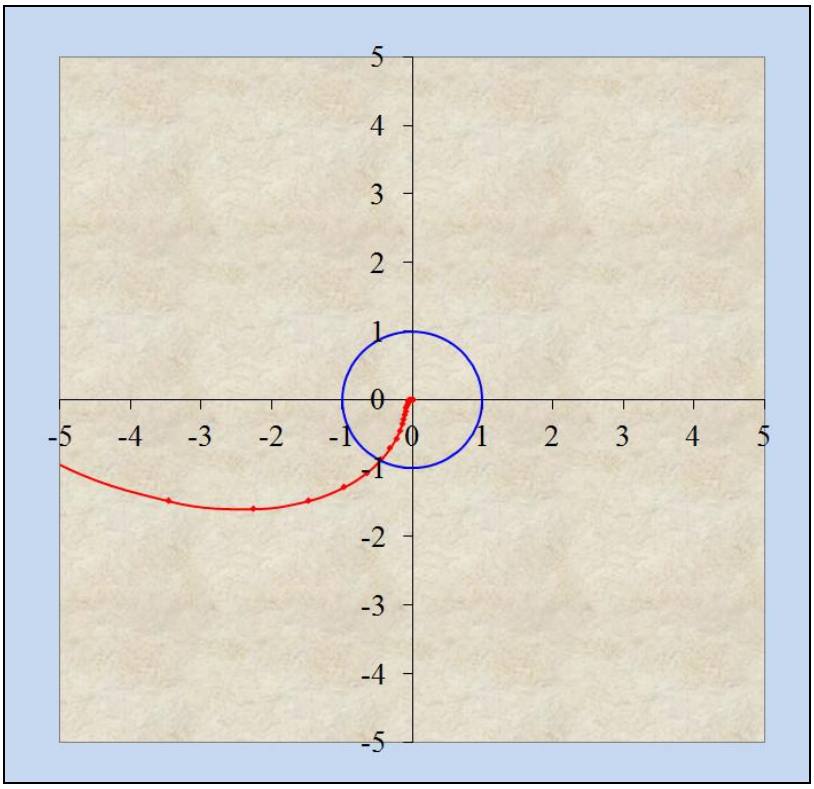
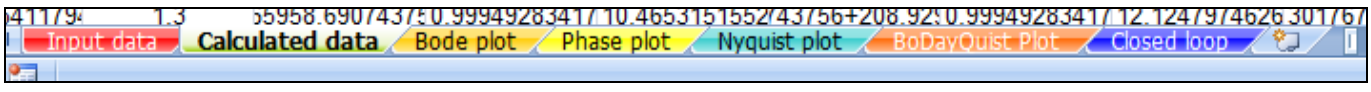


Fig.2.2.2 A conventional Nyquist plot for comparison with fig. 2.2.1

Various plots in these monographs were easily produced with a spreadsheet using complex number arithmetic (IMSUM, IMPRODUCT etc). One worksheet, including slider controls, is provided for entering various parameters (e.g. coefficients of transfer functions, overall gain, start frequency and number of decades). A second worksheet does all the calculations in stages. Further worksheets contain the various plots required (magnitude, phase, Nyquist, BoDayQuist and closed loop magnitude and errors). For a copy please contact the author.



1. Monograph: “High accuracy voltage followers”. See section 6 and fig. 6.1.1.

2.3 Phase margin

The part of the BarDayQuist criterion "too near to the -1 point" is a little vague. A useful and quantitative measure of stability is the phase margin. This is the difference between the actual phase shift, θ_{UG}, as the magnitude of the loop gain passes through 0dB (unity gain), and that required for sustained oscillation. The phase shift for sustained oscillation is 360 degrees but 180 degrees is due to the feedback being applied to the inverting input. The definition of loop gain does not include this hence the critical point in the complex plane is a magnitude of 1 and a phase shift of 180 degrees (the -1 point). One is usually concerned with phase "lag" which, for historical reasons, is defined as a numerically negative angle and clockwise rotation from the real axis. For convenience, however, it is usual to express the phase margin as numerically positive and in degrees. The usual definition of phase margin is: -

$$\theta_{PM}(\text{deg}) = 180 - |\theta_{UG}(\text{deg})|$$

In the Nyquist plot fig. 2.3.1, for example, the phase margin is the angle between the negative real axis and the line joining the centre point to the point at which the locus passes through the unity gain circle (about 60 degrees).

A general rule of thumb is to aim for a minimum (worst case) phase margin of 45 degrees. The result is a resonant peak but this is not a major problem in many applications. Most high accuracy applications employ a low distortion sine wave signal which does not contain components at the frequency that would excite that resonance.

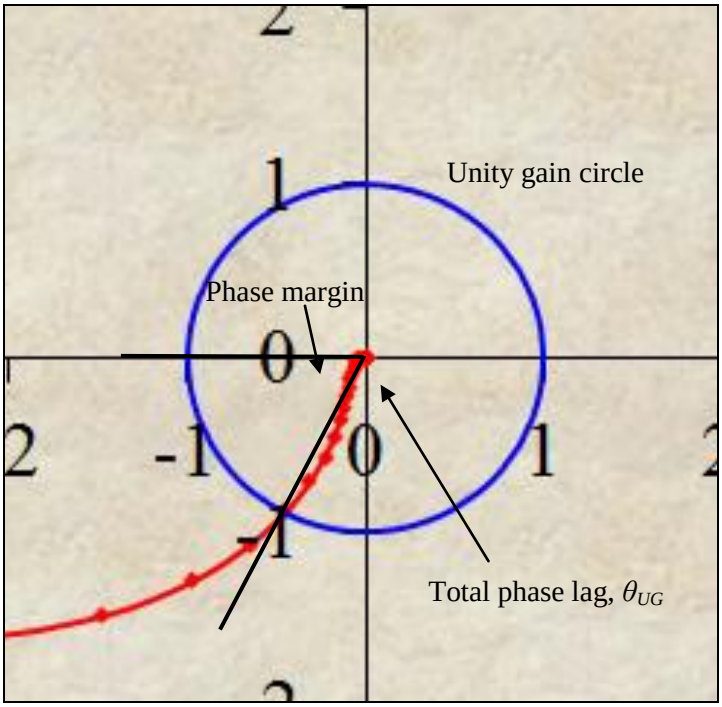


Fig. 2.3.1 The definition of phase margin

Calculations can often be performed most conveniently (with a scientific calculator) using the polar form. For an integrator with an "extra pole", for example (see section 3.1.1): -

$$L(s) = \frac{1}{\tau_B s(1+\tau_E s)} = \frac{1}{\tau_B s(1+j\tan(\theta_E))} = \frac{\exp(-j\pi/2)}{\tau_B \omega}\cos(\theta_E)\exp(-j\theta_E) \quad \text{with} \quad \tan(\theta_E) = \tau_E \omega$$

The integrator contributes -90 degrees and the pole an extra phase lag which is a function of frequency.

$$|L(s)| = 1 \;\Rightarrow\; \omega_{UG} = \frac{\cos(\theta_E)}{\tau_B} \;\Rightarrow\; \tau_E = \tau_B \frac{\tan(\theta_E)}{\cos(\theta_E)}$$

Given τ_B and the extra phase, θ_E (e.g. 45 deg), it is possible to calculate the (max) time constant of the extra pole.

2.4 The size of the resonant peak

The ideal phase margin is 90 degrees – the ideal HGB is an integrator, at least as the loop gain passes through 0dB. A lower phase margin results in a resonant peak in the closed loop response. Computer modelling allows one to explore the trade-off between closed loop accuracy and stability with considerable precision. There is, however, a simple calculation that provides an approximate value for the size of the peak.

I shall assume that as frequency increases the magnitude of $L(s)$ decreases monotonically through 0dB. There should be only one frequency ω_{UG} at which the magnitude is 1. Denote the phase shift at that frequency as θ_{UG}. In polar form the loop gain, with a magnitude of 1 is, therefore: -

$$|L(\omega_{UG})|=1 \;\Rightarrow\; L(\omega_{UG})=\exp(j\theta_{UG})$$

The "D" factor at this particular frequency is, therefore: -

$$D(\omega_{UG})=\frac{L(\omega_{UG})}{L(\omega_{UG})+1}=\frac{\exp(j\theta_{UG})}{\exp(j\theta_{UG})+1}$$

From basic complex number theory: -

$$|D(\omega_{UG})|=\left|\frac{\exp(j\theta_{UG})}{\cos\theta_{UG}+j\sin\theta_{UG}+1}\right|=\frac{1}{\sqrt{(1+\cos\theta_{UG})^2+\sin^2\theta_{UG}}}$$

With a bit of algebra this simplifies: $$|D(\omega_{UG})|=\frac{1}{\sqrt{2+2\cos\theta_{UG}}}$$

Now $1+\cos 2\theta=2\cos^2\theta$ with the remarkably simple result: $$|D(\omega_{UG})|=\left(2\cos\left(\frac{\theta_{UG}}{2}\right)\right)^{-1}$$

Clearly as θ_{UG} approaches $\pm\pi$ the magnitude becomes infinite. The cosine function is even so that it does not matter if one defines the phase shift as positive or negative. In the case that the phase margin is 45 degrees the total phase lag of the loop is 135 degrees: -

$$|\theta_{UG}|=135\,\text{degrees} \;\Rightarrow\; |D(\omega_{UG})|=\frac{1}{2\cos(67.5)}\approx 1.31$$

The peak at this frequency is approximately 31% (+2.3dB).

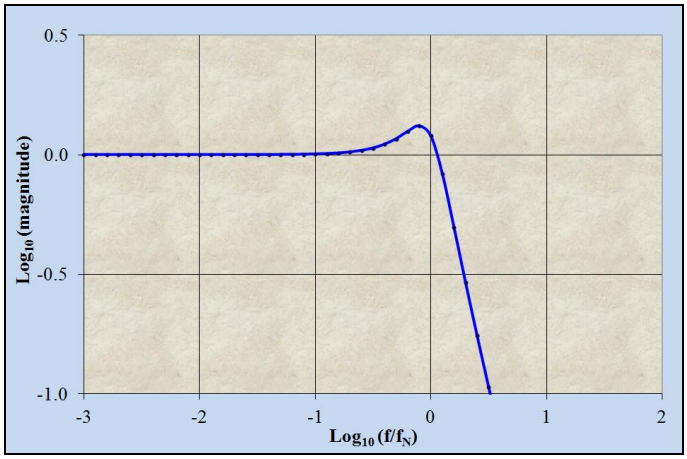


Fig. 2.4.1 The resonant peak with a phase margin of 45 degrees [1]

1. Spreadsheet: "Second order low-pass"

3. Some operational amplifier circuits

3.1 Operational amplifiers

An integrated circuit op-amp typically consists of a number of amplifier stages which combine to provide an open loop gain, at low frequency (typically < 100Hz), between 10^5 and 10^6 (100 – 120dB). For most types the input is differential and the output is single-ended (relative to local 0V). The ideal device has zero noise, infinite input impedance, zero output impedance and is immune to common mode and power supply variations.

Internally compensated op-amps incorporate a small capacitor which introduces a "dominant pole" (typically: $f_P \approx 10-100Hz$) resulting in a first order low-pass characteristic over a wide range of frequency. This ensures stable operation with any amount of feedback. The approximate frequency at which the gain falls below 0dB is known as the "unity gain bandwidth" or "gain-bandwidth product" (GBWP) and is specified in Hz(f_B). Low cost devices have a GBWP which is typically between 1MHz and 10MHz. At higher frequency (f_E)an "extra pole" is responsible for more phase shift resulting in (with 100% feedback) a phase margin less than 90 degrees. An op-amp is thus, in effect, a high gain, second order, low-pass filter with very widely separated real poles. In the complex representation$(s = j\omega)$: -

$$\omega < \omega_E \;\Rightarrow\; H(s) = \frac{G}{(1+\tau_P s)(1+\tau_E s)} \quad \text{with} \quad \tau_P = G\tau_B \quad \text{and} \quad \tau_P >> \tau_B > \tau_E$$

The open loop frequency response of a typical family of devices is illustrated in the following diagrams. The LF356, for example, has a GBWP of 5MHz (at 20ºC) with a generous phase margin of 75 degrees. The LF357 is basically the same device with reduced internal compensation (by a factor of four). As a consequence it has a minimum closed loop gain of four unless external compensation is employed [1]. The advantage of reduced internal compensation is a closed loop response that is, given the same feedback factor, four times faster and four times more accurate. The generous phase margin, combined with a relatively low output resistance, also means that the LF356 can easily drive a load capacitance of 1nF (typically 10m of cable) and remain stable, even with 100% negative feedback.

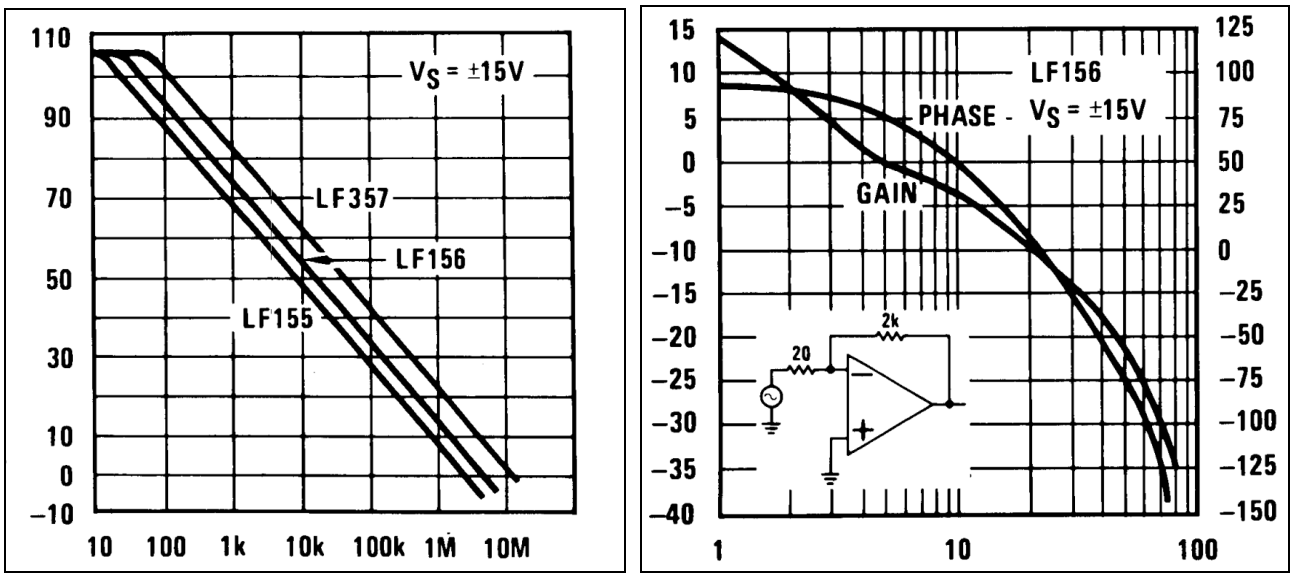


Fig. 3.1 Open loop frequency response for a family of op-amps (gain (dB) and phase (degrees)) [2].

Higher performance op-amps (especially in terms of GBWP) are available but these tend to be quite expensive, dissipate more power and require RF design techniques, skills and test equipment. Non-linear characteristics (e.g. overload and slew rate limiting) also contribute to instability. These effects are difficult to predict or model. It is recommended that prototypes are thoroughly tested: **"the proof of the pudding is in the eating".**

1. Monograph: "High accuracy voltage followers". See section 6.3. See also section 3.6 of this monograph.
2. Courtesy National Semiconductor (data sheet DS005646).

3.1.1 Op-amp models

A very good model of an op-amp, for a small sinusoidal signal (no slew rate limiting), from DC to the unity gain frequency (and a little beyond) is a high gain, first order, low-pass filter followed by an extra low pass characteristic. In the complex representation $(s = j\omega)$: -

$$0 < \omega < \omega_B \;\Rightarrow\; H(s) = \frac{G}{(1+\tau_P s)} P(s) \;\text{ with }\; \tau_P = \frac{G}{\omega_B} = G\tau_B$$

Where G is the gain at very low frequency (i.e. the "DC gain"), τ_P is the time constant of the dominant pole (internal compensation), ω_B is the gain-bandwidth product (GBWP) and $P(s)$ represents the fact that the op-amp transistors are "running out of steam". The extra term, $P(s)$, depends on the type of op-amp and load capacitance. Some manufacturers push the internal compensation to the limit, increasing the GBWP, for marketing reasons, at the expense of closed loop stability. At low frequency the effect of $P(s)$ is negligible and a good model is a high gain first order low pass filter: -

$$\omega << \omega_B \;\Rightarrow\; P(s) \approx 1 \;\Rightarrow\; H(s) = \frac{G}{(1+\tau_P s)}$$

This model is useful for calculating closed loop errors, at low frequency, for circuits employing a single op-amp.

Over a wide range of frequency (typically four decades), above the dominant pole but below the GBWP, the op-amp behaves like a fast integrator. The response is linear but the time constant varies significantly with temperature (typically ≈ -0.5%/deg C): -

$$1 << |\tau_P s| << |\tau_B s| \;\Rightarrow\; H(s) = \frac{1}{\tau_B s} \;\text{ with }\; \tau_B = \frac{\tau_P}{G}$$

In practice the contribution to the magnitude, due to $P(s)$, is small and a sufficiently good model is a low-pass filter or integrator with an "extra pole": -

$$0 < \omega < \omega_B \;\Rightarrow\; H(s) = \frac{G}{(1+\tau_P s)(1+\tau_E s)} \;\text{ and }\; \omega \approx \omega_B \;\Rightarrow\; H(s) = \frac{1}{\tau_B s(1+\tau_E s)}$$

The latter is useful for calculating the stability margin for single op-amp circuits. The polar form (magnitude and phase) proves most convenient: -

$$\frac{1}{\tau_B s} = \frac{\exp(-j\pi/2)}{\tau_B \omega} \;\text{ and }\; \frac{1}{1+\tau_E s} = \frac{1}{1+j\tan(\theta_E)} = \frac{\cos(\theta_E)}{\exp(j\theta_E)} \;\text{ with }\; \tan(\theta_E) = \tau_E \omega$$

Also, with this model, the unity gain frequency is lower than the gain-bandwidth product.

$$|H(s)| = \frac{\cos(\theta_E)}{\tau_B \omega} \;\Rightarrow\; \omega_{UG} = \frac{\cos(\theta_E)}{\tau_B} = \omega_B \cos(\theta_E)$$

HGBs Type 2 and 3 rely on the full open loop characteristic of at least one op-amp and the extra pole model is required.

With type 1 HGBs one need not include the extra pole in the model. The overall gain is designed to be well below 0dB at the GBWP. Type 1 HGBs are employed where robust stability is required, especially when driving cable capacitance.

The LF356**,** for example, has a GBWP of 5MHz and phase margin of 75 degrees with negligible load capacitance. The extra pole is responsible, therefore**,** for an extra 15 degrees of phase lag (the ideal integrator would provide a phase margin of 90 degrees). From section 2.3: -

$$|L(s)|=1 \;\Rightarrow\; \tau_E=\tau_B\frac{\tan(\theta_E)}{\cos(\theta_E)}$$

$$\tau_B=\frac{1}{2\pi f_B}=3.2\times10^{-8}s \;\Rightarrow\; \tau_E=\tau_B\frac{\tan(\theta_E)}{\cos(\theta_E)}\approx 8.9\times10^{-9}s$$

This corresponds to an extra pole at around 18MHz**. The LF356 is a "good" op-amp.**

A large capacitive load results in another pole and an increase in phase shift. The output resistance, *R* and load capacitance, *C*, combine to form another first order, low-pass filter and the high frequency model becomes: -

$$\omega\approx\omega_B \;\Rightarrow\; H(s)=\frac{1}{\tau_B s(1+\tau_E s)(1+\tau_L s)} \text{ with } \tau_L=RC$$

In practice the two poles combine as a single extra pole, as the second order term is usually negligible: -

$$|\tau_E s|<1 \text{ and } |\tau_L s|<1 \;\Rightarrow\; |\tau_L\tau_E s^2|<<1 \text{ and } (1+\tau_E s)(1+\tau_L s)\approx 1+(\tau_E+\tau_L)s$$

The phase shifts from each mechanism simply add and the single "extra pole" model is usually sufficient. The polar form, however, provides a convenient method of calculation without this approximation: -

$$H(s)=\frac{\exp(-j\pi/2)}{\tau_B\omega}\cos(\theta_E)\exp(-j\theta_E)\cos(\theta_L)\exp(-j\theta_L)$$

$$\text{with } \tan(\theta_E)=\tau_E\omega \text{ and } \tan(\theta_L)=\tau_L\omega$$

The total phase shift is: $\theta_T=-\frac{\pi}{2}-\theta_E-\theta_L$ The phase margin (degrees) is, therefore: $\theta_{PM}=90-(\theta_E+\theta_L)$

Consider, for example, the unity gain frequency: -

$$|H(s)|=\frac{1}{\tau_B\omega}\cos(\theta_E)\cos(\theta_L) \;\Rightarrow\; \omega_{UG}=\frac{1}{\tau_B}\cos(\theta_E)\cos(\theta_L)$$

The LF356 has an output resistance of 50Ω. What is the maximum load capacitance for a minimum phase margin of 45 degrees (15 + 30) with a feedback factor of 100%**?**

$$|H(s)|=1 \;\Rightarrow\; \omega_{UG}=\frac{1}{\tau_B}\cos(\theta_E)\cos(\theta_L)=\frac{0.966\times0.866}{3.2\times10^{-8}}=2.61\times10^{7}$$

$$\tau_E=\frac{\tan(15)}{\omega_{UG}}=1.03\times10^{-8}s \text{ and } \tau_L=\frac{\tan(30)}{\omega_{UG}}=2.21\times10^{-8}s$$

The maximum capacitance is, therefore: $C_{MAX}=\frac{\tau_L}{R}=\frac{2.21\times10^{-8}}{50}=442\,pF$

3.2 The integrator

With feedback an op-amp can be used to implement a practical integrator circuit which is an excellent building block as part of a multi-stage HGB.

The main advantage is that the gain falls below 0dB well before the frequency at which the op-amp runs out of steam (the extra pole) and the effect of load capacitance is much reduced. Compared to the op-amp itself the integrator characteristic extends to a much lower frequency and the time constant is more stable over temperature. As the final stage of an HGB overall stability is assured by design, even with a large capacitive load.

One other important characteristic is the effect of the virtual earth. The action of feedback is to maintain both inputs at approximately the same voltage (i.e. 0V). The op-amp common mode input is very small and error due to the op-amp common mode rejection ratio is negligible. The virtual earth can also act as a summing junction - it is possible to add two or more inputs with extra input resistors.

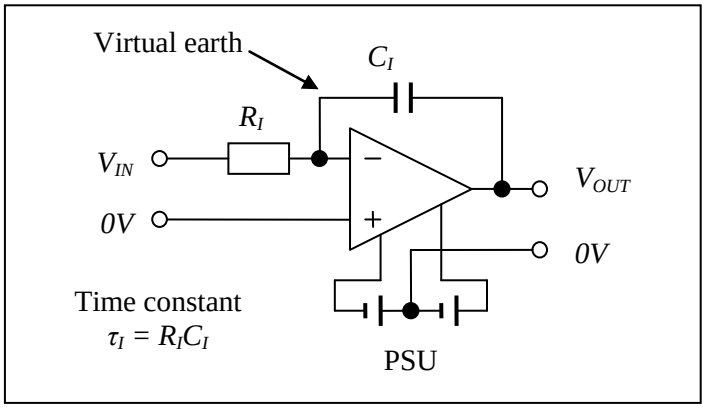


Fig.3.2.1 A simple integrator

With an integrator one expects the gain to continue increasing as frequency decreases limited only by the finite gain of the op-amp. This sets a lower limit to the operating frequency. For a low frequency analysis I shall assume, therefore, the high gain, first order, low-pass filter model for the op-amp (not including the extra pole): -

$$\omega << \omega_B \;\Rightarrow\; H(s) = \frac{G}{1+\tau_P s} \quad \text{with} \quad \tau_P = G\tau_B$$

Where G is the "DC gain" τ_P is the time constant of the dominant pole and $\omega_B = 1/\tau_B$ is the GBWP in rads^{-1}. From section 2.1 the "D" factor is, therefore: -

$$D(s) = \frac{H(s)}{H(s)+\beta(s)} = \frac{G}{G+\beta(s)(1+\tau_P s)}$$

The feedback elements are a capacitor and resistor with a time constant which is usually much larger than the op-amp time constant: -

$$Z_1 = R_I \quad \text{and} \quad Z_2 = \frac{1}{sC_I} \;\Rightarrow\; \beta(s) = 1 + \frac{1}{\tau_I s} \quad \text{with} \quad \tau_I = R_I C_I \quad \text{and} \quad \tau_I >> \tau_B$$

The "D" factor is, therefore:

$$D(s) = \left(1 + \frac{1}{G} + \frac{1}{G\tau_I s} + \tau_B s + \frac{\tau_B}{\tau_I}\right)^{-1}$$

The effects of the real components are usually irrelevant as they do not contribute to the phase error. One could, for example, define a parameter $\left(\delta = \frac{1}{G} + \frac{\tau_B}{\tau_I} << 1\right)$. The transfer function becomes: -

$$T(s) = \frac{-1}{\tau_I s} D(s) = \frac{-1}{\tau_I s}\left(1 + \delta + \frac{1}{G\tau_I s} + \tau_B s\right)^{-1}$$

Extract a factor $(1+\delta)$: $$T(s) = \frac{-1}{\tau'_I s}\left(1 + \frac{1}{G\tau'_I s} + \tau'_B s\right)^{-1}$$

Where: -

$\tau'_I = \tau_I(1+\delta) \approx \tau_I$: indicating a slightly increased time constant of the integrator, compared to the ideal. This could be easily mitigated with a corresponding reduction in the value of the resistor R_I or increase in C_I.

$\tau'_B = \frac{\tau_B}{1+\delta} \approx \tau_B$: indicating a slightly reduced op-amp time constant (a small increase in the GBWP).

Multiply top and bottom by $G\tau'_I s$ and a factor of $\tau'_I s$ cancels and, noting that $G\tau'_I >> \tau'_B$: -

$$T(s) = \frac{-G}{1 + G\tau'_I s + G\tau'_I \tau'_B s^2} \approx \frac{-G}{(1+G\tau'_I s)(1+\tau'_B s)}$$

The result is, as expected, a finite gain at very low frequency (the same as the op-amp) with a second order low-pass characteristic with two widely separated poles. The "*D*" factor has a broad, flat, band-pass characteristic with, to a very good approximation: -

$$D(s) \approx \frac{G\tau'_I s}{(1+G\tau'_I s)(1+\tau'_B s)}$$

As frequency increases (from about 1Hz to 1kHz) and well below the GBWP the term $\tau'_B s$ is negligible and the circuit looks increasingly like an integrator followed by a first order high-pass filter characteristic with a positive phase error: -

$$|\tau'_B s| << 1 \;\Rightarrow\; T(s) \approx \frac{-1}{\tau'_I s}\left(\frac{G\tau'_I s}{1+G\tau'_I s}\right) \approx \frac{-1}{\tau'_I s}\left(1 - \frac{1}{G\tau'_I s}\right)$$

At some critical frequency the phase error is zero: -

$$\frac{1}{G\tau'_I s} + \tau'_B s = 0 \quad \Rightarrow \quad \omega_C = \frac{1}{\sqrt{G\tau_I \tau_B}}$$

Note that the effect of the correction factors $(1+\delta)$ cancel.

Typical values for an integrator as part of an HGB are: -

$$G \approx 10^5 \quad \tau'_I \approx 10^{-5} s \text{ (16kHz)} \quad \text{and} \quad \tau'_B \approx 3.2 \times 10^{-8} s \text{ (5MHz)}$$

The critical frequency is, in more convenient form: $f_C = \sqrt{\frac{f_I f_B}{G}} \approx 900 Hz$

At higher frequency the term $\frac{1}{G\tau'_I s}$ become negligible and the transfer function becomes an integrator followed by a first order low-pass characteristic with negative phase error: -

$$|G\tau'_I s| >> 1 \;\Rightarrow\; T(s) \approx \frac{-1}{\tau'_I s}\left(\frac{1}{1+\tau'_B s}\right) \approx \frac{-1}{\tau'_I s}(1-\tau'_B s)$$

In more convenient form, with frequency in Hz, the phase error is sufficiently small (<15 degrees is tolerable as part of an HGB) over a wide range of frequency (typically < 1Hz to 1MHz): -

$$\frac{f_I}{G} << f << f_B \;\Rightarrow\; T(s) \approx \frac{-1}{\tau_I s}\left(1+j\left(\frac{f_I}{Gf}-\frac{f}{f_B}\right)\right)$$

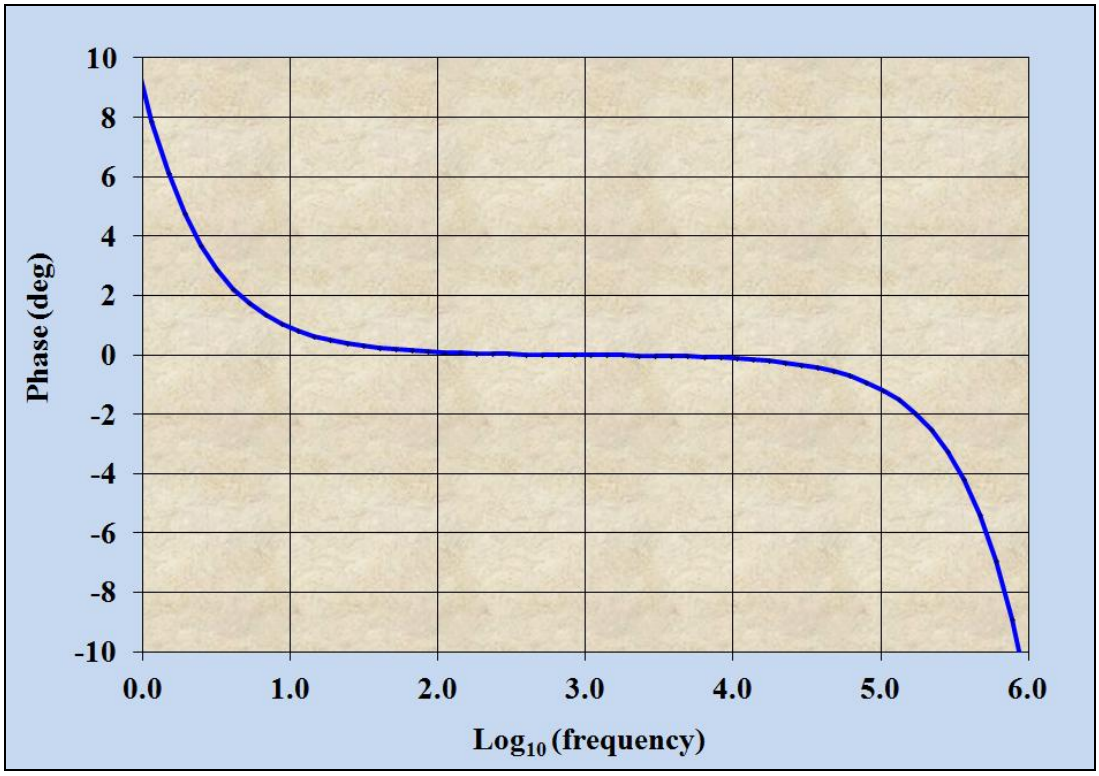


Fig. 3.2.2 Phase error versus frequency (Hz)

One can conclude a rule of thumb: -

With a good op-amp and stable feedback components, a reasonably good model for this circuit (as part of an HGB type 1) is an ideal integrator:-

$$\frac{\omega_I}{G} << \omega << \omega_B \;\Rightarrow\; T(s) \approx \frac{-1}{\tau_I s}$$

If, on the other hand, one requires an accurate integrator, with low phase error (e.g. as part of a high accuracy oscillator or filter circuit) then one should employ a two-stage HGB (type 2) rather than a single op-amp [1 and 2].

1. Monograph: “High accuracy amplifiers, integrators and differentiators”. See section 4.
2. Monograph: “A circuit for measuring tanδ”.

The best way to assess stability is an examination of the loop gain. At high frequency the best model of the op-amp is an integrator with an extra pole (see section 3.1.1): -

$$\omega \approx \omega_B \quad \Rightarrow \quad H(s) = \frac{1}{\tau_B s}\left(\frac{1}{1+\tau_E s}\right)$$

The feedback factor is also a function of frequency: -

$$Z_1 = R_I \text{ and } Z_2 = \frac{1}{sC_I} \quad \Rightarrow \quad F(s) = \frac{Z_1}{Z_1 + Z_2} = \frac{\tau_I s}{1+\tau_I s} \text{ with } \tau_I = R_I C_I \text{ and } \tau_I >> \tau_B$$

The loop gain is, therefore: -

$$L(s) = H(s)F(s) = \frac{1}{\tau_B s}\left(\frac{1}{1+\tau_E s}\right)\left(\frac{\tau_I s}{1+\tau_I s}\right)$$

The phase shift of the feedback network is positive (phase advance) and the total phase shift around the loop is reduced (e.g. compared to a voltage follower). One can employ the polar form for which no approximation is necessary: -

$$\frac{\tau_I s}{1+\tau_I s} = \left(1+\frac{1}{\tau_I s}\right)^{-1} = \left(1 - j\frac{1}{\tau_I \omega}\right)^{-1} = \cos(\theta_I)\exp(j\theta_I) \text{ with } \tan(\theta_I) = \frac{1}{\tau_I \omega}$$

The loop gain, in polar form, is, therefore: -

$$L(s) == \frac{\exp(-j\pi/2)}{\tau_B \omega}\cos(\theta_E)\exp(-j\theta_E)\cos(\theta_I)\exp(j\theta_I)$$

The magnitude is: $$|L(s)| = \frac{\cos(\theta_E)\cos(\theta_I)}{\tau_B \omega} = 1 \quad \Rightarrow \quad \omega_{UG} = \frac{\cos(\theta_E)\cos(\theta_I)}{\tau_B}$$

The total phase shift is: $\theta_T = -\frac{\pi}{2} - \theta_E + \theta_I$ The phase margin (degrees) is: $\theta_{PM} = 90 - \theta_E + \theta_I$

In practice the phase shift due to the feedback network is small and the extra pole is the more significant (typically a maximum of 45 degrees with a capacitive load) and the unity gain frequency is not much lower than the GBWP.

$$\theta_I \approx 0 \text{ and } \theta_E < 45 \quad \Rightarrow \quad \omega_{UG} > \frac{1}{\sqrt{2}\tau_B} = 0.707\omega_B$$

The unity gain frequency is usually much higher than the integrator characteristic frequency so that the phase shift due to the feedback network is small and the phase margin is about the same as a voltage follower: -

$$\tan(\theta_I) = \frac{1}{\tau_I \omega_{UG}} << 1 \quad \Rightarrow \quad \theta_I \approx 0 \quad \Rightarrow \quad \theta_{PM} \approx 90 - \theta_E$$

The result is a resonant peak at high frequency. The closed loop gain is well below 0dB at this frequency so that any resonant peak is unlikely to cause a problem. This is not the case with a one-plus-integrator circuit as the next section demonstrates.

It is recommended, therefore, that the integrator is the final stage of a multi-stage type 1 high gain block, especially with a capacitive load.

3.3 The one-plus-integrator (OPI)

A one-plus-integrator is the same circuit as an integrator with the input connections reversed. The feedback factor is exactly the same as for the integrator as is the low frequency phase error analysis.

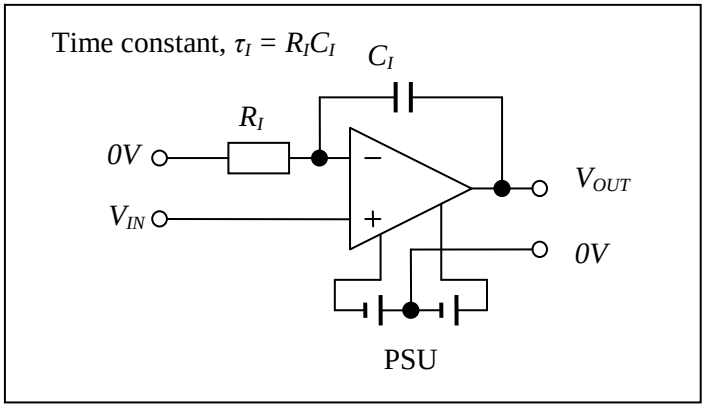


Fig. 3.3.1 A one-plus integrator circuit (OPI)

From the general analysis, section 2.1, the closed loop transfer function is: -

$$Z_1 = R_I \quad \text{and} \quad Z_2 = \frac{1}{sC_I} \quad \text{and} \quad \tau_I = R_I C_I \quad \Rightarrow \quad \frac{V_{OUT}}{V_{IN}} = \left(1 + \frac{Z_2}{Z_1}\right) D(s) = \left(1 + \frac{1}{\tau_I s}\right) D(s)$$

The reciprocal feedback factor and "*D*" factor are the same as for the integrator (section 3.2) and so the transfer function is, from DC to a frequency approaching the GBWP of the op-amp: -

$$|\tau_B s| << 1 \;\Rightarrow\; T(s) = \left(1 + \frac{1}{\tau_I s}\right)\left(1 + \frac{1}{G} + \frac{1}{G\tau_I s} + \tau_B s + \frac{\tau_B}{\tau_I}\right)^{-1}$$

I shall again extract a real factor $(1+\delta)$ resulting in a small overall drop in gain and a band-pass correction factor with two widely separated poles: -

$$T(s) = \left(\frac{1}{1+\delta}\right)\left(1 + \frac{1}{\tau_I s}\right)\left(\frac{G\tau_I' s}{1 + G\tau_I' s + G\tau_I'\tau_B' s^2}\right) \approx \left(\frac{1}{1+\delta}\right)\left(1 + \frac{1}{\tau_I s}\right)\left(\frac{G\tau_I' s}{(1 + G\tau_I' s)(1 + \tau_B' s)}\right)$$

At low frequency the OPI looks like a first order low-pass filter with the same DC gain as the op-amp: -

$$|\tau_I s| << 1 \quad \Rightarrow \quad |\tau_B s| << 1 \;\Rightarrow \quad T(s) \approx \left(\frac{1}{1+\delta}\right)\left(\frac{G}{1 + G\tau_I' s}\right)$$

As frequency increases the gain falls, first order, and the circuit behaves like an integrator: -

$$\frac{1}{G} << |\tau_I s| << 1 \;\Rightarrow\; T(s) \approx \left(\frac{1}{1+\delta}\right)\frac{1}{\tau_I' s}$$

Over a wide range of frequency, the magnitude of the band-pass characteristic is very nearly one and the result is a reasonably accurate one-plus-integrator (see fig. 3.2.2 for typical phase error): -

$$\frac{1}{G} << |\tau_I s| << |\tau_B s| \quad \Rightarrow \quad \frac{G\tau_I' s}{1 + G\tau_I' s + G\tau_I'\tau_B' s^2} \approx 1 \;\Rightarrow\; T(s) \approx \left(\frac{1}{1+\delta}\right)\left(1 + \frac{1}{\tau_I s}\right)$$

At high frequency the feedback factor is very nearly 100% and the closed loop gain levels out to very nearly one. To investigate the phase margin and any resonant peak, therefore, one needs the high frequency model of the op-amp: an integrator with an extra pole: -

$$\omega \approx \omega_B \quad \Rightarrow \quad H(s) = \frac{1}{\tau_B s}\left(\frac{1}{1+\tau_E s}\right)$$

The analysis is exactly the same as for the integrator circuit.

$$L(s) = \frac{\exp(-j\pi/2)}{\tau_B \omega}\cos(\theta_E)\exp(-j\theta_E)\cos(\theta_I)\exp(j\theta_I)$$

$$\text{with} \quad \tan(\theta_E) = \tau_E \omega \quad \text{and} \quad \tan(\theta_I) = \frac{1}{\tau_I \omega}$$

The magnitude and unity gain frequency are: $|L(s)| = \dfrac{\cos(\theta_E)\cos(\theta_I)}{\tau_B \omega} \quad \Rightarrow \quad \omega_{UG} = \dfrac{\cos(\theta_E)\cos(\theta_I)}{\tau_B}$

The total phase shift is: $\theta_T = -\dfrac{\pi}{2} - \theta_E + \theta_L$ The phase margin (degrees) is: $\theta_{PM} = 90 - \theta_E + \theta_L$

The consequence is the possibility of a resonant peak above 0dB which could be a problem. It is recommended, therefore, that an OPI circuit is not used to drive a capacitive load – the last stage of a multi-stage HGB (type 1) should be the integrator.

Calculating the unity gain frequency and phase margin may take a few iterations: -

a). First estimate the unity gain frequency: $\omega_{UG} \approx 1/\tau_B$

b). Calculate the phase shifts based on this estimate: $\tan(\theta_E) = \tau_E \omega$ and $\tan(\theta_I) = 1/\tau_I \omega$

c). Recalculate the unity gain frequency and repeat: $\omega_{UG} = \dfrac{\cos(\theta_E)\cos(\theta_I)}{\tau_B}$

The solution should converge after two or three iterations.

Rule of thumb: -

One can conclude that, as part of an HGB, with a low capacitance load, a reasonably good model of the circuit, over a wide range of frequency (typically < 1Hz to 1MHz) is a one-plus-integrator: -

$$\frac{\omega_I}{G} << \omega << \omega_B \quad \Rightarrow \quad T(s) \approx 1 + \frac{1}{\tau_I s}$$

3.4 The non-inverting amplifier and voltage follower

Both feedback components are resistors which simplifies the analysis. A major source of error, however, is the finite common mode rejection ratio of the op-amp. This circuit, therefore, has limited high accuracy applications. For a voltage follower $R_2 = 0$ or $R_1 = \infty$ (or both).

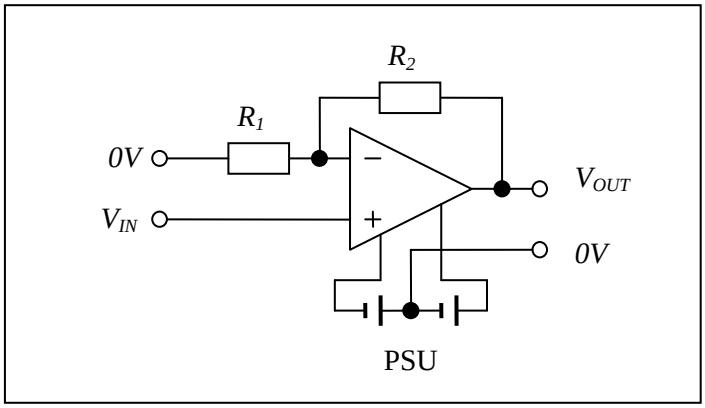


Fig. 3.4.1 A non-inverting amplifier circuit

From the circuit analysis, section 2.1 the closed loop transfer function is: -

$$Z_1 = R_1 \quad \text{and} \quad Z_2 = R_2 \;\Rightarrow\; \frac{V_{OUT}}{V_{IN}} = \left(1 + \frac{R_2}{R_1}\right) D(s) \quad \text{with} \quad \beta = 1 + \frac{R_2}{R_1}$$

I shall assume the high gain, low-pass filter model for the op-amp for low frequency (below the unity gain frequency). The reciprocal feedback factor is now real and constant so that the "*D*" factor is: -

$$\omega << \omega_B \;\Rightarrow\; H(s) = \frac{G}{1 + \tau_P s} \;\Rightarrow\; D(s) = \frac{H(s)}{H(s) + \beta} = \left(1 + \frac{\beta}{G} + \beta\tau_B s\right)^{-1}$$

The transfer function is, therefore: $$T(s) = \left(1 + \frac{R_2}{R_1}\right)\left(1 + \frac{\beta}{G} + \beta\tau_B s\right)^{-1}$$

One can extract a factor $(1+\delta)$: -

$$\delta = \frac{\beta}{G} << 1 \;\Rightarrow\; T(s) = \left(1 + \frac{R_2}{R_1}\right)\left(\frac{1}{1+\delta}\right)\left(\frac{1}{1 + \beta\tau'_B s}\right) \quad \text{with} \quad \tau'_B = \frac{\tau_B}{1+\delta}$$

The result is a slightly reduced gain with a first order low-pass characteristic. The bandwidth is approximately inversely proportional to gain (hence the gain-bandwidth product is approximately constant). At low frequency, to a very good approximation: -

$$|\beta\tau_B s| << 1 \;\Rightarrow\; T(s) \approx \left(1 + \frac{R_2}{R_1}\right)\left(\frac{1}{1+\delta}\right)\left(1 - \beta\tau'_B s - (\beta\tau'_B s)^2\right)$$

In more convenient form with frequency in Hz: -

$$f << f_B \;\Rightarrow\; T(s) \approx \left(1 + \frac{R_2}{R_1}\right)\left(\frac{1}{1+\delta}\right)\left(1 - j\beta\frac{f}{f_B} + \left(\beta\frac{f}{f_B}\right)^2\right)$$

The phase error is first order with respect to frequency and too large for some amplifier/follower applications. Also, this does not include errors due to the finite CMRR.

With a low gain ($R_2 < R_1$ and a feedback factor approaching 100%) the extra pole results in a low phase margin and a resonant peak at high frequency. The high frequency model of the op-amp is required: an integrator with an extra pole: -

$$\omega \approx \omega_B \quad \Rightarrow \quad H(s) = \frac{1}{\tau_B s}\left(\frac{1}{1+\tau_E s}\right)$$

From section 2.1 the closed loop response is: -

$$T(s) = -\frac{R_2}{R_1} D(s) \quad \text{with} \quad D(s) = \frac{H(s)}{H(s)+\beta} = \frac{1}{1+\beta\tau_B s(1+\tau_E s)} \quad \text{and} \quad \beta = 1+\frac{R_2}{R_1}$$

The result is a second order response: $$D(s) = \frac{1}{1+\beta\tau_B s + \beta\tau_B\tau_E s^2}$$

The transfer function is, in standard normalised form $(s = j\omega/\omega_N)$: -

$$T(s) = \frac{1}{1+2\xi s + s^2} \quad \text{with natural frequency} \quad \omega_N = \frac{1}{\sqrt{\beta\tau_B\tau_E}} \quad \text{and damping ratio} \quad \xi = \frac{1}{2}\sqrt{\frac{\beta\tau_B}{\tau_E}}$$

To achieve the minimum phase margin of 45 degrees the unity gain frequency must be the same as the extra pole. The total phase shift is 135 degrees (90 degrees from the integrator and 45 degrees from the extra pole). For a typical bode plot, with 45 degrees phase margin, see fig. 2.4.1. With a loop gain magnitude of one the result is: -

$$\omega_{UG} = \frac{1}{\tau_E} \quad \Rightarrow \quad |L(s)| = \frac{|H(s)|}{\beta} = \left|\frac{1}{\beta\tau_B s}\left(\frac{1}{1+\tau_E s}\right)\right| = \left|\frac{\tau_E}{j\beta\tau_B}\left(\frac{1}{1+j}\right)\right| = 1 \quad \Rightarrow \frac{\tau_E}{\beta\tau_B} = \sqrt{2}$$

$$\Rightarrow \quad \xi = \frac{1}{2}\sqrt{\frac{\beta\tau_B}{\tau_E}} = \frac{1}{2}\sqrt{\frac{1}{\sqrt{2}}} = 0.42$$

Worst case is the voltage follower (100% feedback: $\beta = 1$). If one assumes that the extra pole is entirely due to the output resistance and capacitive load the capacitance to produce this peak can be estimated. The LF356, for example, has output resistance of around 50Ω and a GBWP of 5MHz: -

$$\beta = 1 \quad \text{and} \quad \tau_B = \frac{1}{2\pi f_B} = 3.2\times10^{-8} s \quad \Rightarrow \quad \tau_E = \sqrt{2}\tau_B = 4.5\times10^{-8} s$$

$$\text{and} \quad \tau_E = RC \quad \Rightarrow C_{MAX} = \frac{\tau_E}{R} \approx \frac{4.5\times10^{-8}}{50} \approx 900\,pF$$

In practice the op-amp internal extra pole contributes about 15 degrees of extra phase lag so that the maximum contribution from the load capacitance is 30 degrees. The maximum capacitance is about half this estimate. See section 3.1.1 for the more precise calculation. For improved accuracy and stability see section 3.6.

As a rule of thumb a voltage follower based on an internally compensated single op-amp does not have sufficient accuracy or stability for many high accuracy applications. Fortunately a two or three-stage HGB with 100% feedback and floating power supply provide a simple and cost-effective solution [1].

To eliminate CMRR errors employ the inverting mode (see section 3.5). To reduce phase error employ a two-stage HGB [2].

1. Monograph: "High accuracy voltage followers"
2. Monograph: "High accuracy amplifiers, integrators and differentiators".

3.5 The Inverting amplifier

With an inverting amplifier both op-amp inputs are maintained at local 0V and common mode is not an issue. With two or more input resistors the circuit can also perform addition.

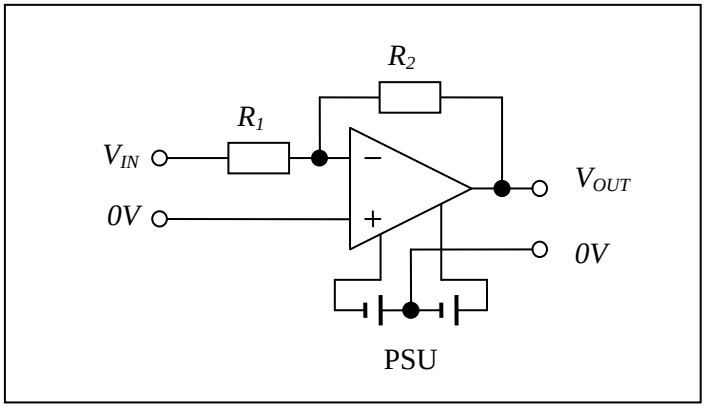


Fig. 3.5.1 The inverting amplifier circuit

From the circuit analysis, section 2.1 the closed loop transfer function is: -

$$Z_1 = R_1 \quad \text{and} \quad Z_2 = R_2 \;\Rightarrow\; \frac{V_{OUT}}{V_{IN}} = -\frac{R_2}{R_1} D(s)$$

The reciprocal feedback factor and "*D*" factor are the same as the non-inverting amplifier so that, at low frequency, the transfer function is: -

$$\omega << \omega_B \;\Rightarrow\; H(s) = \frac{G}{1+\tau_P s} \;\Rightarrow\; T(s) = -\frac{R_2}{R_1}\left(1+\frac{\beta}{G}+\tau_B s\right)^{-1}$$

I shall again extract a factor $(1+\delta)$: -

$$\delta = \frac{\beta}{G} << 1 \;\Rightarrow\; T(s) = -\frac{R_2}{R_1'}\left(\frac{1}{1+\beta\tau_B' s}\right) \text{ with } R_1' = (1+\delta)R_1 \text{ and } \tau_B' = \frac{\tau_B}{1+\delta}$$

The result is an inverting amplifier with a slightly reduced gain and a first order low-pass characteristic. At low frequency, to a very good approximation: -

$$|\beta\tau_B s| << 1 \;\Rightarrow\; T(s) \approx -\frac{R_2}{R_1'}\left(1-\beta\tau_B s-(\beta\tau_B s)^2\right)$$

In more convenient form with frequency in Hz: -

$$f << f_B \;\Rightarrow\; T(s) \approx -\frac{R_2}{R_1}\left(\frac{1}{1+\delta}\right)\left(1- j\beta\frac{f}{f_B}+\left(\beta\frac{f}{f_B}\right)^2\right)$$

Once again the main problem is the phase error (quadrature). The in-phase error is likely to be negligible compared to the resistor tolerance and drift. For lower phase error employ a two-stage HGB (usually type 2) [1 and 2].

The high frequency analysis (resonant peak due to the extra pole) is the same as for the non-inverting mode but unlikely to be an issue. The minimum feedback factor is typically $0.5\,(\beta \geq 2)$ for an inverter (gain = -1).

1. Monograph: "High accuracy amplifiers, integrators and differentiators". See section 3.
2. Monograph: "A circuit for measuring tanδ".

3.6 A voltage follower with external compensation

Voltage followers, based on a single (internally compensated) op-amp, do not usually provide sufficient accuracy for many applications (see section 3.4). A de-compensated op-amp (reduced internal compensation) can, however, with a suitable feedback network, boost the closed loop accuracy and stability. When combined with an advanced OPI stage, for example, the result is a useful two-stage high accuracy voltage follower (HAVF) [1].

The basic idea is to employ the higher gain to boost accuracy at low frequency but reduce the feedback factor, at high frequency, to maintain stability. This is often referred to as "external compensation": -

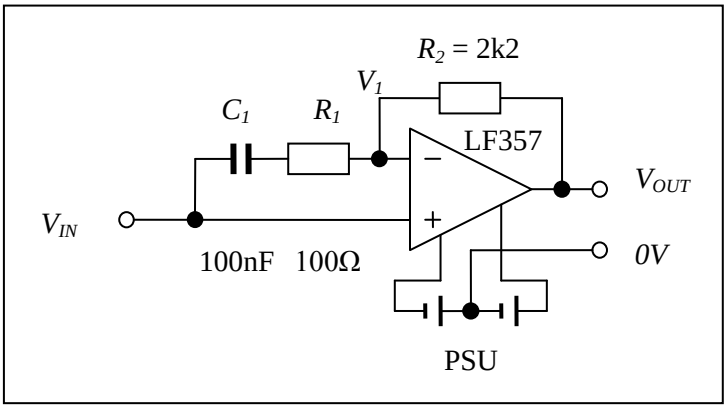


Fig. 3.6.1 A voltage follower with a de-compensated op-amp

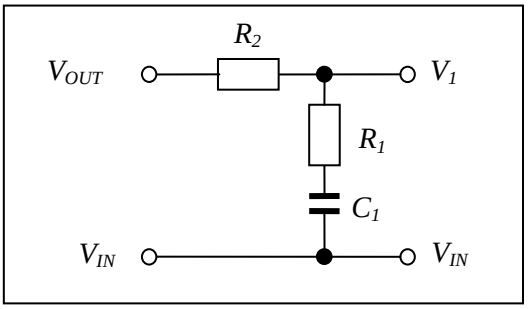


Fig. 3.6.2 The feedback network

The best way to analyse the circuit is to think of the feedback network as a simple RC filter (with V_{IN} = 0) so that it can be defined in the normal way: -

$$V_{IN}=0 \;\Rightarrow\; \frac{V_1}{V_{OUT}}=F(s)=\frac{1+R_1C_1s}{1+(R_1+R_2)C_1s}=\frac{1+\tau_1 s}{1+\tau_2 s} \quad \text{with} \quad \tau_1=R_1C_1, \quad \tau_2=(R_1+R_2)C_1 \quad \text{and} \quad \tau_2 >> \tau_1$$

More generally, therefore, based on voltage differences: $V_1-V_{IN}=F(s)(V_{OUT}-V_{IN})$

Combine the definitions of *F(s)* and *H(s)* to eliminate V_1: -

$$V_{OUT}=H(s)(V_{IN}-V_1) \;\Rightarrow\; V_{OUT}=H(s)(V_{IN}-F(s)(V_{OUT}-V_{IN})-V_{IN})$$

With a little algebra the closed loop transfer function is: $T(s)=\frac{V_{OUT}}{V_{IN}}=\frac{H(s)F(s)}{H(s)F(s)+1}$

The loop gain is, therefore (from section 2.1): $T(s)=\frac{L(s)}{L(s)+1} \;\Rightarrow\; L(s)=H(s)F(s)$

1. Monograph: "High accuracy voltage followers". See section 6.3

A very good model of the op-amp, from DC to the unity gain frequency, is a high gain, low-pass filter with an extra pole: -

$$0 < \omega < \omega_B \;\Rightarrow\; H(s) = \frac{V_{OUT}}{V_{IN} - V_1} = \frac{G}{1+\tau_P s}\left(\frac{1}{1+\tau_E s}\right) \quad \text{with} \quad \tau_P = G\tau_B = \frac{G}{\omega_B}$$

The closed loop transfer function is, therefore: -

$$T(s) = \frac{G(1+\tau_1 s)}{G(1+\tau_1 s) + (1+\tau_P s)(1+\tau_2 s)(1+\tau_E s)}$$

This looks a little daunting but the time constants (and relevant frequencies) are usually widely separated: -

$$\tau_P > \tau_2 >> \tau_1 >> \tau_B \approx \tau_E$$

With reduced internal compensation the maximum gain (minimum typically 10^5) and the frequency of the extra pole remain the same but the time constant of the dominant pole is reduced. With the LF357, for example, the GBWP is increased to 20MHz (compared to 5MHz for the LF356) which is also about the same as the extra pole (see section 3.1.1). The relevant parameters are: -

$$G \approx 10^5 \quad \text{and} \quad \tau_E \approx \tau_B \approx 8\times10^{-9}s \;\Rightarrow\; \tau_P = G\tau_B = \frac{G}{\omega_B} \approx 8\times10^{-4}s \quad \text{and} \quad f_P = \frac{1}{2\pi\tau_P} \approx 200Hz$$

For the recommended component values in fig. 3.6.1, for example, the external compensation kicks in at a frequency determined by the time constant τ_2.

Typical values are: $\tau_2 = (R_1 + R_2)C_1 \approx 2.3\times10^{-4}s$ and $f_2 = \frac{1}{2\pi(R_1+R_2)C_1} \approx 700Hz$

Similarly the time constant of the zero is $\tau_1 = R_1C_1 \approx 10^{-5}s$ and $f_1 = \frac{1}{2\pi R_1 C_1} \approx 16kHz$

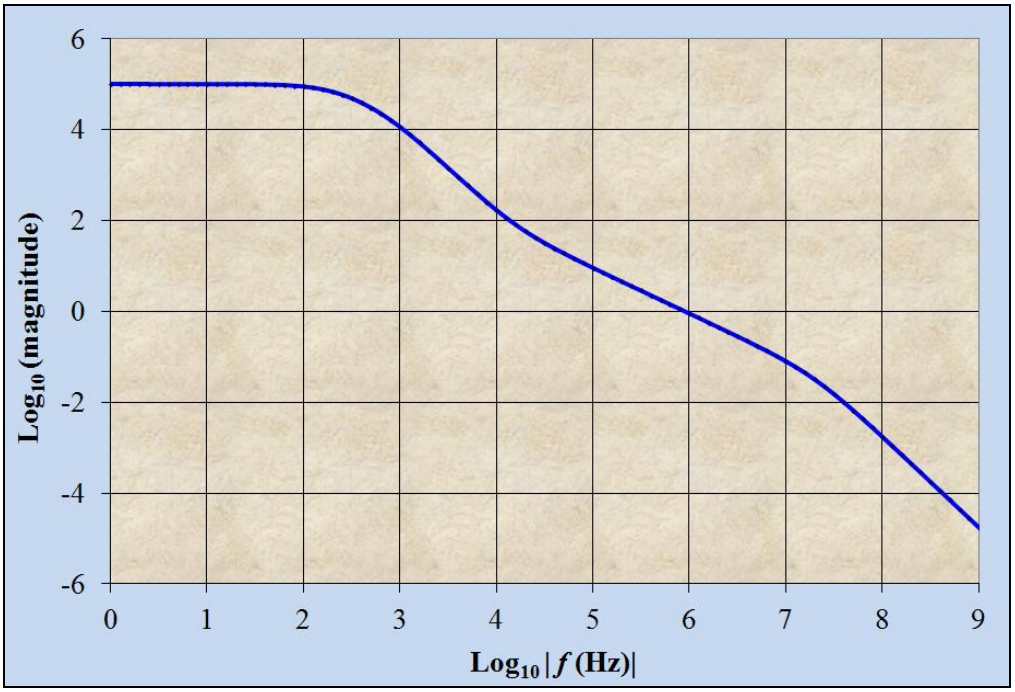


Fig. 3.6.3 Loop gain with typical component values (see fig. 3.6.1)

The resulting loop gain is flat at low frequency (typically DC – 200Hz). As frequency increases the op-amp dominant pole and the feedback pole cause the gain to drop with a second order slope. The zero at 16kHz causes the slope to revert to first order (causing the phase lag to reduce to nearly 90 degrees) as the loop gain magnitude falls below 0dB. This occurs well below the frequency of the extra pole resulting in a phase margin approaching 90 degrees. The unity gain frequency (of the loop gain) is sufficiently low that this circuit can easily drive substantial load capacitance (≈ 1nF) with no stability problems.

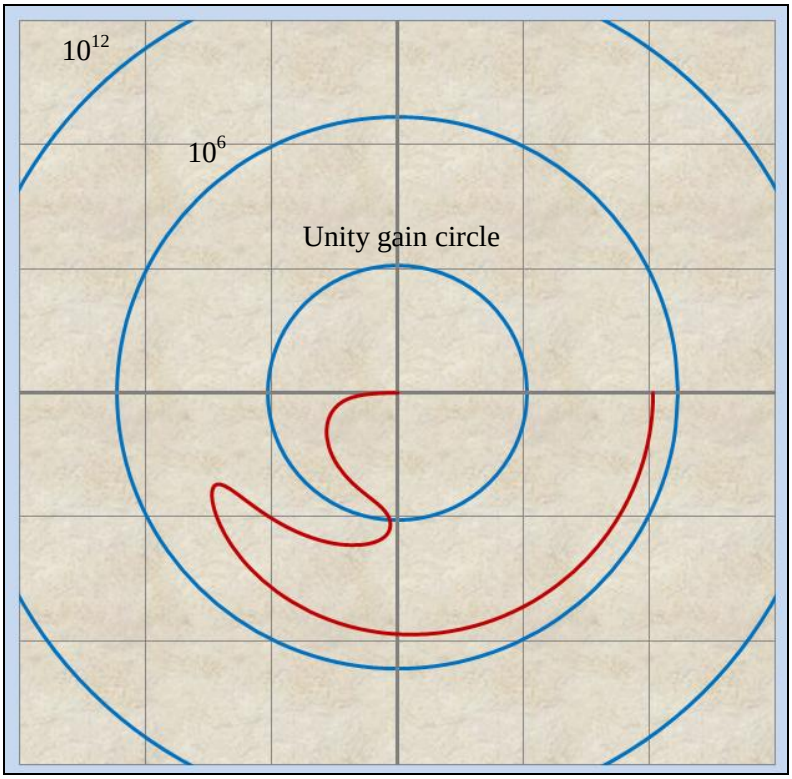


Fig. 3.6.4 The BoDayQuist plot of loop gain with typical component values

The detailed analysis is not as complicated as it looks. At low frequency the R_1C_1 series snubber circuit has high impedance and the resistor R_2 provides 100% feedback. In practice a moderately low value for R_2 is required due to the op-amp input capacitance. The closed loop response is the same as a voltage follower with the LF357 gain-bandwidth product $\left(f_B \approx 20MHz\right)$. Repeated from above, for convenience: -

$$F(s)=\frac{1+\tau_1 s}{1+\tau_2 s} \text{ and } H(s)=\frac{G}{1+\tau_P s}\left(\frac{1}{1+\tau_E s}\right) \text{ and } T(s)=\frac{H(s)F(s)}{H(s)F(s)+1}$$

$$|\tau_E s|<<|\tau_1 s|<<|\tau_2 s|<<1 \;\Rightarrow\; F(s)\approx 1 \text{ and } T(s)\approx\frac{H(s)}{H(s)+1}=\frac{G}{G+1+\tau_P s}$$

The result is lower in-phase and quadrature error, by a factor of four, compared to the fully compensated LF356$\left(f_B \approx 5MHz\right)$: -

$$|\tau_E s|<<|\tau_1 s|<<|\tau_2 s|<<1 \;\Rightarrow\; T(s)\approx\left(1+\frac{1}{G}+\tau_B s\right)^{-1}\approx 1-\frac{1}{G}-\tau_B s$$

In more convenient form, with frequency in Hz: -

$$f<<f_2 \;\Rightarrow\; T(s)\approx 1-\frac{1}{G}-j\frac{f}{f_B}$$

With a minimum DC gain of 10^5 the in-phase error is less than 10ppm and, in many applications, the effect of quadrature error is negligible with a reasonably good phase sensitive null detector.

Even if one takes into account the rapid fall in gain due to the second pole (≈700Hz) the next significant error term is small. Repeated from above: -

$$T(s)=\frac{G(1+\tau_1 s)}{G(1+\tau_1 s)+(1+\tau_P s)(1+\tau_2 s)(1+\tau_E s)}$$

$$|\tau_E s|<<|\tau_1 s|<<1 \;\Rightarrow\; T(s)\approx\frac{G}{G+(1+\tau_P s)(1+\tau_2 s)}$$

$$\Rightarrow\; T(s)\approx\left(1+\frac{1}{G}+\frac{(\tau_P+\tau_2)}{G}s+\frac{\tau_P\tau_2}{G}s^2\right)^{-1}\approx 1-\frac{1}{G}-\tau_B s-\tau_B\tau_2 s^2$$

In more convenient form, with frequency in Hz: -

$$f<<f_1 \;\Rightarrow\; T(s)\approx 1-\frac{1}{G}-j\frac{f}{f_B}+\frac{f^2}{f_B f_2}$$

The main application is designed to operate at a frequency of 400Hz [1]. The resulting extra in-phase error is around +11ppm, though one should not rely on this to cancel, with any great accuracy, the -10ppm error due to the finite open loop gain.

At high frequency (above f_1: the zero of the feedback network) and approaching the extra pole: -

$$\omega\approx\omega_B \;\Rightarrow\; |\tau_P s|>>|\tau_2 s|>>|\tau_1 s|>>1 \;\Rightarrow\; T(s)\approx\frac{G\tau_1 s}{G\tau_1 s+\tau_P s(1+\tau_E s)\tau_2 s}$$

With a little algebra: $$T(s)\approx\frac{1}{1+\tau'_B s(1+\tau_E s)} \quad \text{with} \quad \tau'_B=\frac{\tau_P}{G}\frac{\tau_2}{\tau_1}=\tau_B\frac{\tau_2}{\tau_1}>>\tau_B$$

This is precisely the same second order, low-pass response of the basic voltage follower (see section 3.4) with a GBWP reduced by the factor $\tau_2/\tau_1>>1$. With the values recommended this is typically: -

$$\frac{\tau_2}{\tau_1}=\frac{R_1+R_2}{R_1}\approx 23$$

The LF357 GBWP of 20MHz is reduced to below 1MHz (≈ 870kHz) and well below the extra pole, even with significant load capacitance. The stability margin is an almost perfect 90 degrees (see fig. 3.6.4).

The stability margin is best analysed in the polar form of the loop gain. The unity gain frequency is much higher than the dominant pole and a sufficiently accurate model for the op-amp is an integrator with an extra pole. The contribution to phase shift from the feedback may be significant, however, and the loop gain model is: -

$$|\tau_P s|>>1 \;\Rightarrow\; L(s)=\frac{1}{\tau_B s}\left(\frac{1}{1+\tau_E s}\right)\left(\frac{1+\tau_1 s}{1+\tau_2 s}\right)$$

1. Monograph: “High accuracy voltage followers”. See section 6.3

At high frequency the phase shift of the feedback network is expected to be almost 90 degrees phase lag due to the pole and approaching 90 degree phase advance due to the zero. The network is best described, therefore, with (small) corresponding angles $(\phi = 90 - \theta)$ etc: -

$$|\tau_2 s| >> |\tau_1 s| >> 1 \;\Rightarrow\; 1 + \tau_1 s = 1 + j\tan(\theta_1) \;\text{ with }\; \theta_1 \approx \frac{\pi}{2} \;\text{ etc.}$$

In which case do the following: -

$$F(s) = \left(\frac{1+\tau_1 s}{1+\tau_2 s}\right) = \frac{\tau_1 s}{\tau_2 s}\left(\frac{1+1/\tau_1 s}{1+1/\tau_2 s}\right) = \frac{\tau_1}{\tau_2}\left(\frac{1 - j\tan(\phi_1)}{1 - j\tan(\phi_2)}\right) = \frac{\tau_1}{\tau_2}\exp\left(j(\phi_2 - \phi_1)\right)\frac{\cos(\phi_2)}{\cos(\phi_1)}$$

$$\text{with }\; \tan(\phi_1) = \frac{1}{\tau_1 \omega} \;\text{ and }\; \tan(\phi_2) = \frac{1}{\tau_2 \omega} \;\text{ and }\; \tau_2 > \tau_1 \;\Rightarrow\; \phi_2 < \phi_1 \text{ and } \phi_1 = \frac{\pi}{2} - \theta_1 \approx 0 \;\text{ etc.}$$

The net phase shift is always negative (phase lag). The loop gain is: -

$$L(s) = \frac{1}{\tau_B s}\left(\frac{1}{1+\tau_E s}\right)F(s) = \frac{\tau_1}{\tau_2 \tau_B \omega}\exp\left(j\left(-\frac{\pi}{2} - \theta_E + \phi_2 - \phi_1\right)\right)\frac{\cos(\theta_E)\cos(\phi_2)}{\cos(\phi_1)} \;\text{ with }\; \tan(\theta_E) = \tau_E \omega$$

The total phase shift is: $\theta_T = -\frac{\pi}{2} - \theta_E + \phi_2 - \phi_1$ The phase margin (degrees) is: $\theta_{PM} = 90 - (\theta_E + \phi_1 - \phi_2)$

$$|L(s)| = \frac{\tau_1}{\tau_2 \tau_B \omega}\frac{\cos(\theta_E)\cos(\phi_2)}{\cos(\phi_1)}$$

All the angles are small so that a reasonably good estimate for the unity gain frequency is: -

$$|L(s)| = 1 \;\Rightarrow\; \omega_{UG} \approx \frac{\tau_1}{\tau_2 \tau_B} \approx 5.43 \times 10^6 \quad \text{(about 865kHz and consistent with the previous estimate)}$$

From which a reasonably good estimate of the angles are, in degrees: -

$$\tau_E \approx 8 \times 10^{-9} s \;\Rightarrow\; \theta_E \approx 2.5 \;\Rightarrow\; \cos(\theta_E) \approx 0.999$$

$$\tau_2 \approx 2.3 \times 10^{-4} s \;\Rightarrow\; \phi_2 \approx 0.0 \;\Rightarrow\; \cos(\theta_2) \approx 1.000$$

$$\tau_1 \approx 10^{-5} s \Rightarrow\; \phi_1 \approx 1.0 \;\Rightarrow\; \cos(\theta_E) \approx 1.000$$

The total phase margin is, therefore, a generous: $\theta_{PM} = 90 - (3.5) \approx 86.5$ degrees

This example confirms the assumption that the phase contribution due to the dominant pole (≈ 200Hz) is negligible. Also, the 90 degree phase lag of the pole due to the feedback network (≈ 700Hz) is almost entirely cancelled by the 89 degree phase advance of the zero, at the unity gain frequency.

External compensation makes it possible to employ a de-compensated op-amp to boost accuracy, at low frequency, <u>and</u> achieve much better stability margin with a capacitive load.

3.7 An advanced OPI[2]

A useful variant of the one-plus-integrator is a quasi-differential input, single-stage HGB based on an op-amp with a two-stage RC network [1]. In this case the network is used to provide feedback with reduced phase error compared to a single-stage RC filter. At low frequency the feedback signal passes through the network $(V_1 \approx V_F)$ and any errors in the subsystem (e.g. non-linearity, offset, noise) are thereby eliminated. At high frequency the capacitors of the network provide 100% feedback with stability margin inherent to the op-amp (the same as a voltage follower). A typical application is in combination with a voltage follower (boosted accuracy: see section 3.6) with overall 100% feedback to implement a moderately high accuracy voltage follower [2] capable of driving significant capacitance.

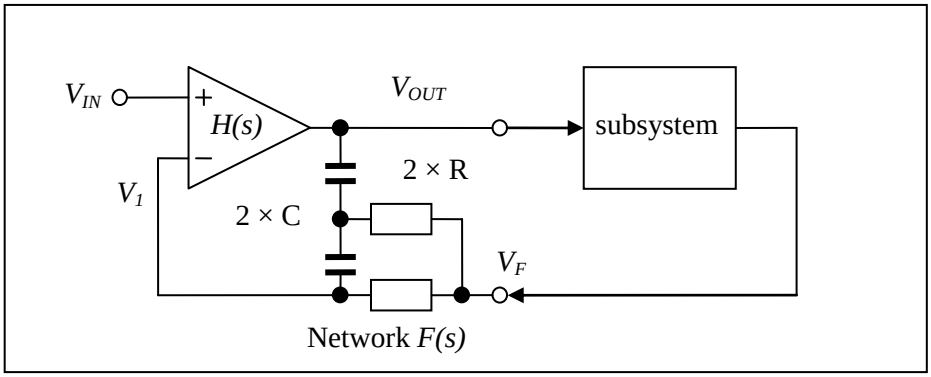


Fig. 3.7.1 An advanced OPI provides feedback to correct subsystem error

The output of the feedback network can be derived by the same method of the previous section: -

$$V_1 = V_{OUT} + F(s)(V_F - V_{OUT})$$

$$\Rightarrow \quad V_{OUT} = H(s)(V_{IN} - V_1) = H(s)(V_{IN} - V_{OUT} - F(s)(V_F - V_{OUT}))$$

With a little algebra:

$$V_{OUT} = \frac{H(s)V_{IN} - H(s)F(s)V_F}{1 + H(s)(1 - F(s))}$$

The circuit is designed to provide feedback around a subsystem for which, at low frequency, the output is nearly the same as the input with a small error represented by the parameter δ. This may be a complex number, possibly a function of frequency, or a random number to represent noise: -

$$\omega_F \approx \frac{1}{RC} \quad \text{and} \quad \omega << \omega_F \quad \Rightarrow \quad V_F = (1+\delta)V_{OUT} \quad \text{with} \quad \delta << 1$$

With a little algebra the overall transfer function is:

$$T(s) = \frac{V_F}{V_{IN}} = \frac{H(1+\delta)}{1 + H + HF\delta}$$

At low frequency the feedback network transfer function is nearly one with a small error represented by the parameter ε: -

$$\omega << \omega_F \quad \Rightarrow \quad F(s) = 1 + \varepsilon \quad \text{with} \quad \varepsilon << 1$$

$$\Rightarrow \quad T(s) = \frac{H(s)(1+\delta)}{1 + H(s) + H(s)\delta + H(s)\varepsilon\delta} = \frac{1+\delta}{1 + \delta + 1/H(s) + \varepsilon\delta}$$

1. Monograph: “Two-stage filters”. For a basic low phase error filter see section 2.2.
2. Monograph: “High accuracy voltage followers”. See section 6.3

At low frequency the open loop gain of the op-amp is large and the term $\alpha\delta$ is likely to be very small. Dividing top and bottom by $1+\delta$ has negligible effect on the size of the error terms. To a good approximation, therefore: -

$$\omega << \omega_F \;\Rightarrow\; |H(s)| >> 1 \text{ and } \alpha\delta <<< 1 \;\Rightarrow\; T(s) \approx 1 - \frac{1}{H(s)} - \alpha\delta$$

The dominant error term is due to the finite open loop gain of the op-amp. Closed loop accuracy is limited, therefore, by the advanced OPI[2] (open loop gain and CMRR) and not the subsystem.

For a more detailed analysis for a typical application see the monograph "High accuracy voltage followers" [1].

A basic OPI could be used in the same way – as a quasi-differential input single-stage HGB in a feedback loop. It is interesting, therefore, to compare the open loop characteristic of both circuits - with V_F connected to 0V. From above, repeated for convenience: -

$$V_{OUT} = \frac{H(s)V_{IN} - H(s)F(s)V_F}{1 + H(s)(1 - F(s))}$$

$$V_F = 0 \;\Rightarrow\; T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{H(s)}{1 + H(s)(1 - F(s))}$$

For a basic OPI, over a wide range of frequency (typically < 1Hz to 1MHz), one can simplify the algebra with the basic integrator model of the op-amp (see section 3.3 for detail): -

$$H(s) = \frac{1}{\tau_B s} \text{ and } F(s) = \frac{1}{1 + \tau_F s} \text{ with } \tau_F = RC \;\Rightarrow\; 1 - F(s) = \frac{\tau_F s}{1 + \tau_F s}$$

$$\Rightarrow\; \frac{\omega_F}{G} << \omega << \omega_B \;\Rightarrow\; T(s) \approx 1 + \frac{1}{\tau_F s}$$

With a two-stage low-pass filter one obtains a significantly different result. From the monograph "High accuracy filters" [2]: -

$$F(s) = \frac{3\tau_F s + 1}{\tau_F^2 s^2 + 3\tau_F s + 1} \text{ with } \tau_F = RC \;\Rightarrow\; 1 - F(s) = \frac{\tau_F^2 s^2}{\tau_F^2 s^2 + 3\tau_F s + 1}$$

With the same model for the op-amp and a little algebra: -

$$1 << |\tau_P s| << |\tau_B s| \;\Rightarrow\; H(s) = \frac{1}{\tau_B s}$$

$$\Rightarrow\; T(s) = \frac{\tau_F^2 s^2 + 3\tau_F s + 1}{\tau_B s(\tau_F^2 s^2 + 3\tau_F s + 1) + \tau_F^2 s^2}$$

At a frequency well above the filter cut-off but below the op-amp GBWP the $\tau_F^2 s^2$ terms dominate: -

$$|\tau_F s| >> 1 \text{ and } |\tau_B s| < 1 \;\Rightarrow\; T(s) \approx \frac{\tau_F^2 s^2}{\tau_B s(\tau_F^2 s^2) + \tau_F^2 s^2} = \frac{1}{\tau_B s + 1}$$

1. Monograph: "High accuracy voltage followers". See section 6.3
2. Monograph: "Two-stage filters". For a basic low phase error filter see section 2.2.

Once again the result is a voltage follower with a bandwidth limited only by the op-amp. The gain is very nearly one (the O in OPI[2]). More generally, for a frequency well below the GBWP the s^3 term is negligible: -

$$|\tau_B s| << 1 \quad \Rightarrow \quad |\tau_B s \tau_F^2 s^2| << |\tau_F^2 s^2| \quad \Rightarrow \quad T(s) \approx \frac{\tau_F^2 s^2 + 3\tau_F s + 1}{\tau_B s(3\tau_F s + 1) + \tau_F^2 s^2}$$

Also, the filter cut-off frequency is much lower than the GBWP so the term $3\tau_B \tau_F s^2$ is also negligible: -

$$\tau_B << \tau_F \quad \Rightarrow \quad |3\tau_B \tau_F s^2| << |\tau_F^2 s^2| \quad \Rightarrow \quad T(s) \approx \frac{\tau_F^2 s^2 + 3\tau_F s + 1}{\tau_B s + \tau_F^2 s^2}$$

There is a transition frequency depending on the filter cut-off frequency relative to the GBWP. For a frequency above this value: -

$$\omega >> \frac{\tau_B}{\tau_F^2} \quad \Rightarrow \quad |\tau_B s| << |\tau_F^2 s^2| \quad \Rightarrow \quad T(s) \approx \frac{\tau_F^2 s^2 + 3\tau_F s + 1}{\tau_F^2 s^2} = 1 + \frac{3}{\tau_F s} + \frac{1}{\tau_F^2 s^2}$$

The result resembles an OPI plus an extra double integrator. A more appropriate name could be an OPI+I^2.

At lower frequency the double integrator term becomes small compared to the single integrator: -

$$|\tau_F s| << 1 \text{ and } \omega << \frac{\tau_B}{\tau_F^2} \quad \Rightarrow \quad |\tau_B s| >> |\tau_F^2 s^2| \quad \Rightarrow \quad T(s) \approx \frac{\tau_F^2 s^2 + 3\tau_F s + 1}{\tau_B s} \approx \frac{3\tau_F}{\tau_B} + \frac{1}{\tau_B s}$$

At very low frequency the slope reduces to first order.

For an LF356 (GBWP = 5MHz), for example, with a filter cut-off frequency of 16kHz [1]: -

The op-amp and filter time constants are: $\tau_B = 3.2 \times 10^{-8} s$ and $\tau_F = 10^{-5} s$

The transition frequency from first order to second order roll-off is: -

$$f_T = \frac{1}{2\pi} \frac{\tau_B}{\tau_F^2} \approx 51 Hz$$

This is about the same as the dominant pole and above the lowest possible operating frequency. For this circuit, unlike the basic OPI, one needs the low frequency model of the op-amp, including finite DC gain, if one wishes to predict the low frequency response accurately: -

The resulting algebra is complicated and it is necessary to resort to a spreadsheet model. One may as well employ the fullest model, from DC to the GBWP, with the extra pole, as this will also model the high frequency response, including any resonant peak.

$$0 < \omega < \omega_E \quad \Rightarrow \quad H(s) = \frac{G}{(1 + \tau_P s)(1 + \tau_E s)}$$

With a little algebra: $$T(s) = \frac{G(\tau_F^2 s^2 + 3\tau_F s + 1)}{(\tau_F^2 s^2 + 3\tau_F s + 1)(1 + \tau_B s)(1 + \tau_E s) + G\tau_F^2 s^2}$$

1. Monograph: “High accuracy voltage followers”. See section 6.3

Typical values are [1]: -

Op-amp DC gain: $G \approx 10^5$

Gain-bandwidth product (5MHz): $\tau_B \approx 3.2 \times 10^{-8} s$

Filter frequency (16kHz): $\tau_F = 10^{-5} s$

Extra pole (with some capacitive load also ≈ 5MHz): $\tau_E \approx 3.2 \times 10^{-8} s$

The result is slightly under-damped second order low-pass characteristic which levels out to a gain of one at the filter frequency. There are two resonant peaks: at the transition from flat to second order roll-off (at ≈ 50Hz) and at high frequency, due to the extra pole (at ≈ 5MHz).

Fig. 3.7.3 Open loop characteristic of an advanced OPI+I^2 [2]

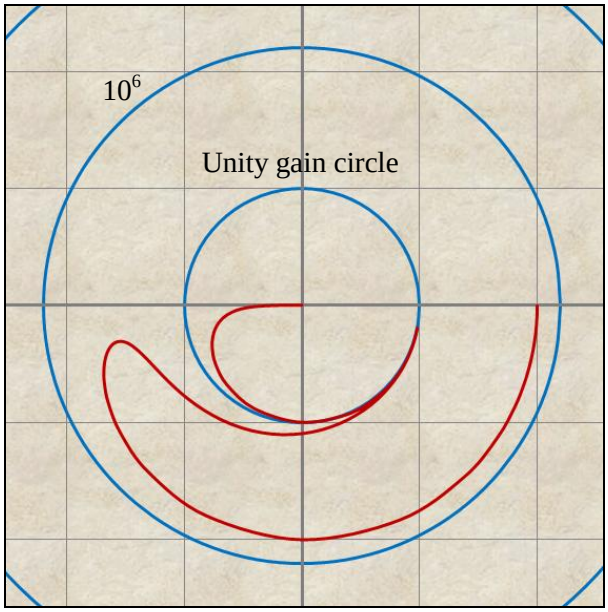


Fig. 3.7.4 The BoDayQuist plot (phase margin ≈ 90 deg.)

1. Monograph: “High accuracy voltage followers”. See section 6.3
2. Spreadsheet: “Advanced OPI”

4. Multi-stage HGBs – some practical circuits

4.1 Type 1 HGBs

A type 1 multi-stage HGB consists of one or more one-plus-integrators followed by an integrator, each based on an op-amp with external feedback. The final stage is the integrator as it is better able to drive a capacitive load (see section 3.2).

Almost any type of low cost internally compensated op-amp will work well though a moderately fast type is recommended (e.g. the JFET input LF356 with a gain-bandwidth product of 5MHz and slew rate of 12V/μs). The main applications are for high accuracy voltage followers [1] and active guard circuits [2].

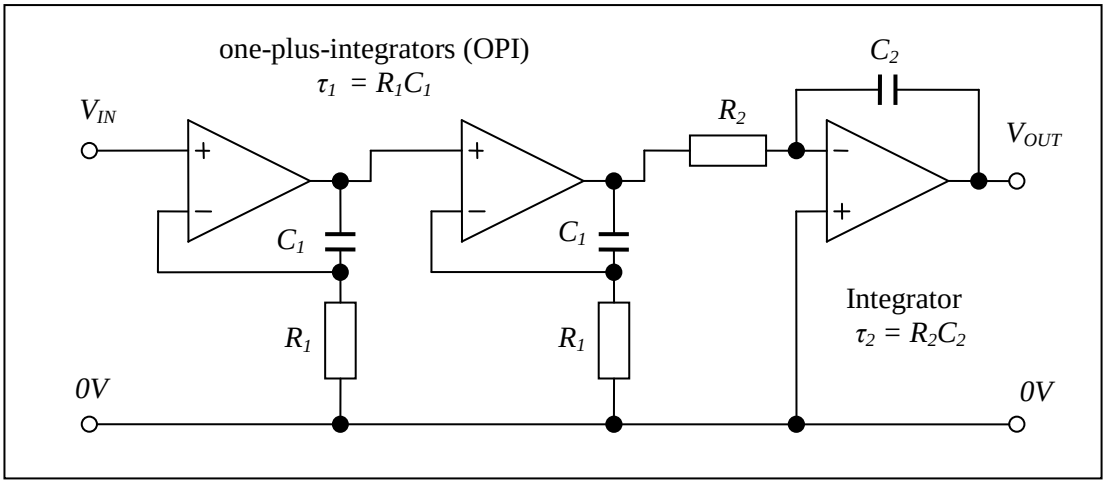


Fig. 4.1.1 A three-stage HGB (type 1)

It is shown in section 3 that the OPIs and integrators can be described, over a wide range of frequency (with reasonably good op-amps typically 1Hz to 1MHz), by relatively simple transfer functions. These do not need to include the finite op-amp gain at low frequency or the extra pole at high frequency. The overall open loop characteristic can be described, therefore, by the following transfer function in the usual complex representation $(s = j\omega)$: -

$$\frac{\omega_1}{G} << \omega << \omega_B \quad \Rightarrow \quad H_N(s) = \left(1 + \frac{1}{\tau_1 s}\right)^{N-1} \frac{1}{\tau_2 s}$$

With two or more stages ($N > 1$) the algebra is simplified if one adopts the normalised form $(s = j\omega\tau_1)$: -

$$H_N(s) = \frac{(1+s)^{N-1}}{\alpha s^N} \text{ with } \quad \alpha = \frac{\tau_2}{\tau_1}$$

The value of α determines the stability margin and, to a certain extent, low frequency errors. These are investigated in detail in the relevant monographs [1 and 2].

The physical layout of the circuit is important – coupling from the output stage to the input stage must be kept to a minimum (especially magnetic – use twisted pairs and triples). Fig. 4.1.2 illustrates an example [1].

A low noise front end is also important – for best performance use a composite amplifier based on a matched BJT long tail pair to implement the first OPI (see fig. 4.1.3).

1. Monograph: “High accuracy voltage followers”.
2. Monograph: “High accuracy amplifiers, integrators and differentiators”.

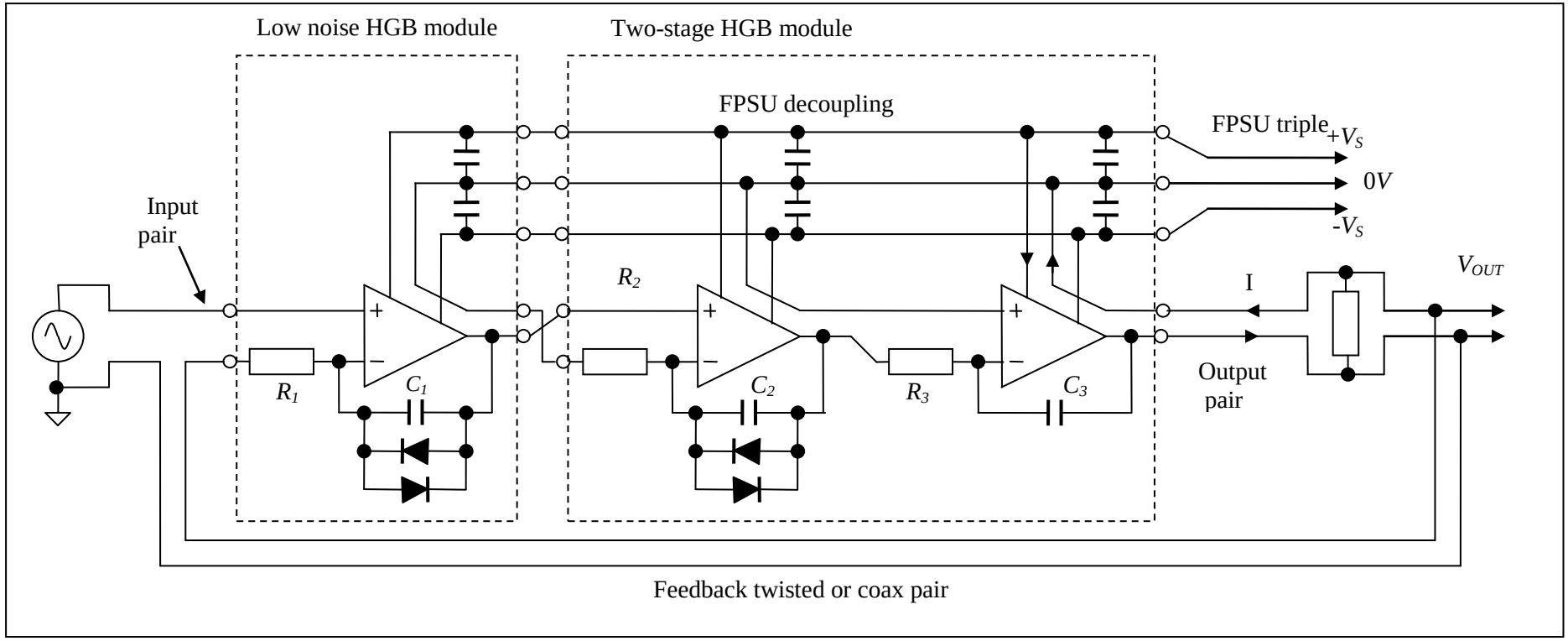


Fig. 4.1.2 A three-stage HGB configured as an inside-out voltage follower

In normalised form $(s = j\omega\tau_1)$: $\quad H_3(s) = \dfrac{(1+s)^2}{\alpha s^3}$ with $\quad \alpha = \dfrac{\tau_2}{\tau_1}$

$R_1 = 100\Omega \quad C_1 = 100nF \quad R_2 = 10k\Omega \quad C_2 = 1nF \quad R_3 = 3.9k\Omega \quad C_3 = 1nF$

$f_1 = 16kHz \quad \tau_1 = 10^{-5}s \quad f_2 = 41kHz \quad \tau_2 = 3.9\times10^{-6}s \quad \alpha = 0.39$

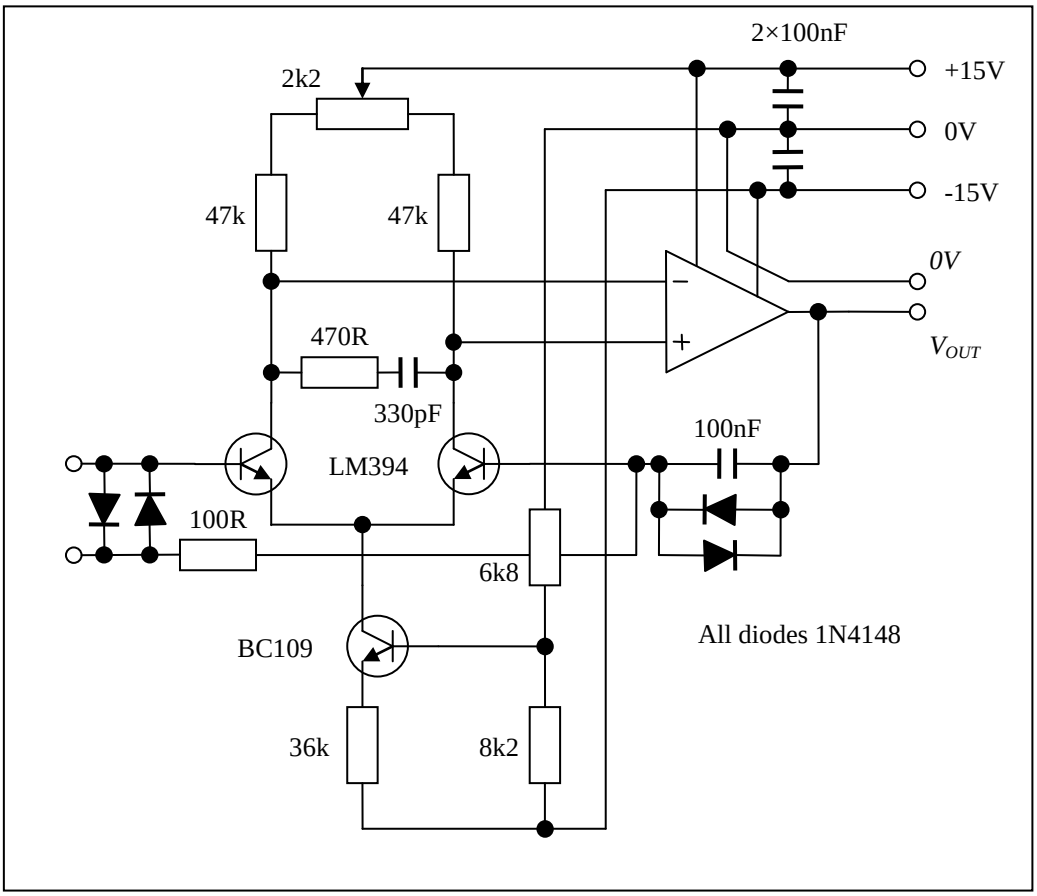


Fig. 4.1.3 A low noise OPI front-end

The BJTs are operating at 0.1mA each and generate less than $2nV/\sqrt{Hz}$ (RMS).

The feedback capacitor is quite large as the resistor needs to have a low value (to keep Johnson noise low).

For more details see the monograph “Low noise BJT pre-amplifiers” by the same author [1].

1. Monograph: “Low noise BJT pre-amps”.

4.2 A type 2 two-stage HGB

Type 2 HGBs employ the inherent integrator characteristic of typical op-amps to achieve a similar result to the type 1 HGB. Unlike the type 1 HGB, however, closed loop stability is not so predictable. A circuit that works well with one type of op-amp may oscillate when another type is plugged in. The LF353 dual JFET op-amp is recommended (GBWP of 4MHz and 13V/μs slew rate). The main applications are for inverting configurations: low phase error amplifiers, integrators and differentiators [1] where the load capacitance is small (< 50pF).

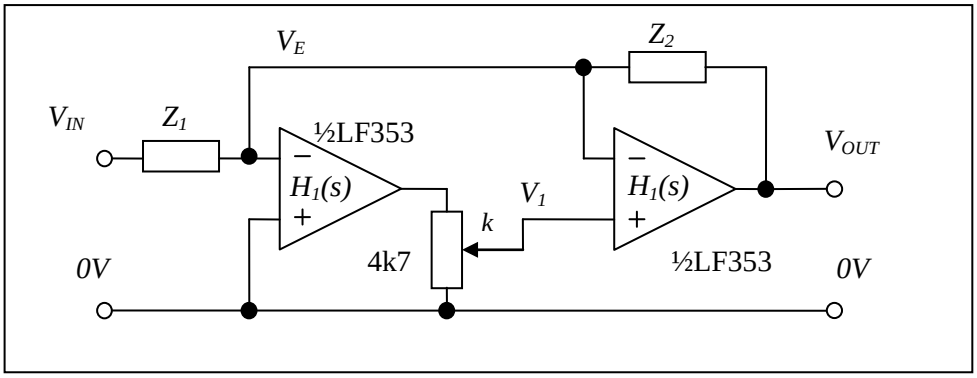


Fig 4.2.1 A two-stage (type 2) HGB in inverting mode

The combined gain of two op-amps, at low frequency (< 100Hz), is so high that the closed loop error is too small to be measured unless the feedback factor is extremely low (very high closed loop gain). There is usually no point, therefore, in employing the very low frequency model. For the low frequency analysis I shall assume the basic integrator model for each op-amp: -

$$\left|\tau_B s\right| << 1 \quad \Rightarrow \quad H_1(s) = \frac{1}{\tau_B s} \quad \text{with} \quad \tau_P = G\tau_B$$

For the first op-amp: $$V_1 = -kH_1(s)V_E = -\frac{k}{\tau_B s}V_E$$

For the second op-amp: $$V_{OUT} = H_1(s)(V_1 - V_E) = \frac{1}{\tau_B s}(V_1 - V_E)$$

Substitute the first equation into the second: -

$$s = j\omega\tau_B \qquad \Rightarrow \qquad V_{OUT} = \frac{1}{\tau_B s}\left(-\frac{k}{\tau_B s}V_E - V_E\right) = -\frac{1}{\tau_B s}\left(1 + \frac{k}{\tau_B s}\right)V_E$$

One normally defines the open loop transfer function from the non-inverting input to the output: -

$$H_2(s) = -\frac{V_{OUT}}{V_E} = \frac{1}{\tau_B s}\left(1 + \frac{k}{\tau_B s}\right)$$

For a direct comparison with type 1 HGBs normalise to the OPI time constant $\left(s = j\omega\tau_B/k\right)$: -

$$\tau_1 = \frac{\tau_B}{k} \quad \text{and} \quad \tau_2 = \tau_B \quad \Rightarrow \quad H_2(s) = \frac{1+s}{ks^2}$$

This is exactly the same as the type 1 two-stage HGB with: $k = \alpha$

High frequency analysis (extra pole) can be found for various circuits in the relevant sections [1].

1. Monograph: “High accuracy amplifiers, integrators and differentiators”.

4.3 A type 3 two-stage HGB

With a type 3 HGB the order is reversed (integrator first) to take advantage of the differential input stage of the op-amp. Latch-up is avoided and recovery from overload ensured with a limiter circuit based on a pair of zener diodes. The main application is for a high gain differential amplifier [1] as part of a quadrature servo [2]. If a low gain is required (×1 or ×10) the feedback factor must be reduced with a low value resistor across the inputs [1].

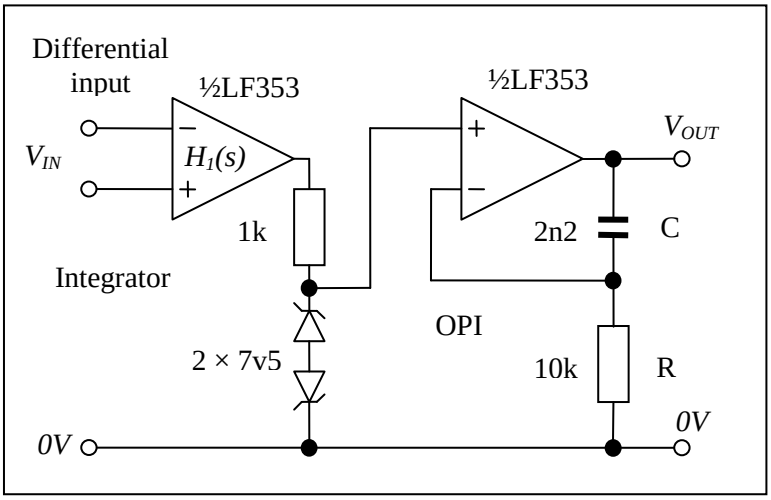


Fig. 4.3.1 A two-stage HGB (type 3)

As with a type 2 HGB, even with a closed loop gain of 1000, the loop gain with two op-amps is so high that the low frequency (<100Hz) error is too small to be measured. The low frequency characteristic can be modelled, therefore, with an integrator followed by a one-plus integrator: -

$$|\tau_B s| << 1 \quad \Rightarrow \quad H_2(s) = \frac{1}{\tau_B s}\left(1 + \frac{1}{\tau_1 s}\right) \quad \text{with} \quad \tau_1 = RC$$

Over most of the frequency range of interest (typically 100Hz to 1MHz), in normalised form $(s = j\omega\tau_1)$: -

$$\tau_2 = \tau_B \qquad \Rightarrow \quad H_2(s) \approx \frac{1+s}{\alpha s^2} \quad \text{with} \quad \alpha = \frac{\tau_B}{\tau_1}$$

This is the same as the type 1 two-stage HGB. The optimum value for α depends on the feedback factor [1]. The OPI time component values in fig. 4.3.1 are for a maximum gain of 1000 [1].

A type 3 HGB should be used with a low feedback factor (typically < 0.01) - the extra pole of the first op-amp and the resonant peak of the OPI could result in instability at high frequency. A low feedback factor ensures that the loop gain is well below 0dB at the extra pole frequency and the effect of the extra poles is negligible.

1. Monograph: “High accuracy amplifiers, integrators and differentiators”. See section 6.
2. Monograph: “Null detectors – the basics”. See section 6.

High accuracy voltage followers (HAVFs)

1. Introduction

High accuracy voltage followers (HAVFs) have numerous applications. Some resistance bridges, for example, employs a pair of HAVFs to drive the primary windings of a ratio transformer.

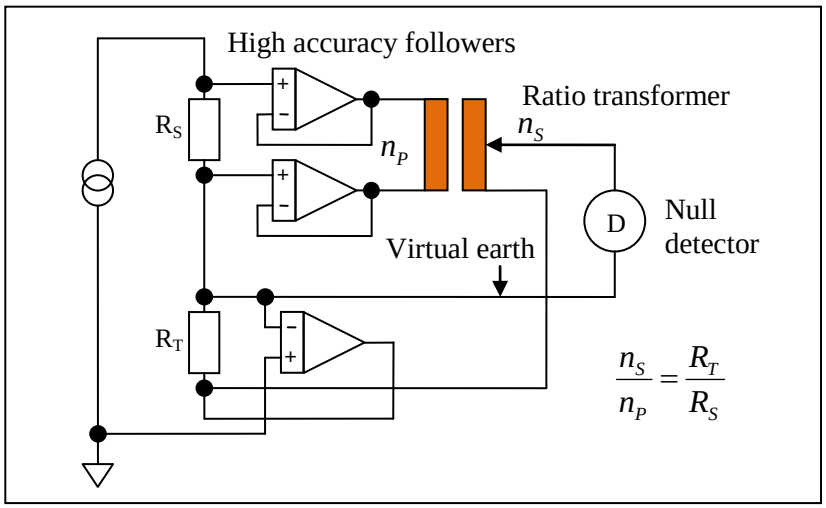


Fig. 1.1 Outline schematic of the ASL F16 and F17 series resistance bridges

Followers can also be used to eliminate the problem of loading of a ratio transformer secondary winding by providing the necessary load current. The HAVF forces the transformer secondary voltage across the two-terminal-pair reference resistor R_S. With virtually no current passing through the transformer secondary much thinner wire and semiconductor switches can be used.

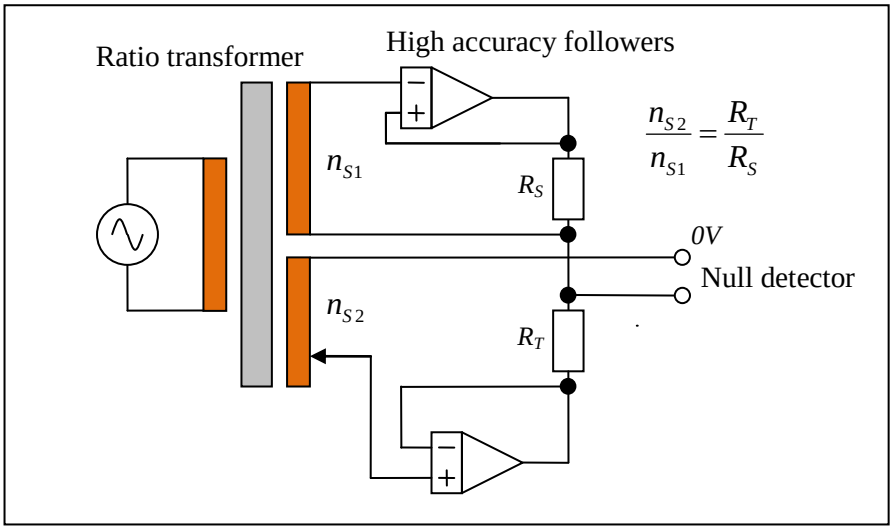


Fig. 1.2 HAVF being used to accurately drive a load (ASL model F25)

It is also possible to implement very accurate active guarding, often necessary when working at the level of parts per billion. The follower provides the drive current flowing through the stray capacitance from the active guard to local 0V (“ground”). The active guard screens the sensitive connection.

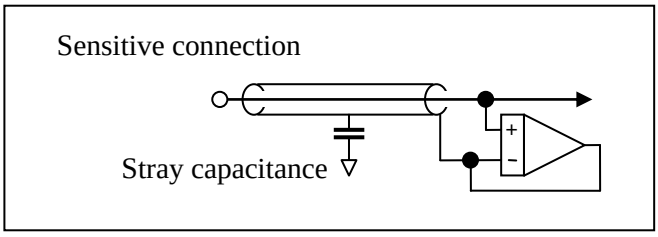


Fig. 1.3 HAVF used as an active guard

2. Theory of operation

For more detailed analysis of feedback and stability see the monograph “High gain blocks” [1].

2.1 The limitations of a conventional voltage follower

The two main problems with a conventional voltage follower, employing a single op-amp, are the limited common mode rejection ratio, open loop gain and stability margin.

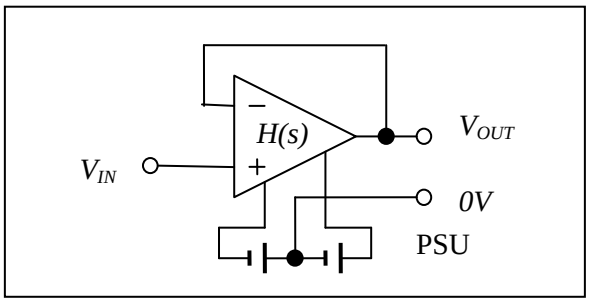


Fig. 2.1.1 A conventional voltage follower based on a single op-amp

It is shown elsewhere [2] that the closed loop transfer function, at low frequency, based on the high gain low-pass filter model of the op-amp is, to a very good approximation: -

$$\omega << \omega_B \;\Rightarrow\; H(s) = \frac{G}{1+\tau_P s} \;\Rightarrow\; T(s) = \left(1 + \frac{1}{G} + \tau_B s\right)^{-1}$$

Where G is the gain at very low frequency (i.e. the “DC gain”), τ_P is the time constant of the dominant pole (internal compensation) and $\omega_B = 1/\tau_B$ is the gain-bandwidth product (GBWP) in radians/s.

In more convenient form with frequency in Hz: -

$$f << f_B \;\Rightarrow\; T(s) \approx \left(1 - \frac{1}{G}\right)\left(1 - j\frac{f}{f_B} + \left(\frac{f}{f_B}\right)^2\right)$$

The real (in-phase) error term is a minimum $1/G$ and, therefore, limited by the DC gain of the op-amp. The imaginary (quadrature) error, increases in direct proportion to the frequency and is usually much less of a problem.

With an LF356 op-amp, for example, the DC gain can be as low as 100dB $\left(G \approx 10^5\right)$ resulting in an in-phase error of up to 10ppm and too large for many applications. The GBWP is 5MHz so that, at an operating frequency of 75Hz, the quadrature error is approximately 15ppm (15µrad).

The finite common mode rejection ratio, which could also be as low as 100dB, contributes similar errors. See the next section.

The very wide bandwidth and low stability margin, especially when driving a capacitive load (e.g. long cables) results in a resonant peak at high frequency (typically 5MHz) [2] and severe practical problems. The solution is to employ a robustly stable (type 1) multi-stage high gain block (HGB).

1. Monograph: “High gain blocks”. See, especially, section 2.
2. Ibid: See section 3.4

2.2 Common mode rejection and the "inside-out" configuration

The main problem with the conventional follower is common mode due to mismatch in the differential input stage. The inputs of the HGB vary with respect to the power supply, resulting in a significant error. Fortunately there is a simple and elegant solution to this problem – the "inside-out" mode of operation. The basic principle is to place the input signal in series with the feedback connection. The action of the feedback is the same – the output of the HGB ramps up or down until its inputs are the same. Only then does the output stabilise. In the case of the inside-out follower the power supply is "floating" and is driven to follow the input voltage, thus eliminating the common mode completely.

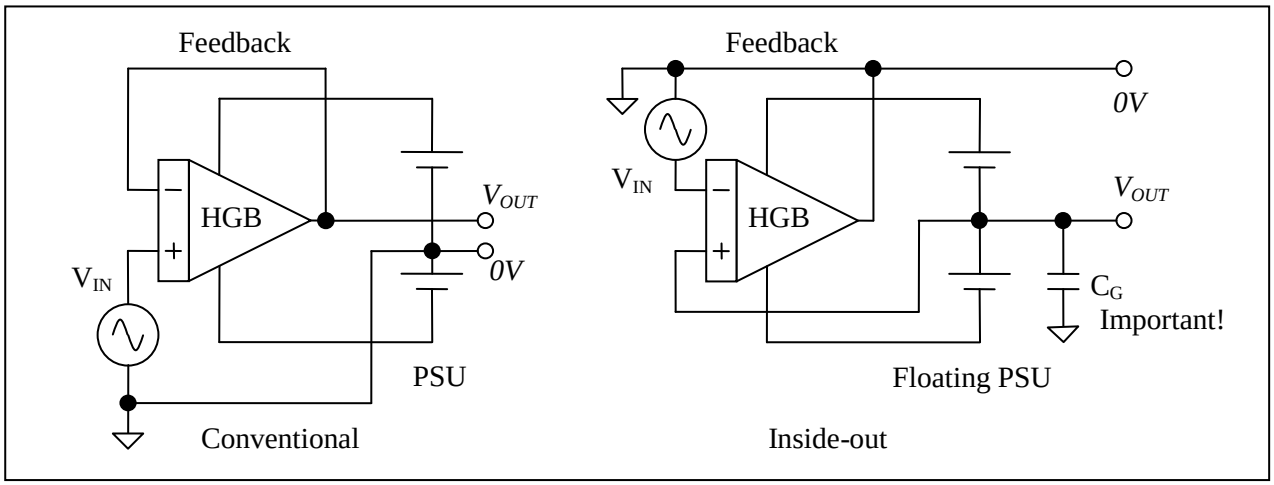


Fig. 2.2.1 Conventional follower and the inside-out follower

The earth symbols indicate the part of the circuit which may be connected to local 0V while testing (e.g. the signal generator and/or oscilloscope inputs) or are at or near local earth potential. In practice a high accuracy system should have only one earth connection.

At first sight it appears very strange that the output pin of the op-amp (final stage of the HGB) is connected to earth. Old habits die hard and this author admits to many head-scratching moments before realising that the 'scope test lead earth was clipped to the PSU 0V.

The capacitance from the centre of the PSU to ground is also marked as important. This is because it is usually quite large, due mainly to the capacitance across the power supply mains transformer. The capacitive load can combine with the op-amp output resistance resulting in extra phase shift and reduced stability margin. The open loop frequency response of the HGB is designed to mitigate this problem. For more details on stability see the monograph "High gain blocks" [1].

2.3 The floating power supply (FPSU)

Each follower requires its own floating power supply. This need not be expensive (the power required is usually quite low) but the capacitance to ground is an important factor. The best type of commercially available PSU for this task employs linear regulators with low noise and ripple and a mains transformer with low capacitance between primary and secondary. The author has found that a power supply transformer with split sections is perfectly adequate for all but the most demanding applications. For the ultimate in performance a special PSU can be constructed using a small (high frequency) transformer, with separated or screened windings, operating from a higher frequency AC power source. It is also possible to use rechargeable batteries with guarded screening to minimise this problem but this is not very practicable.

1. Monograph: "High gain blocks".

2.4 Closed loop transfer functions

The open loop transfer function for an *N*-stage HGB is: $$H_N(s)=\left(1+\frac{1}{\tau_1 s}\right)^{N-1}\frac{1}{\tau_2 s}$$

I.e. the HGB consists of *N*-1 "one-plus-integrators" (OPIs), followed by a final integrator. The algebra is simplified if one employs the normalised form (normalised to the OPI time constant): -

$$s=j\omega\tau_1 \qquad \text{with} \qquad \alpha=\frac{\tau_2}{\tau_1}\le 1 \qquad \Rightarrow \qquad H_N(s)=\frac{(1+s)^{N-1}}{\alpha s^N}$$

The closed loop transfer function for an N-stage voltage follower is, therefore: -

$$T_N(s)=\frac{H(s)}{H(s)+1}=\frac{(1+s)^{N-1}}{\alpha s^N+(1+s)^{N-1}}$$

The parameter $\alpha=\tau_2/\tau_1$ has a significant effect on stability margin (*N* > 1). The values chosen are at the upper limit of what is practical and must, therefore, be regarded as a maximum in each case: -

In normalised form $(s=j\omega\tau_1)$		α for 60% or +4dB peaking
Single-stage: *N=1*	$T_1(s)=\frac{1}{\alpha s+1}$	No limit
Two-stage: *N=2*	$T_2(s)=\frac{s+1}{\alpha s^2+s+1}$	1
Three-stage: *N=3*	$T_3(s)=\frac{s^2+2s+1}{\alpha s^3+s^2+2s+1}$	0.4
Four-stage: *N=4*	$T_4(s)=\frac{s^3+3s^2+3s+1}{\alpha s^4+s^3+3s^2+3s+1}$	0.2
Five-stage: *N=5*	$T_5(s)=\frac{s^4+4s^3+6s^2+4s+1}{\alpha s^5+s^4+4s^3+6s^2+4s+1}$	0.15

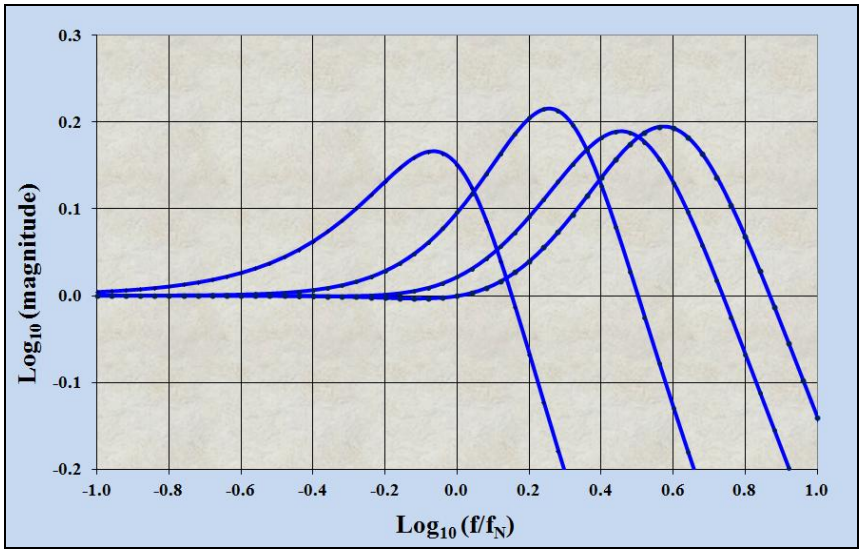


Fig. 2.4.1 Frequency response of *N*-stage voltage followers (*N* = 2 to 5 left to right).

2.5 Error analysis

As above the algebra is made easier if one normalises to the time constant of the OPIs $(s = j\omega\tau_1)$.

The transfer function for an *N*-stage follower is: $$T_N(s) = 1 - \frac{\alpha s^N}{\alpha s^N + (1+s)^{N-1}}$$

For a single op-amp ($N = 1$) the approximation is fairly simple. At low frequency: -

$$N = 1 \quad \text{and} \quad |s| << 1 \quad \Rightarrow \quad T_1(s) = 1 - \frac{\alpha s}{\alpha s + 1} \approx 1 - \alpha s + (\alpha s)^2$$

This simplifies to a more convenient form, with frequency in Hz: -

$$T_1(s) \approx 1 - j\frac{f}{f_2} + \left(\frac{f}{f_2}\right)^2 \quad \text{with} \quad f_2 = \frac{1}{2\pi\tau_2} \quad \text{(the gain-bandwidth product of the op-amp)}$$

For two-stage HGBs and higher the algebra is a little more subtle. The denominator of the error term is approximately 1 so that for $N > 1$: -

$$N > 1 \quad \text{and} \quad \alpha|s|^N << 1 \quad \Rightarrow \quad T_N(s) \approx 1 - \frac{\alpha s^N}{(1+s)^{N-1}} \approx \frac{\alpha s^N}{1 + (N-1)s} \approx 1 - \alpha s^N\left(1 - (N-1)s + O(s^2)\right)$$

To a very good approximation, therefore: -

$$N > 1 \quad \text{and} \quad |s| << 1 \quad \Rightarrow \quad T_N(s) \approx 1 - \alpha s^N + \alpha(N-1)s^{N+1}$$

The first two error terms are sufficient for both the real (in-phase) and imaginary (quadrature) components: -

In more convenient form with frequency in Hz $(f_1 = 1/2\pi\tau_1)$: -

$$N > 1 \quad \text{and} \quad f << f_1 \quad \Rightarrow \quad T_N(s) \approx 1 - \alpha\left(j\frac{f}{f_1}\right)^N + \alpha(N-1)\left(j\frac{f}{f_1}\right)^{N+1}$$

I shall employ the maximum values for α.

Two-stage: $N=2,\ \alpha = 1 \quad \Rightarrow \quad T_1(f) \approx 1 + \left(\frac{f}{f_1}\right)^2 - j\left(\frac{f}{f_1}\right)^3$

The largest term is real (in-phase). The imaginary term is third order and represents a much smaller phase error than would be the case with a follower based on a single op-amp.

Three-stage: $N=3,\ \alpha = 0.4 \quad \Rightarrow \quad T_3(f) \approx 1 - j0.4\left(\frac{f}{f_1}\right)^3 + 0.8\left(\frac{f}{f_1}\right)^4$

It is unlikely that N > 3 designs offer any further advantage.

The reader may notice that the transfer functions and, therefore, the error analysis have the same form as an *N*-stage low pass filter [1]. See part 2 [1].

1. Monographs: “Two-stage filters” and “Three-stage filters”.

3. Power supply and signal wiring.

The HGB is carefully constructed so that the input and output signal are twisted or co-axial pairs. The power supply also enters the HGB as a twisted triple. The output pair and FPSU triple must be kept a safe distance (a few cm) away from the input and feedback pair (certainly not bunched in the same loom). This ensures minimal problems with stray magnetic flux. The following diagram illustrates the correct layout for power supply and signal wiring. The arrows indicate the presence of go/return current which, being physically close, means a minimum of external magnetic flux.

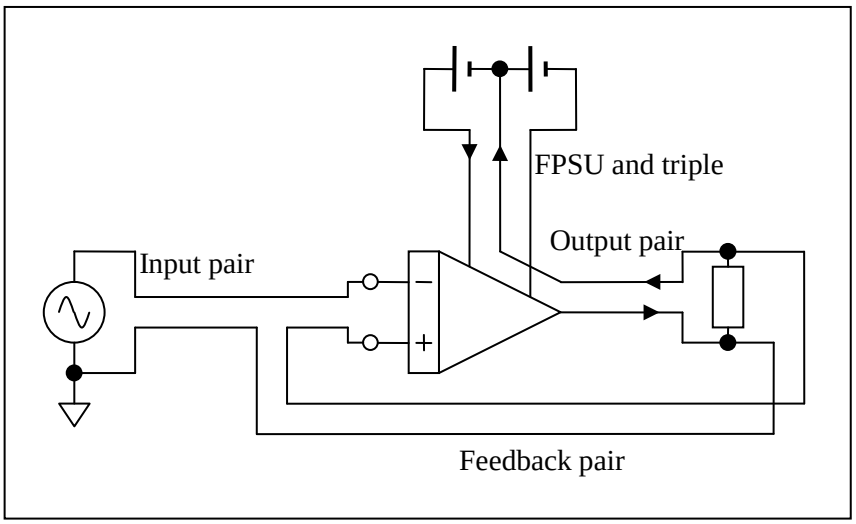


Fig. 3.1 HAVF with more detail of signal pair and PSU triple routing

4. Measuring the accuracy of an inside-out follower

For testing purposes it is possible to reverse the input signal connections so that the 0V (earth potential) is on the HGB inverting input side and the final stage op-amp output is driven, relative to the power supply. The action of the feedback is the same and there is no difference, as far as the internal circuitry of the follower is concerned. Both inputs are now at, or close to, 0V (very low common mode) and a low noise, high gain, differential amplifier can be used to boost the error signal so that it can be measured with an AC voltmeter or spectrum analyser. The amplifier needs its own power supply. The HAVF floating power supply must still be allowed to float - one must not connect FPSU 0V to earth.

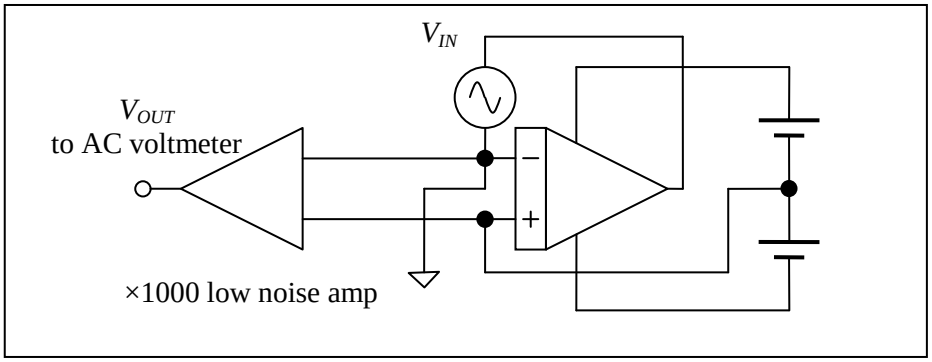


Fig. 4.1 Error test configuration (signal generator connections reversed)

5. Stability considerations

HAVF stability is a complex topic when one considers non-linear effects such as slew rate limiting and recovery from power-up and overload conditions. It has been found in practice, however, that a number of simple measures can be taken so that stability is ensured even with complex and variable load impedance in a wide range of applications.

The main trade-off is, not surprisingly, between frequency of operation and accuracy. One of the main (and most demanding) applications of high accuracy voltage followers, for example, is platinum resistance thermometry, operating at 25Hz or 75Hz for a 50Hz supply (30Hz or 90Hz for a 60Hz supply). At this low frequency it is possible to achieve sub-ppm accuracy while maintaining stability, even while driving a ratio transformer energising winding (with resonant leakage inductance and inter-winding capacitance).

Capacitance bridges usually operate at a frequency of 1-10 kHz. At the higher end it is still possible to achieve sufficient accuracy for HAVFs to be useful (e.g. as drivers for active guards).

The general principle is to design for an HGB with the lowest possible unity gain frequency while still achieving the necessary open loop gain at the frequency of operation. In many cases a two or three stage design is adequate and the gain is well below 0dB at a frequency much lower than the range where capacitive and resonant loads become an issue.

Recovery from overload is made possible by connecting a pair of back-back diodes in parallel with the feedback capacitors (apart from the last stage). In normal operation the voltage across the diodes is very small and they present high impedance. When overloaded, however, the diodes conduct and the one-plus integrators revert to a ×1 follower, allowing the circuit to recover. This can be understood with the aid of a BarDayQuist plot of the open loop gain of an extreme example (4-stage HGB) [1]. As each OPI overloads the phase shift reduces by 90 degrees and the locus collapses to the centre without passing through the critical -1 point.

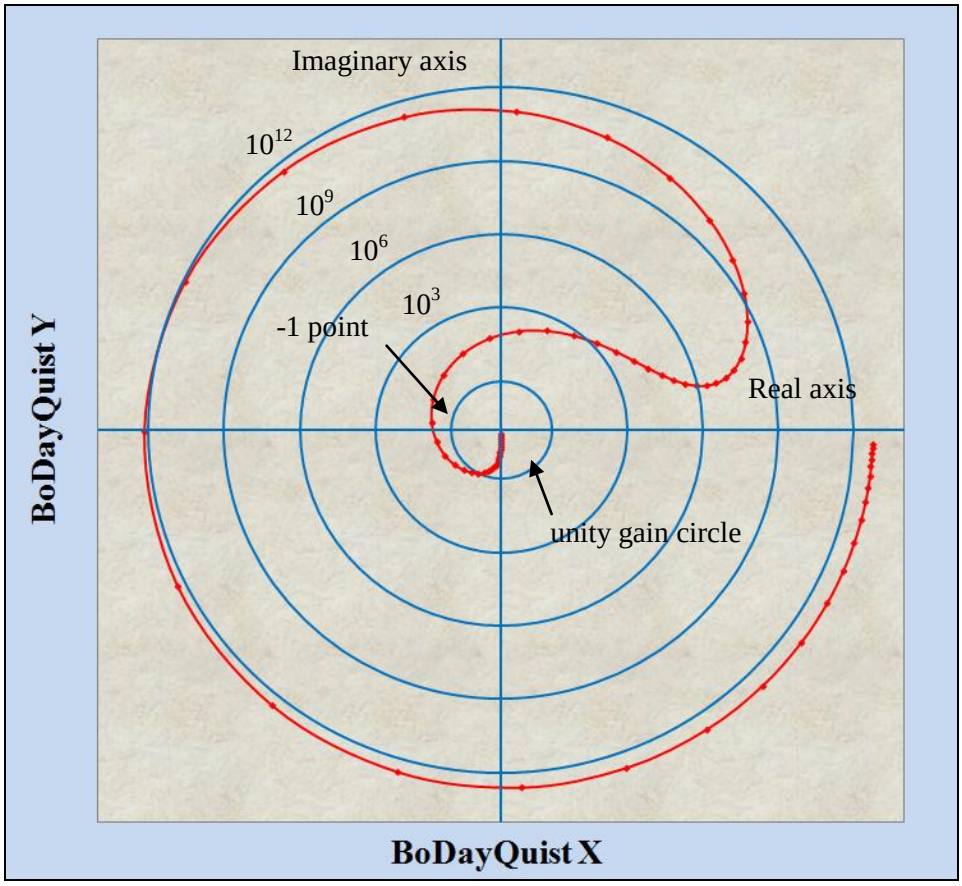


Fig. 2.4.1 The "BoDayQuist" plot illustrates how the stability criterion is satisfied (4-stage HGB)

1. Monographs: "High gain blocks". See section 2

6. Practical circuits

6.1 A two-stage HAVF

A two-stage HAVF has sufficient accuracy for an active guard and providing a virtual earth (see section 1). The circuit, fig. 6.1.1 is shown in normal test configuration (not error test) with a load resistor (see section 3).

Op-amps: 2 × LF356 or similar

Time constants $R_1 = 10k\Omega \quad C_1 = 1nF \quad \tau_1 = 10^{-5}s$

$R_2 = 10k\Omega \quad C_2 = 330pF \quad \tau_2 = 10^{-5}s \quad \alpha = \frac{\tau_2}{\tau_1} = 1$

FPSU decoupling: 2×100nF ceramic per op-amp.

6.2 A three stage HAVF with low noise front end

This three-stage design employs a low noise front end (see fig. 6.2.1 and 6.2.2) with collector current set at 0.1mA. It has sufficiently high accuracy, low noise current and low DC offset for driving the energising windings of an F18 type ratio transformer [1] (see also fig. 1.1). It is also shown in normal test configuration with a load resistor: -

Op-amps: 3 × LF356 or similar

Time constants $R_1 = 100\Omega \quad C_1 = 100nF \quad \tau_2 = 10^{-5}s$ (reduced resistance for lower noise)

$R_2 = 10k\Omega \quad C_2 = 1nF \quad \tau_1 = 10^{-5}s$

$R_3 = 10k\Omega \quad C_3 = 330pF \quad \tau_2 = 3.3\times10^{-6}s \quad \alpha = \frac{\tau_2}{\tau_1} = 0.33$

FPSU decoupling: 2×100nF ceramic per op-amp.

With the current source (0.2mA) and collector resistances shown the noise resistance is approximately 1kΩ.

The same design can be used to drive the energising windings and ratio primary of a two or three-stage ratio transformer [2]. Best noise performance is achieved when the front-end is matched to the source resistance. This usually involves increased collector current with lower voltage noise and higher current noise (lower noise resistance). Collector current should be no higher than 1 or 2mA (each), however, to ensure low self-heating and maintain low DC offset [3].

1. Monograph: “An F18 type ratio transformer bridge”.
2. Monograph: “An F17 type ratio transformer bridge”.
3. Monograph: “Low noise BJT pre-amps”.

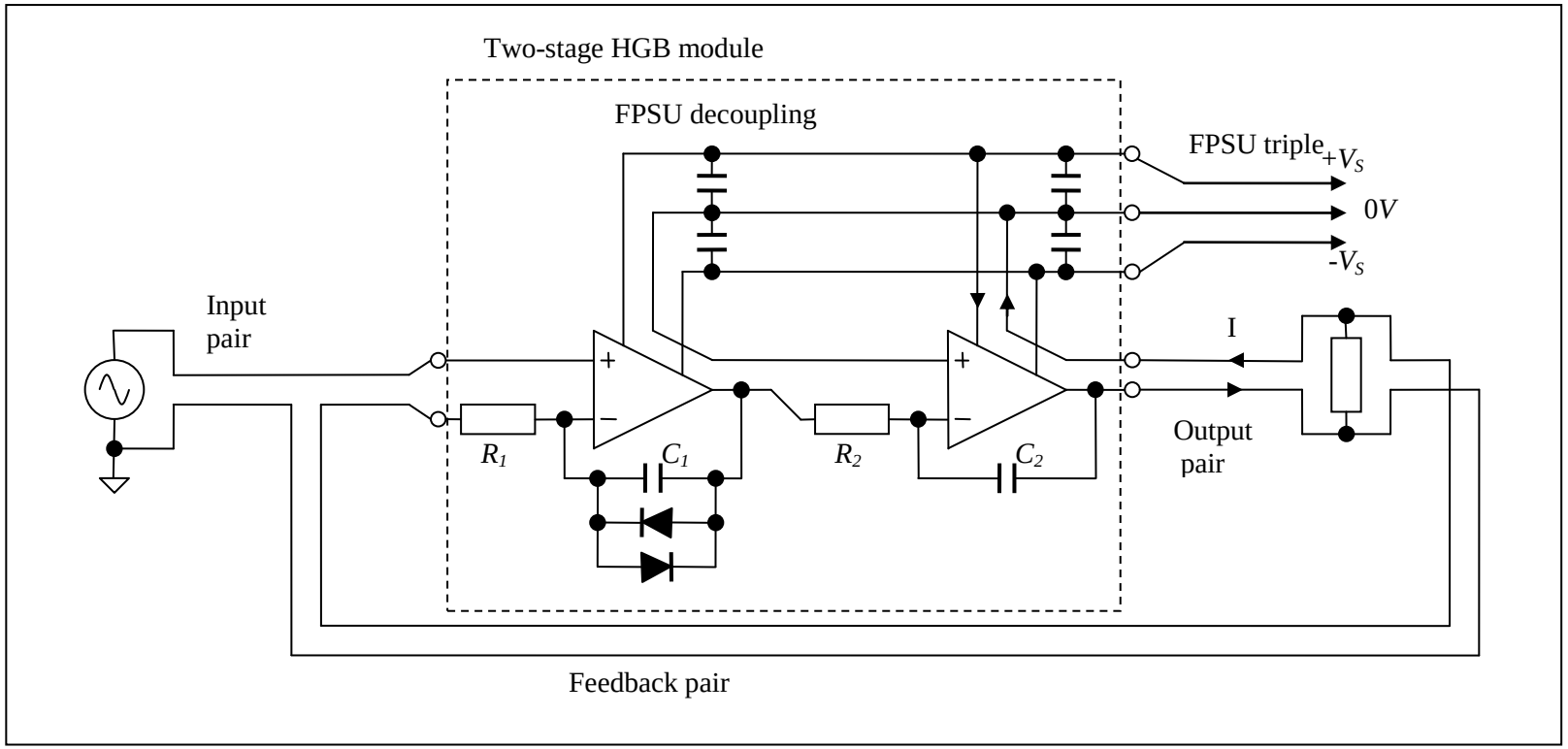


Fig. 6.1.1 A two-stage HGB with routing detail (normal operation)

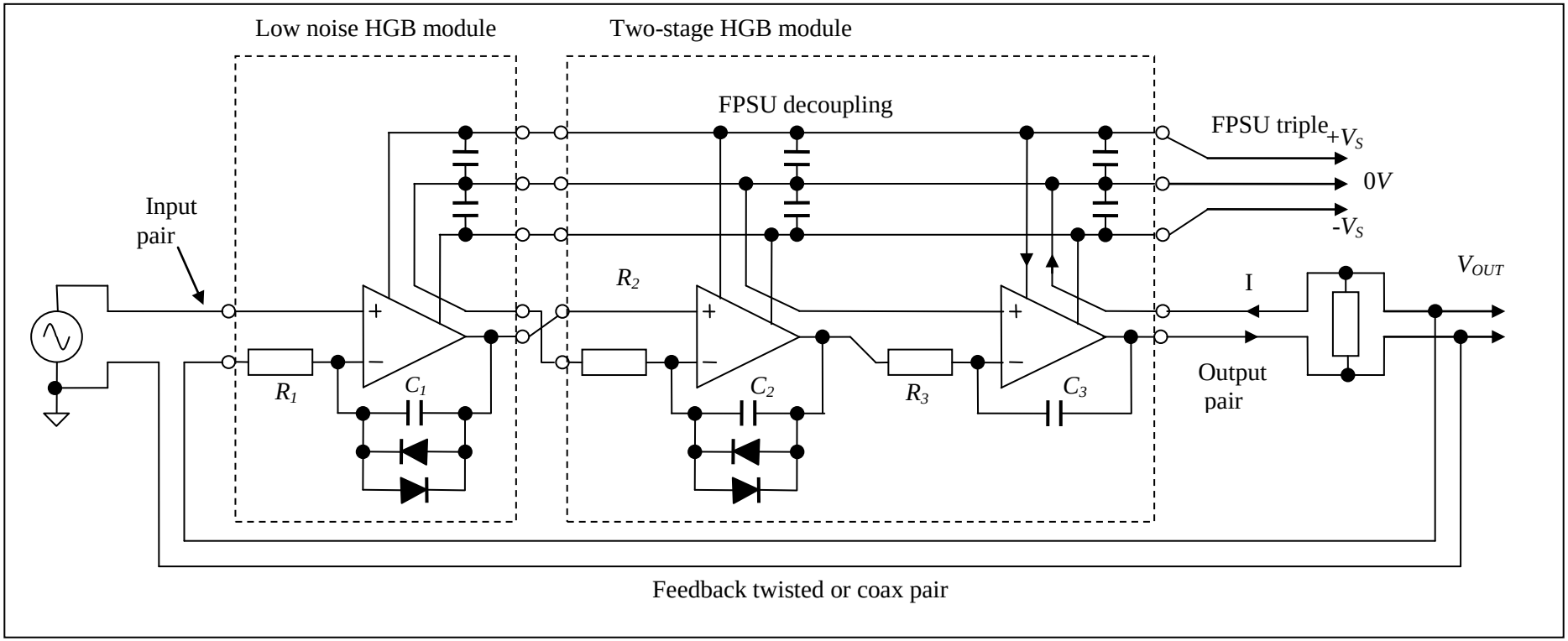


Fig. 6.2.1 Three-stage HGB with low noise front end

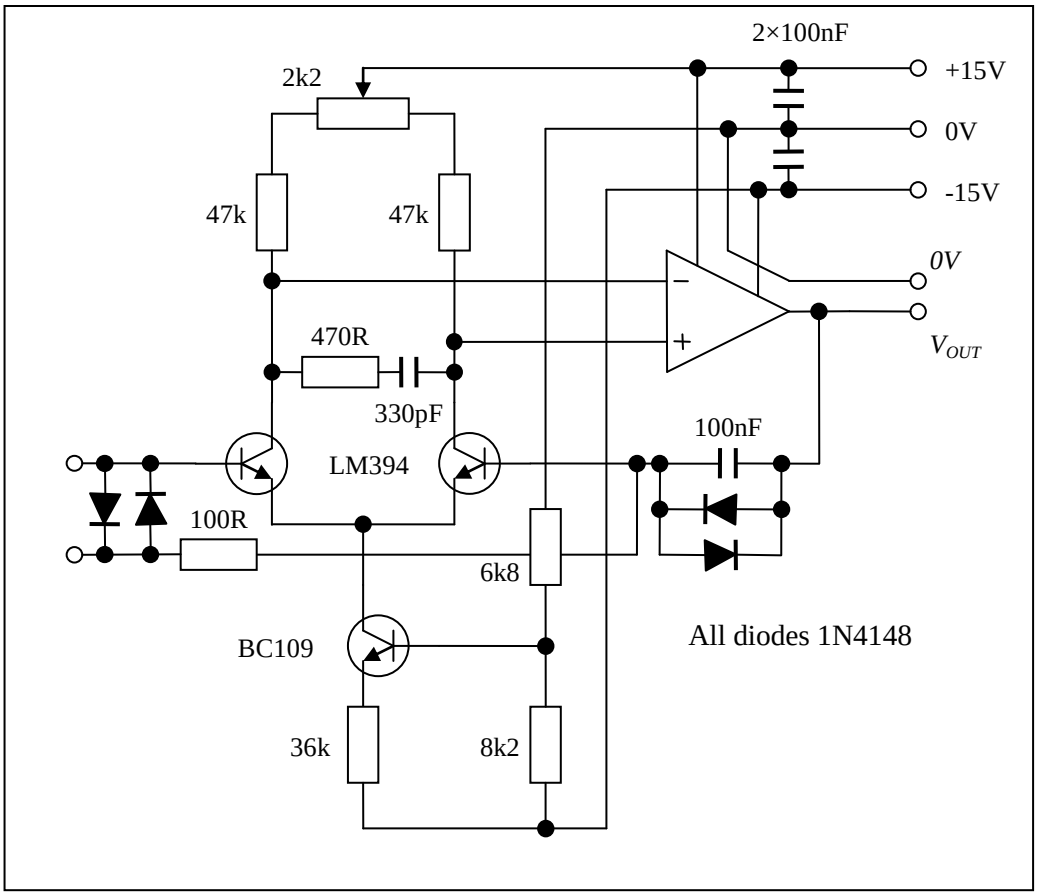


Fig. 6.2.2 A low noise OPI front-end

The BJTs are operating at 0.1mA each and generate less than $2nV/\sqrt{Hz}$ (RMS).

For more details see the monograph “Low noise BJT pre-amplifiers” by the same author [1].

1. Monograph: “Low noise BJT pre-amps”.

6.3 A low cost follower with moderately high accuracy

Whereas not as accurate as a multi-stage inside-out follower this circuit may be found useful with an accuracy approaching 1ppm. The cost is reduced by operating from a normal power supply so that the main limitation is the common mode rejection ratio (CMRR) of the first op-amp. The main application is for a low cost resistance bridge (c.f. ASL model F25) depicted in fig. 1.2.

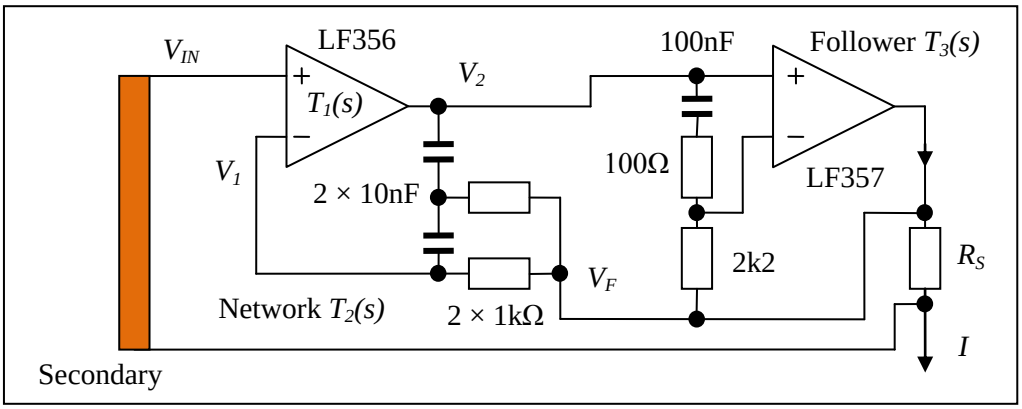


Fig. 6.3.1 A low cost voltage follower

The first stage is an op-amp with a two-stage feedback network employed as a single-stage HGB. I have chosen the name "advanced OPI2" by analogy with the one-plus integrator [1] which is described in some detail elsewhere [2]. The signal being fed back is the voltage sensed across the reference resistor. At high frequency the network provides 100% feedback with plenty of stability margin despite the second stage follower. At low frequency the overall error (difference between V_{IN} and V_F) is very small due to the full open loop gain of the first op-amp (small difference between V_{IN} and V_1), the low phase error of the two-stage network (small difference between V_1 and V_F) and the moderately high accuracy of the second stage follower (small difference between V_2 and V_F).

The second stage follower employs an op-amp with reduced internal compensation so that the effective gain-bandwidth product (at low frequency) is typically 20MHz (e.g. LF357). The external compensation ensures plenty of stability margin even with a highly capacitive load (long cables to the reference resistor) [2]. The purpose of the second stage follower is to use feedback to overcome the resistance of the current carrying wire to the reference resistor. The very small difference between V_2 and V_F also ensures that the current flowing in the voltage sensing wire is negligible. The external compensation requires a low source resistance for the non-inverting input.

The open loop transfer function of the first op-amp over most of the frequency range (up to the gain-bandwidth product of typically 5MHz) can be accurately described as an integrator: -

$$T_1(s) = \frac{1}{\tau_{B1} s} \quad \text{with} \quad |T_1| >> 1$$

The feedback network is a two-stage low-pass filter and can be accurately described by [1]: -

$$T_2(s) = \frac{3\tau_N s + 1}{\tau_N^2 s^2 + 3\tau_N s + 1}$$

At a frequency significantly below the natural frequency: $|\tau_N s| << 1 \quad \Rightarrow \quad T_2(s) \approx 1 - \tau_N^2 s^2 + 3\tau_N^3 s^3$

If one assumes that the lead resistance is small compared to the reference resistor the second stage follower is (see section 2.1):

$$T_3(s) = \frac{1}{1 + \tau_{B2} s} \quad \text{so that} \quad |T_3| \approx 1$$

1. Monograph: "Two-stage filters". See section 2.2.
2. Monograph: "High gain blocks". See sections 3.6 and 3.7

Referring to fig. 6.3.1: -

$$V_2 = T_1(V_{IN} - V_1) \quad \text{and} \quad V_1 = V_2 + T_2(V_F - V_2) \quad \text{and} \quad V_F = T_3 V_2$$

Use the second and third equation to eliminate V_1 and V_2: -

$$\Rightarrow \quad \frac{V_F}{T_3} = T_1 V_{IN} - T_1\left(\frac{V_F}{T_3} + T_2\left(V_F - \frac{V_F}{T_3}\right)\right)$$

$$V_F(1 + T_1 - T_1T_2 + T_1T_2T_3) = T_1T_3V_{IN}$$

The overall transfer function is, therefore: -

$$T(s) = \frac{V_F}{V_{IN}} = \frac{T_1T_3}{1 + T_1(1 - T_2) + T_1T_2T_3}$$

$$\Rightarrow \quad T(s) = \frac{1}{T_2}\left(\frac{T_1T_2T_3}{1 + T_1(1 - T_2) + T_1T_2T_3}\right) = \frac{1}{T_2}\left(1 + \frac{1 - T_2}{T_2T_3} + \frac{1}{T_1T_2T_3}\right)^{-1}$$

The second and third terms in the brackets are small corrections and dividing by T_2T_3 makes a negligible difference so that, to a very good approximation: -

$$|T_2| \approx 1 \quad \text{and} \quad |T_3| \approx 1 \quad \Rightarrow \quad T(s) \approx \frac{1}{T_2}\left(1 + 1 - T_2 + \frac{1}{T_1}\right)^{-1}$$

From above, at low frequency: $|\tau_N s| << 1 \quad \Rightarrow \quad T_2(s) \approx 1 - \tau_N^2 s^2 + 3\tau_N^3 s^3$

$$\Rightarrow \quad T(s) \approx (1 - \tau_N^2 s^2 + 3\tau_N^3 s^3)^{-1}(1 + \tau_N^2 s^2 - 3\tau_N^3 s^3 + \tau_{B1} s)^{-1}$$

The lower order terms in $\tau_N s$ cancel and it is likely, therefore, that the accuracy is limited mainly by the finite gain-bandwidth product of the first op-amp: -

$$|\tau_N s|^4 << |\tau_{B1} s| \quad \text{and} \quad |\tau_{B1} s| << 1 \quad \Rightarrow \quad T(s) \approx (1 + \tau_{B1} s)^{-1} \approx 1 - \tau_{B1} s + \tau_{B1}^2 s^2$$

In practice the second order term is negligible and the CMRR is the main source of in-phase error (typically 100dB resulting in < 10ppm). The contribution due to the second stage follower appears to be negligible!

In the example above the first op-amp has a gain-bandwidth of 5MHz, the feedback network a natural frequency of 16kHz and the operating frequency is 400Hz. The relevant factors are: -

$$|\tau_{B1} s| = \frac{f}{f_{B1}} \approx 8 \times 10^{-5} \quad \text{and} \quad |\tau_N s| = \frac{f}{f_N} \approx 2.5 \times 10^{-2}$$

The main source of error is 80ppm quadrature which is easily rejected by a null detector with reasonable phase accuracy and a synchronous rectifier [1].

1. Monograph: “Null detectors – the basics”.

High accuracy amplifiers, integrators and differentiators

1. Introduction

This monograph is concerned with feedback around a two-stage high gain block (HGB) in the inverting configuration with a small capacitive load. The action of feedback is such that the inverting input is maintained at very nearly local 0V ("virtual earth") and the common mode voltage is negligible. Closed loop accuracy is limited only by the feedback components and the gain of the HGB. There are four main applications: -
a). Inverting amplifiers and a differential amplifier - the latter as part of a quadrature servo.
b). Integrators: These are useful building blocks for active filters and two-phase oscillators.
c). Differentiator: For generating an accurate 90 degree phase shift as part of a quadrature servo.
d). Charge amplifiers: The feedback components are small capacitors and a very large resistor [3].

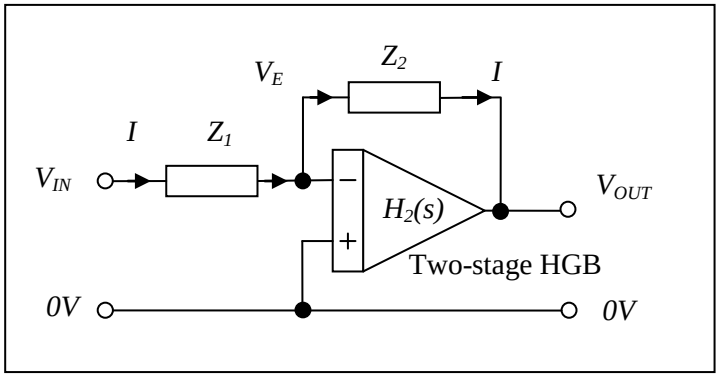


Fig. 1 The general inverting configuration with a high gain block (HGB).

Whereas a single high performance op-amp may offer sufficient in-phase accuracy for some applications they can be quite expensive and phase error (quadrature) can be an issue. A more advanced two-stage HGB (types 1 or 2) offers significantly lower phase error as well as lower cost. A differential amplifier can be implemented with a two-stage type 3 HGB. For circuits employing a single op-amp see section 3 of "High gain blocks" [1].

It is unlikely that a three-stage HGB, or higher, offers any advantage. The improved in-phase accuracy is of no benefit as this is limited by the feedback components.

For a high accuracy amplifier the resistors need to be high precision, low temperature coefficient types. Matched pairs of resistors, usually fabricated on the same substrate, are a good choice. If a high value feedback resistor (>10kΩ) is necessary then a small capacitor in parallel eliminates the extra phase shift (interaction with the input capacitance) which could otherwise affect stability.

For a high accuracy integrator a low phase error ("low tanδ") capacitor should be used (e.g. polypropylene) for the feedback component. A good example is a circuit designed to measure tanδ [2]. Even better is to include a compensator for the phase error based on its repeatability and stability.

For a high accuracy differentiator, as part of a quadrature servo, a low phase error transformer, based on a gapped low loss ferrite pot core, is the preferred choice. The circuit converts the input (sinusoidal) voltage to a current, through the transformer primary so that the output of the transformer is accurately in quadrature (precisely 90 degrees phase shift relative to the input). The transformer also provides an isolated output so that it can be added to the output of a ratio transformer.

Charge amplifiers are mostly used as low noise null detectors and so accuracy is seldom an issue [3]. The option is included here as the analysis is almost the same as for an inverting amplifier and high accuracy may prove useful in some applications.

1. Monograph: "High gain blocks".
2. Monograph: "A circuit for measuring tanδ".
3. Monograph: "Null detectors – the basics".

2. General analysis

For the more general discussion of feedback and stability for three-stage HGBs and higher see the monograph "High gain blocks" [1]. There it is argued that a sufficiently good model of the open loop transfer function of a two-stage HGB need not include the very low frequency (finite DC gain) model. The loop gain is usually so high that it is not possible to measure the error at very low frequency. For the low frequency error analysis I shall assume, therefore, an integrator and one-plus-integrator model: -

$$H_2(s)=\left(1+\frac{1}{\tau_1 s}\right)\frac{1}{\tau_2 s}$$

This model works well over a wide frequency range (typically 1Hz to 1MHz) for HGB type 1. It works less well, at high frequency, for types 2 and 3, especially when driving a capacitive load [1] due to the "extra poles". The analysis is simplified if one normalises to the time constant of the one-plus integrator $(s = j\omega\tau_1)$: -

$$H_2(s)=\frac{(1+s)}{\alpha s^2} \quad \text{with} \quad \alpha=\frac{\tau_2}{\tau_1}$$

The parameter α is chosen for the required stability margin and depends on the feedback factor.

At low frequency the HGB open loop gain can be enormous and the error voltage (at the inverting input) is very small. The open loop transfer function is defined from the non-inverting input to the output so that $(s = j\omega\tau_1)$: -

$$|s|<<1 \;\Rightarrow\; V_{OUT}=-H_2(s)V_E \approx -\frac{V_E}{\alpha s^2} \;\Rightarrow\; V_E \approx -\alpha s^2 V_{OUT}$$

This is included as it is readily measured with a low noise null detector [2]. In more convenient form, with frequency in Hz: -

$$f << f_1 \;\Rightarrow\; \frac{V_E}{V_{OUT}} \approx \alpha\left(\frac{f}{f_1}\right)^2 \quad \text{with} \quad f_1=\frac{1}{2\pi\tau_1}$$

The closed loop transfer function is, in normalised form $(s = j\omega\tau_1)$ [1]: -

$$T_2(s)=-\frac{Z_2}{Z_1}\left(\frac{H_2(s)}{H_2(s)+\beta(s)}\right)=-\frac{Z_2}{Z_1}\left(\frac{s+1}{\alpha\beta(s)s^2+s+1}\right) \quad \text{with} \quad \beta(s)=1+\frac{Z_2}{Z_1}$$

The parameter $\beta(s)$ is the reciprocal of the feedback factor.

The term in brackets is the "*D*" factor [1]: $T_2(s)=-\frac{Z_2}{Z_1}D_2(s)$

At low frequency the "*D*" factor is very nearly one and closed loop accuracy is limited by the quality of the passive component Z_1 and Z_2.

With a little algebra the error term can be extracted: $T_2(s)=-\frac{Z_2}{Z_1}\left(1-\frac{\alpha\beta(s)s^2}{\alpha\beta(s)s^2+s+1}\right)$

1. Monograph: "High gain blocks".
2. Monograph: "Null detectors – the basics".

3. High accuracy amplifiers

3.1 The closed loop transfer function

For an inverting amplifier both feedback components are resistors. The reciprocal of the feedback factor does not vary with frequency and the analysis is simplified. In normalised form $(s = j\omega\tau_1)$: -

$$\beta = 1 + \frac{R_2}{R_1} \quad \Rightarrow \quad T_2(s) = -\frac{R_2}{R_1} D_2(s) \quad \text{with} \quad D_2(s) = \frac{s+1}{\alpha\beta s^2 + s + 1} \text{ and } \alpha = \frac{\tau_2}{\tau_1}$$

Stability depends on the value of the product $\alpha\beta$. A higher value results in lower phase margin (a higher resonant peak at the OPI normalised frequency). In practice there can also be a lower limit due to the limited bandwidth of the op-amps but the range is very wide for a type 1 HGB. A tolerable degree of peaking is about 60% (+4dB) and a spreadsheet model can be used to estimate a suitable value. A typical maximum value is $\alpha\beta = 1$

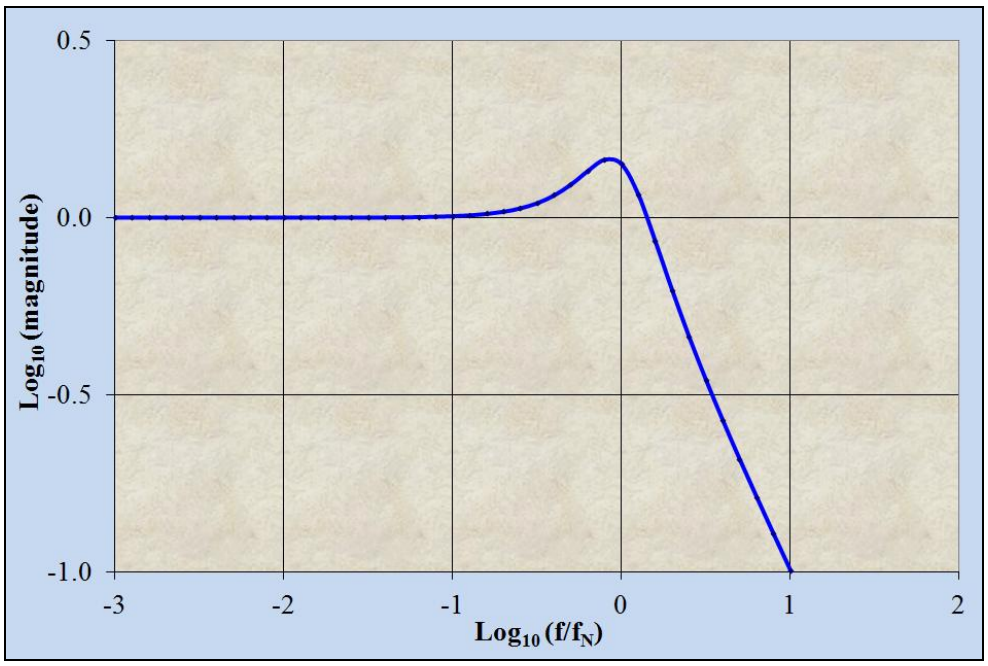


Fig.3.1.1 Closed loop response for $\alpha\beta = 1$ [1]

For a type 2 HGB this model works well, at low frequency, but high frequency stability depends on the type of op-amp and the capacitance of the load. With feedback factor approaching 100% there is usually a resonant peak in the frequency response at around the unity gain frequency. If the resonance is a problem one can easily reduce the feedback factor with a resistor across the HGB inputs or, with a type 2 HGB, setting the potentiometer a bit lower (see fig. 3.3.1).

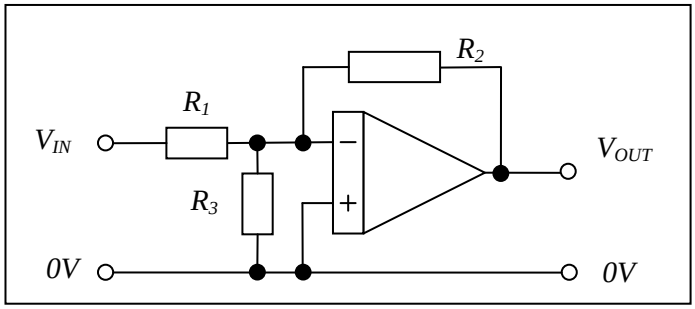


Fig. 3.1.2 Reduced feedback factor with R_3

1. Spreadsheet: "A two-stage HGB"

If one includes the very low and very high frequency characteristics of the op-amps the algebra gets a bit messy but a spreadsheet model [1] reveals the open loop characteristic. For a typical two-stage HGB, based on LF356 op-amps, the BoDayQuist plot illustrates how the stability criterion (avoiding the -1 point) is achieved: -

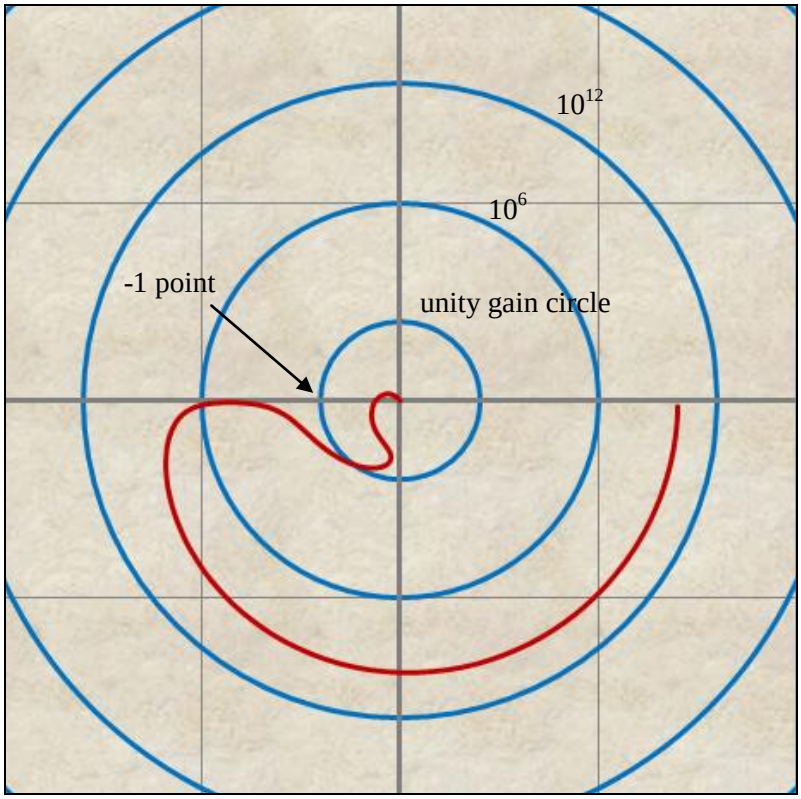


Fig. 3.1.3 BoDayQuist plot for $\alpha\beta = 1$ [1]

3.2 Error analysis

For a two-stage (type 1) HGB, in normalised form $(s = j\omega\tau_1)$ the low frequency model is: -

$$T_2(s) = -\frac{R_2}{R_1}\left(1 - \frac{\alpha\beta s^2}{\alpha\beta s^2 + s + 1}\right) \quad \text{with} \quad \alpha = \frac{\tau_2}{\tau_1}$$

We need only the real (in-phase) and imaginary (quadrature) error components. At low frequency: -

$$\alpha\beta|s|^2 << 1 \Rightarrow \quad D_2(s) \approx 1 - \frac{\alpha\beta s^2}{1+s} \approx 1 - \alpha\beta s^2(1-s) = 1 - \alpha\beta s^2 + \alpha\beta s^3$$

In more convenient form with frequency in Hz $(f_1 = 1/2\pi\tau_1)$: -

$$f << f_1 \quad \Rightarrow \quad D_2(s) \approx 1 + \alpha\beta\left(\frac{f}{f_1}\right)^2 - j\alpha\beta\left(\frac{f}{f_1}\right)^3$$

The quadrature error is now third order with respect to frequency – compared to first order for a single op-amp. The reader may notice that the error analysis has the same form as a two-stage low pass filter [2].

1. Spreadsheet: "Multi two-stage HGB"
2. Monograph: "Two-stage filters"

3.3 Example calculation

A circuit for measuring tanδ provides a useful example [1]. The HGB (type 2) is based on a dual op-amp (LF353) with 4MHz gain-bandwidth product: -

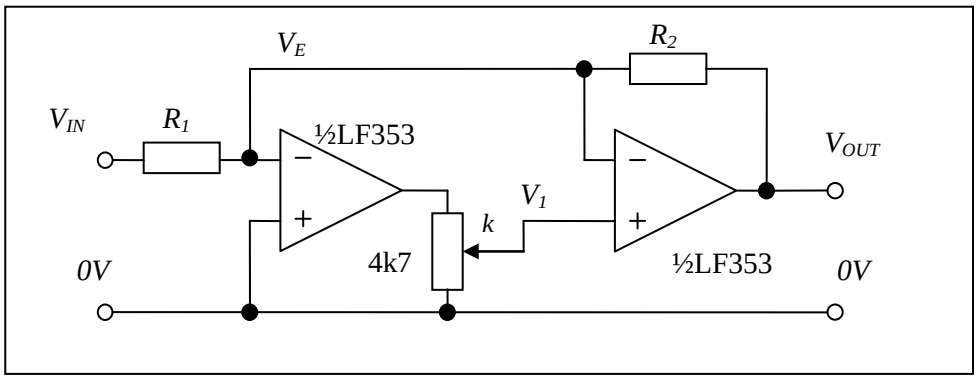


Fig. 3.3.1 A simple inverting amplifier

The open loop transfer function of a two-stage HGB is: -

$$H_2(s)=\left(1+\frac{1}{\tau_1 s}\right)\frac{1}{\tau_2 s}$$

The gain of the first op-amp is reduced with a trimpot by a factor 0.25 so that the GBWP (and the OPI frequency) is reduced to 1MHz. The optimum potentiometer setting was found with a step response (square wave input) on the actual circuit.

$$\tau_1=\tau_B/k\approx 1.6\times10^{-7}s \quad \text{and} \quad \tau_2=\tau_B=\frac{1}{2\pi f_B}\approx 4\times10^{-8}s \quad \text{and} \quad \alpha=\frac{\tau_2}{\tau_1}=k=0.25$$

$$R_1=R_2=4k7 \;\Rightarrow\; \beta(s)=1+\frac{R_2}{R_1}=2 \;\Rightarrow\; \alpha\beta=0.5$$

This is a factor of two lower than what one would predict for a reasonably stable circuit. The lower value, however, improves high frequency stability by reducing the loop gain.

For a two-stage HGB: $$D_2(s)\approx 1+\alpha\beta\left(\frac{f}{f_1}\right)^2-j\alpha\beta\left(\frac{f}{f_1}\right)^3$$

The frequency f_1 corresponds to the time constant τ_1 (i.e. 1MHz).

At 1kHz the ratio $f/f_1\approx10^{-3}$ so that the errors, both in-phase and quadrature, are negligible compared to the capacitor tanδ being measured (≈10^{-4}): -

$$f=1kHz \text{ and } f_1=1MHz \;\Rightarrow\; D_2(s)\approx 1+5\times10^{-7}-j5\times10^{-10}$$

1. Monograph: “A circuit for measuring tanδ”.

4. High accuracy integrators

4.1 The closed loop transfer function

For an integrator the feedback components are a resistor and capacitor.

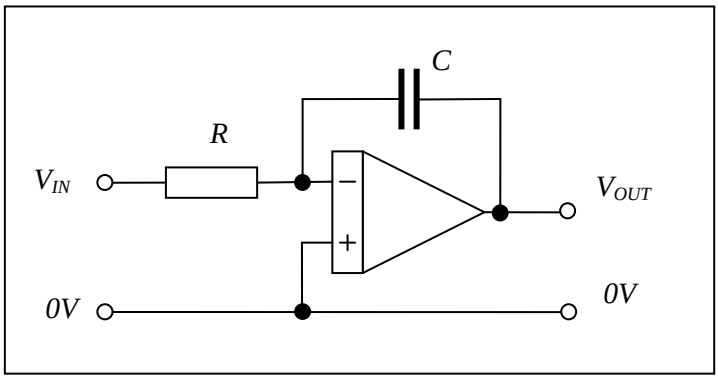


Fig. 4.1.1 Circuit diagram of a high accuracy integrator

$$Z_2 = \frac{1}{sC} \quad \text{and} \quad Z_1 = R \quad \Rightarrow \quad \beta(s) = 1 + \frac{1}{\tau_I s} \quad \text{with} \quad \tau_I = RC$$

From above the closed loop transfer function is: $T_2(s) = -\frac{1}{\tau_I s}\left\{\frac{H_2(s)}{H_2(s)+\beta(s)}\right\}$

In normalised form $(s = j\omega\tau_1)$: $\beta(s) = 1 + \frac{\tau_1}{\tau_I s}$

The closed loop transfer function is, in normalised form $(s = j\omega\tau_1)$: -

$$T_2(s) = -\frac{1}{\tau_I s}\left(\frac{s+1}{\alpha\beta(s)s^2 + s + 1}\right) \quad \text{with} \quad \alpha = \frac{\tau_2}{\tau_1}$$

$$\Rightarrow \quad T_2(s) = -\frac{1}{\tau_I s}\left(\frac{s+1}{\alpha s^2 + (1+\delta)s + 1}\right) \quad \text{with} \quad \delta = \alpha\frac{\tau_1}{\tau_I} << 1$$

In practice the integrator time constant is much higher than the HGB time constant and the effect on the "*D*" factor is negligible (in terms of peaking) compared to the inverting amplifier.

$$D_2(s) = \frac{s+1}{\alpha s^2 + (1+\delta)s + 1}$$

The result is a much reduced phase error compared to an integrator based on a single op-amp. The analysis is repeated here for convenience [1]. The open loop transfer function, except at very high frequency, is: -

$$\omega << \omega_B \quad \Rightarrow \quad H(s) = \frac{G}{1+\tau_P s} \quad \text{with} \quad \tau_P = G\tau_B$$

1. Monograph: "High gain blocks". See section 3.2

The gain-bandwidth product of the op-amp is: $$f_B = \frac{1}{2\pi\tau_B}$$

With a little algebra the result is of the form: $$T(s) \approx -\frac{1}{\tau'_I s}\left(\frac{G\tau'_I s}{(1+G\tau'_I s)(1+\tau'_B s)}\right)$$

Where: -

$\tau'_I = \tau_I\left(1+\frac{\tau_B}{\tau_I}\right)$ This indicates a slightly increased time constant of the integrator, compared to the ideal.

$\tau'_B = \tau_B\left(1+\frac{\tau_B}{\tau_I}\right)^{-1}$ This indicates a slightly reduced time constant compared to the gain-bandwidth product.

The first effect could be easily mitigated with a corresponding reduction in the value of the resistor.

In more convenient form, with frequency in Hz: -

$$\frac{f_I}{G} << f << f_B \;\Rightarrow\; T(s) \approx \frac{-1}{\tau_I s}\left(1+ j\left(\frac{f_I}{Gf} - \frac{f}{f_B}\right)\right)$$

The "D" factor has a second order band-pass characteristic [1] with a wide bandwidth: -

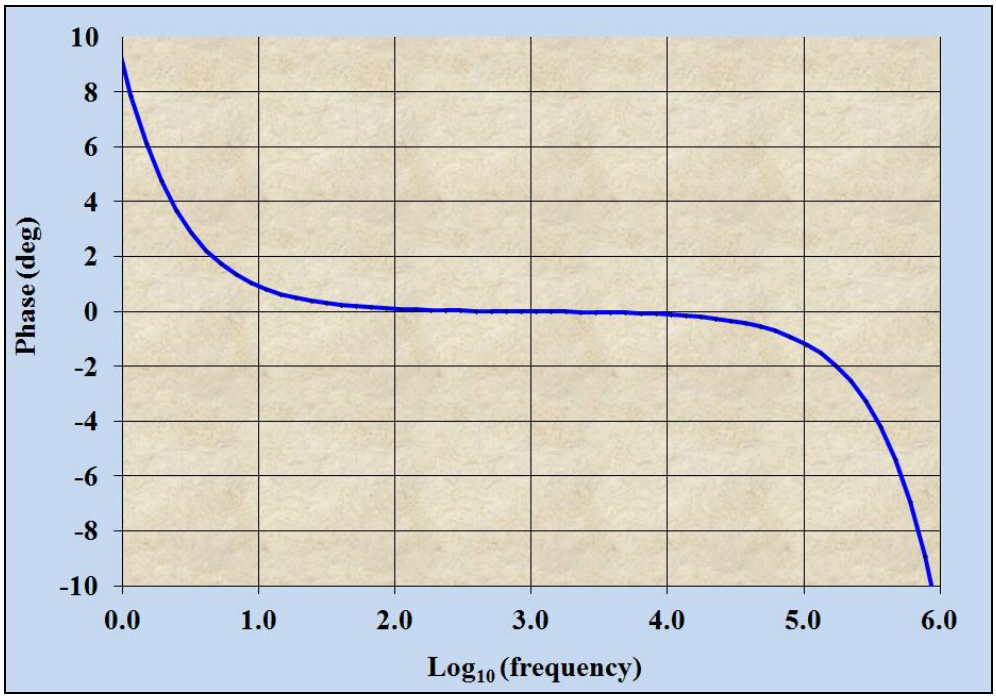


Fig. 4.1.2 Typical phase error of an integrator based on a single op-amp [1]

The result is a positive phase error at low frequency and a negative phase error at high frequency. The phase error is zero only at a mid frequency, typically around 1kHz [2].

With a two-stage HGB the phase error is considerably reduced.

1. Spreadsheet: "Band-pass filter"
2. Monograph: "High gain blocks". See section 3.2

4.2 Error analysis

For a two-stage HGB the "*D*" factor is, from section 2, in normalised form $(s = j\omega\tau_1)$: -

$$D_2(s) = 1 - \frac{\alpha\beta(s)s^2}{\alpha\beta(s)s^2 + s + 1}$$

$$\beta(s) = 1 + \frac{\tau_1}{\tau_I s} \;\Rightarrow\; D_2(s) = 1 - \frac{\alpha s^2 + \delta s}{\alpha s^2 + (1+\delta)s + 1} \text{ with } \delta = \alpha\frac{\tau_1}{\tau_I}$$

The integrator time constant is usually much greater than the OPI time constant so that: $\delta << \alpha$

We need only retain the significant real (in-phase) and imaginary (quadrature) error components. The effect of the s^2 term in the denominator is negligible: -

$$|\alpha s^2| << |s| \;\Rightarrow\; D_2(s) \approx 1 - \frac{\alpha s^2 + \delta s}{(1+\delta)s + 1} = 1 - (\alpha s^2 + \delta s)(1 - (1+\delta)s)$$

To a very good approximation, therefore: -

$$\delta << \alpha \;\Rightarrow\; D_2(s) \approx 1 - \alpha s^2 - \delta s + \alpha s^3$$

In more convenient form, with frequency in Hz: -

$$\Rightarrow \qquad D_2(f) \approx 1 + \alpha\left(\frac{f}{f_1}\right)^2 - j\left\{\delta\left(\frac{f}{f_1}\right) + \alpha\left(\frac{f}{f_1}\right)^3\right\}$$

4.3 Example calculation

As with the previous case a circuit for measuring tanδ provides a useful example [1]. The HGB time constants are the same (LF353 op-amps): -

$$\tau_1 \approx 1.6\times10^{-7}s \text{ and } \tau_2 \approx 4\times10^{-8}s \;\Rightarrow\; \alpha = \frac{\tau_2}{\tau_1} = 0.25$$

The integrator time constant is (4k7 and 10nF): $\tau_I = R_I C_I \approx 4.7\times10^{-5}s \;\Rightarrow\; \delta = \alpha\frac{\tau_1}{\tau_I} \approx 2.1\times10^{-4}$

$$D_2(f) \approx 1 + \alpha\left(\frac{f}{f_1}\right)^2 - j\left\{\alpha\left(\frac{f}{f_1}\right)^3 + \delta\left(\frac{f}{f_1}\right)\right\}$$

At 1kHz the ratios are: $f / f_1 \approx 10^{-3}$ so that the errors, both in-phase and quadrature, are negligible compared to the capacitor tanδ being measured (≈10^{-4}).

$$f = 1kHz \text{ and } f_1 = 1MHz \;\Rightarrow\; D_2(f) \approx 1 + 2.5\times10^{-7} - j(2.5\times10^{-10} + 2.1\times10^{-7})$$

1. Monograph: "A circuit for measuring tanδ". See section 4.

5. A high accuracy differentiator

A convenient method of implementing a differentiator is to employ a low phase error transformer primary winding as the feedback element in an inverting amplifier configuration. The action of feedback is to convert the input voltage to a current through the primary. The output voltage is proportional to the rate of change of the flux/current and the mutual inductance of the transformer (Faraday's law). It is also, therefore, proportional to the rate of change of the input voltage (i.e. differentiation). The resistance of the primary winding is relatively unimportant as the current is accurately in-phase with the input voltage. The main application is as part of a quadrature servo with a mutual inductance of typically 100μH [1].

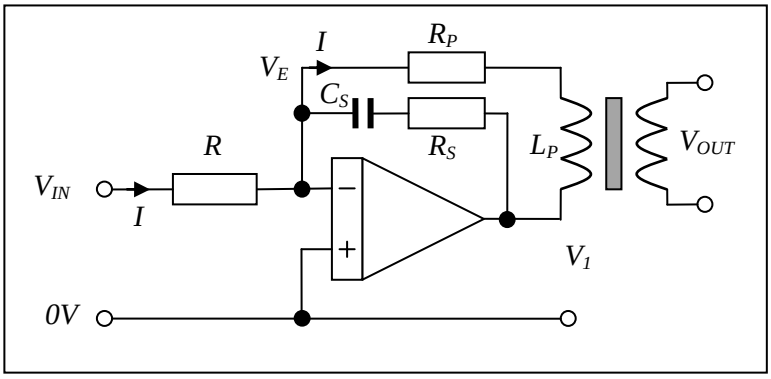


Fig.5.1 A differentiator circuit

The output is isolated from local 0V and can be easily added to the output of a main ratio transformer, as required in a typical quadrature servo.

The purpose of the series resistor/capacitor snubber in parallel with the transformer primary is to limit the gain at high frequency. Even a small amount of distortion, noise or interference at the input could threaten to overload the bridge null detector pre-amplifier. The snubber capacitor can resonate with the primary winding inductance but the peak value is limited by the damping effect of R_S. Typical component values are 1μF and 27Ω.

In practice the best available ferrite material (e.g. MMG type P12) has a phase error of the order 0.2mrad [2] and, therefore, the phase error due to the HGB, at low frequency, is negligible. Fortunately this is sufficiently accurate (and cost-effective) for most quadrature servo applications. The in-phase error is irrelevant as the servo automatically adjusts the amount of quadrature injected for a null balance.

A quadrature servo based on a low phase error capacitor and resistor would be possible but it would also require a low phase error isolating transformer (i.e. two-stage) and be less cost-effective.

5.1 The closed loop transfer function

The basic analysis (omitting R_S and C_S) is fairly simple. If one assumes an ideal HGB, with perfect virtual earth, the current passing through the input resistor and transformer primary is: -

$$R_S = \infty \quad \Rightarrow \quad I = \frac{V_{IN}}{R}$$

The flux in the transformer core is [3]: $\Phi = N_P I A_L$

Where: $N_P =$ Number of turns on the primary and $A_L =$ Permittance of the magnetic circuit (reciprocal of reluctance).

1. Monograph: "Null detectors – the basics". To explain 100μH see section 6.4.
2. Measured at low flux density (B < 1mT). Higher flux density results in increasing tanδ.
3. Monograph: "Single-stage IVDs and RTs". See section 2.

The output of the transformer is, according to Faraday's law: -

$$V_{OUT} = N_S \frac{d\Phi}{dt} = \frac{N_S N_P A_L}{R} \frac{dV_{IN}}{dt}$$

In the complex representation $(s = j\omega \equiv d/dt)$ the transfer function is: $T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{sL_M}{R}$

Where: $L_M = N_S N_P A_L =$ Mutual inductance of the transformer

In practice the inductance is not perfect but one can employ the complex model [1]: $L_M \rightarrow L_M(1 + j\tan\delta)$

The choice of input resistance is limited by the maximum current that can be delivered by the HGB. A maximum of 1mA is a reasonable compromise resulting in manageable power dissipation in the output stage and low flux density in the core. With a maximum input voltage of 10V a suitable value is, therefore, 10kΩ.

5.2 Error analysis

From the general analysis (section 2) the open loop transfer function and "*D*" factor for a two-stage HGB is, in normalised form $(s = j\omega\tau_1)$: -

$$H_2(s) = \frac{(1+s)}{\alpha s^2} \quad \text{and} \quad D_2(s) = \frac{H_2(s)}{H_2(s) + \beta(s)} \quad \text{with} \quad \alpha = \frac{\tau_2}{\tau_1} \text{ and } \beta(s) = 1 + \frac{Z}{R}$$

Z is the complex impedance of the entire feedback network (transformer primary and parallel snubber). The transfer function to the output of the HGB is, therefore (referring to fig. 5.1): -

$$\frac{V_1}{V_{IN}} = \frac{-Z}{R} D_2(s) \quad \text{with "virtual earth" voltage } V_E = \frac{-V_1}{H_N(s)}$$

The total current through Z is: -

$$I = \frac{V_E - V_1}{Z} = \frac{-V_1}{Z}\left(1 + \frac{1}{H_2(s)}\right) = \frac{-V_{IN}}{R}\left(1 + \frac{1}{H_2(s)}\right) D_2(s)$$

$$\Rightarrow \quad I = \frac{-V_{IN}}{R}\left(1 + \frac{1}{H_2(s)}\right)\left(\frac{H_2(s)}{H_2(s) + \beta(s)}\right)$$

$$\Rightarrow \quad I = \frac{-V_{IN}}{R}\left(\frac{H_2(s) + 1}{H_2(s) + \beta(s)}\right)$$

The complex representation for a mutual inductor is: $V_{OUT} = IsL_M$

The transfer function to the transformer output is, therefore $(s = j\omega)$: -

$$T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{sL_M}{R}\left(\frac{H_2(s) + 1}{H_2(s) + \beta(s)}\right)$$

1. Monograph: "Low phase error capacitors and inductors". See section 3.

At low frequency the HGB gain is very high and the term in brackets is very nearly 1. The feedback impedance is very low (the transformer primary has very low impedance and the snubber has negligible effect).

$$\omega\tau_1 << 1 \;\Rightarrow\; |Z| << R \;\Rightarrow\; |\beta(s)| \approx 1 \quad \text{and} \quad |H_N(s)| >> 1$$

$$\Rightarrow \quad \frac{H_2(s)+1}{H_2(s)+\beta(s)} = \left(1+\frac{1}{H_2(s)}\right)\left(1+\frac{\beta(s)}{H_2(s)}\right)^{-1} \approx \left(1+\frac{1}{H_2(s)}\right)\left(1-\frac{\beta(s)}{H_2(s)}\right) \approx 1+\frac{1-\beta(s)}{H_2(s)}$$

To a very good approximation, therefore: $1-\beta(s) = -\frac{Z}{R} \quad \Rightarrow \quad T(s) \approx \frac{sL_M}{R}\left(1-\frac{Z}{R}\frac{1}{H_2(s)}\right)$

The impedance of the feedback network is: $$Z = \frac{\left(\frac{1}{sC_S}+R_S\right)(R_P+sL_P)}{\left(\frac{1}{sC_S}+R_S\right)+(R_P+sL_P)}$$

Multiply top and bottom by sC_S: $$Z = \frac{(1+sC_sR_S)(R_P+sL_P)}{1+sC_S(R_S+R_P)+s^2C_SL_P}$$

The resistance of the primary winding (easily < 0.1Ω) is much smaller than the snubber resistance so that, at a frequency well below resonance, The feedback impedance is only the transformer primary winding (resistance plus inductance). The impedance Z is also much lower than the input resistance: -

$$\left|s^2C_SL_P\right| << 1 \quad \text{and} \quad R_S+R_P \approx R_S \;\Rightarrow\; Z \approx R_P+sL_P \quad \text{and} \quad \left|\frac{Z}{R}\right| << 1 \;\Rightarrow\; \left|\frac{Z}{R}\frac{1}{H_2(s)}\right| <<< 1$$

In practice, therefore, the error due to the snubber and HGB is truly negligible compared to the phase error of the ferrite transformer. **One could easily justify employing a single (good quality) op-amp!** An accurate model is: -

$$L_M \rightarrow L_M(1+j\tan\delta) \quad \Rightarrow \quad T(s) \approx \frac{sL_M}{R}(1+j\tan\delta)$$

5.3 Example calculation

A typical transformer has a mutual inductance of 100μH with a winding resistance of <0.1Ω. With a snubber capacitance of 1μF the natural frequency is: -

$$f_N = \frac{1}{2\pi\sqrt{L_PC_S}} \approx 16kHz$$

A snubber resistance of 27Ω provides sufficient damping with a characteristic frequency of: -

$$f_S = \frac{1}{2\pi R_SC_S} \approx 6kHz$$

At high frequency the snubber resistance becomes the dominant feedback component and the gain levels out: -

$$2\pi fL_P > 27\Omega \quad \Rightarrow \quad \frac{V_1}{V_{IN}} \approx \frac{27}{10^4} = 2.7\times10^{-3}$$

The maximum operating frequency is almost certainly less than 1.6kHz so that the maximum impedance of the transformer primary is at most 0.1Ω, compared to an input resistance of 10kΩ. The open loop gain of even a basic op-amp is at least 1000: -

$$f < 1.6kHz \;\Rightarrow\; |Z| \approx 2\pi fL < 0.1\Omega \;\Rightarrow\; \frac{|Z|}{R} < 10^{-4} \;\Rightarrow\; \left|\frac{Z}{R}\frac{1}{H_1(s)}\right| < 10^{-7}$$

Frequency response with component values: $L_P = 100\mu H \quad R_P = 0.1\Omega \quad R_S = 27\Omega \quad C_S = 1\mu F$

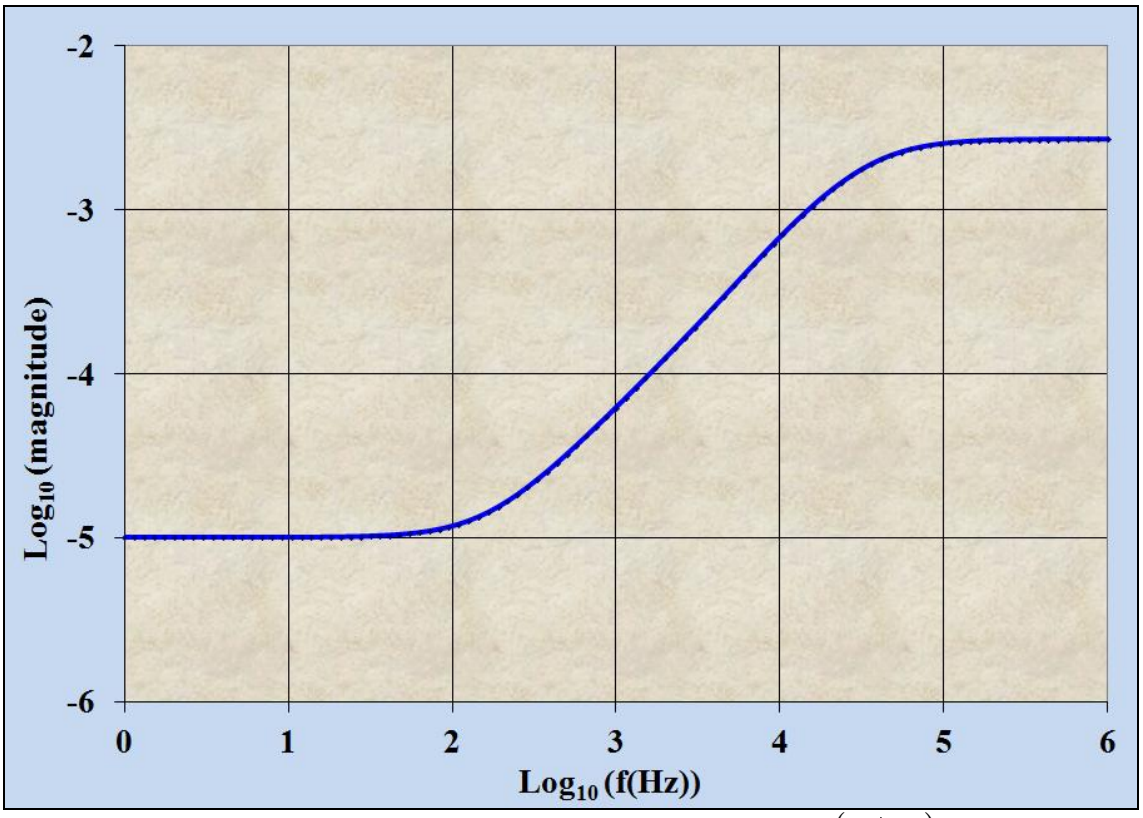


Fig. 5.3.1 Frequency response to the output of the HGB (V_1/V_{IN}) [1]

It is necessary to check that the flux density is low enough for the assumption about tanδ. With 16 turns and a maximum current of 1mA the flux is: -

$$\Phi = N_P I A_L = 16 \times 10^{-3} \times 400 \times 10^{-9} = 6.4 \times 10^{-9} Wb$$

The RM10 pot core has a magnetic circuit with an effective area of 83mm^2 so that the flux density is: -

$$B = \frac{\Phi}{A_E} = \frac{6.4 \times 10^{-9}}{83 \times 10^{-6}} = 77\mu T$$

This is well below the recommended test condition for tanδ (<1mT) for this type of core.

1. Spreadsheet: “A differentiator”.

6. A high accuracy differential amplifier

One form of differential amplifier consists of a high gain block (HGB) and four resistors. Low phase error can be achieved with a two-stage HGB. High common mode rejection ratio (CMRR), however, requires very accurately matched resistor pairs.

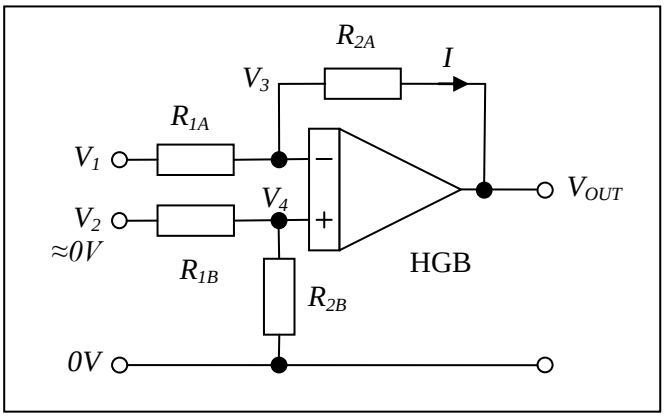


Fig. 6.1 A differential amplifier

Frequently one of the inputs is at approximately 0V and the matching requirement can be much reduced. One notable application is as part of a quadrature servo [1]. The voltage on one side of the standard resistor in a bridge is often no more than a few mV (see fig. 6.2.1). It is then possible to employ a high accuracy differential amplifier as long as the nominally 0V side is connected to the non-inverting input. The action of feedback is to drive the inverting input to be the same as the non-inverting input and the CMRR requirement is much reduced.

6.1 The closed loop transfer function

According to Ohm's law and Kirchhoff's laws, referring to fig. 6.1: $V_4 = \dfrac{R_{2B}}{R_{1B}+R_{2B}}V_2$

Similarly on the inverting side: $I = \dfrac{V_1 - V_3}{R_{1A}} = \dfrac{V_3 - V_{OUT}}{R_{2A}} \quad \Rightarrow \quad V_{OUT} = \dfrac{R_{2A}}{R_{1A}}(V_3 - V_1) + V_3$

For an ideal HGB the action of feedback is to cause: $V_3 = V_4$

$$V_{OUT} = \frac{R_{2A}}{R_{1A}}\left(V_2\frac{R_{2B}}{R_{1B}+R_{2B}} - V_1\right) + V_2\frac{R_{2B}}{R_{1B}+R_{2B}}$$

With a little algebra: $$V_{OUT} = \left(\frac{R_{2A}}{R_{1A}}+1\right)\left(\frac{R_{2B}}{R_{1B}+R_{2B}}\right)V_2 - \frac{R_{2A}}{R_{1A}}V_1$$

$$\Rightarrow \quad V_{OUT} = \left(\frac{R_{1A}+R_{2A}}{R_{1B}+R_{2B}}\right)\left(\frac{R_{2B}}{R_{1A}}\right)V_2 - \frac{R_{2A}}{R_{1A}}V_1$$

$$\Rightarrow \quad V_{OUT} = \left(\frac{R_{1A}+R_{2A}}{R_{1B}+R_{2B}}\right)\left(\frac{R_{2B}}{R_{2A}}\right)\left(\frac{R_{2A}}{R_{1A}}\right)V_2 - \frac{R_{2A}}{R_{1A}}V_1$$

1. Monograph: "Null detectors – the basics". See section 6.

We can take out the gain factor: $V_{OUT} = \frac{R_{2A}}{R_{1A}}\left(\left(\frac{k_A+1}{k_B+1}\right)V_2 - V_1\right)$ with $k_A = \frac{R_{1A}}{R_{2A}}$ and $k_B = \frac{R_{1B}}{R_{2B}}$

If the resistors are perfectly matched the result is, as expected: -

$$k_A = k_B \quad \Rightarrow \quad \left(\frac{k_A+1}{k_B+1}\right) = 1 \text{ and } \quad V_{OUT} = \frac{R_2}{R_1}(V_2 - V_1)$$

If one or the other inputs is 0V then one has either an inverting or non-inverting amplifier and it is clear that the dynamic response is the same (see section 2). The low feedback factor ensures that the open loop gain is well below 0dB at the extra pole frequency and the effect of the extra pole on closed loop stability is negligible.
For a two-stage HGB in normalised form $(s = j\omega\tau_1)$: -

$$T(s) = \frac{V_{OUT}}{V_2 - V_1} = \frac{R_2}{R_1}D_2(s) \quad \text{with} \quad D_2(s) = \frac{s+1}{\alpha\beta s^2 + s + 1}$$

$$f << f_1 \;\Rightarrow\; D_2(f) \approx 1 + \alpha\beta\left(\frac{f}{f_1}\right)^2 - \alpha\beta j\left(\frac{f}{f_1}\right)^3 \quad \text{with} \quad f_1 = \frac{1}{2\pi\tau_1}$$

6.2 Errors due to resistor tolerance

One can take the gain on the inverting side as the precise definition for an ideal amplifier: $G = \frac{R_{2A}}{R_{1A}} = \frac{1}{k_A}$

The gain is typically a minimum of 10 and a reasonably good approximation is: -

$$k_A < 0.1 \quad \Rightarrow \quad \left(\frac{k_A+1}{k_B+1}\right) \approx (1+k_A)(1-k_B) \approx 1 + k_A - k_B$$

The formula simplifies to: -

$$V_{OUT} = G((1 + k_A - k_B)V_2 - V_1)$$

The error, in terms of volts, can then be defined with: $\Delta V_{OUT} = G(V_2 - V_1) + \Delta V_{OUT}$

$$\Rightarrow \quad \Delta V_{OUT} = G(k_A - k_B)V_2$$

The error is proportional to the (small) input voltage on the inverting side and the mismatch of resistance ratios.

Define: $R_{1A} = R_1 \pm \Delta R_1$ etc $\Rightarrow \quad R_{1A} = R_1\left(1 \pm \frac{\Delta R_1}{R_1}\right) = R_1(1 \pm \delta_A)$ etc.

δ_A is the resistor tolerance in terms of ratio (e.g. 1%) on the inverting side. The effect on the resistance ratio is then $\pm 2\delta_A$. Consider the worst case (R_2 high and R_1 low): -

$$\delta_A << 1 \;\Rightarrow\; \frac{R_{2A}}{R_{1A}} = \frac{R_2(1+\delta_A)}{R_1(1-\delta_A)} \approx \frac{R_2}{R_1}(1+\delta_A)(1+\delta_A) \approx \frac{R_2}{R_1}(1+2\delta_A)$$

The effect on the error term $(k_A - k_B)$ is, therefore: -

$$k_A - k_B = k(1 \pm 2\delta_A) - k(1 \pm 2\delta_B) = \pm 2\delta_A k \pm 2\delta_B k$$

Each side contributes the same to the overall error so that one may as well specify the same tolerance.

$$\delta_A = \delta_B = \delta \quad \Rightarrow \quad \Delta V_{OUT} = \pm 4\delta V_2$$

The magnitude of the error, expressed as a ratio is, worst case: -

$$\left|\frac{\Delta V_{OUT}}{V_{OUT}}\right| = \frac{4\delta}{G}\frac{V_2}{(V_2 - V_1)}$$

Moreover, in the quadrature servo application [1] the purpose of the amplifier is to boost the signal developed across the reference resistor, R_S. The voltage V_2 is proportional to V_1 and the result is a gain error and not important as the following demonstrates: -

6.3 Example application and calculations

In a typical resistance bridge the voltage developed across the reference resistor, R_S, is used by the quadrature servo. The high accuracy voltage followers provide convenient low impedance sources for the inputs to the differential amplifier.

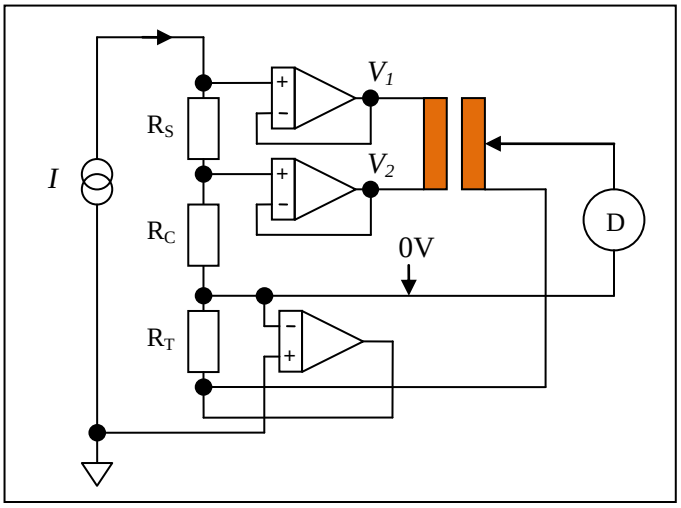


Fig. 6.3.1 A typical resistance bridge configuration

One of the potential leads of the unknown resistor, R_T, is held at local 0V by an active guard. The signal, V_2, on the lower potential lead of the reference resistor is determined, therefore, by the operating current and the lead resistance R_C.

According to Ohm's and Kirchhoff's laws: -

$$V_2 = IR_C \quad \text{and} \quad I = \frac{V_1 - V_2}{R_S} \quad \Rightarrow \quad V_2 = \frac{R_C}{R_S}(V_1 - V_2)$$

$$\Rightarrow \quad \Delta V_{OUT} = \pm 4\delta \frac{R_C}{R_S}(V_1 - V_2)$$

The result is a gain error:

$$\Rightarrow \quad V_{OUT} = \left(G \pm 4\delta \frac{R_C}{R_S}\right)(V_2 - V_1)$$

For this application low cost resistors (1% tolerance) should suffice.

1. Monograph: "Null detectors – the basics". See section 6.

The two-stage HGB required is slightly unusual (a hybrid type 3 with reversed order) [1]. The first stage is an open loop op-amp/integrator, required for its differential input. The second stage is a one-plus-integrator (OPI) with external feedback components for an accurate and stable dynamic response. Latch-up due to overload of the first stage is prevented by a limiter (a pair of zener diodes). The loop gain is reduced (and stability maintained) for the lowest gain setting (×10) with resistor R_3.

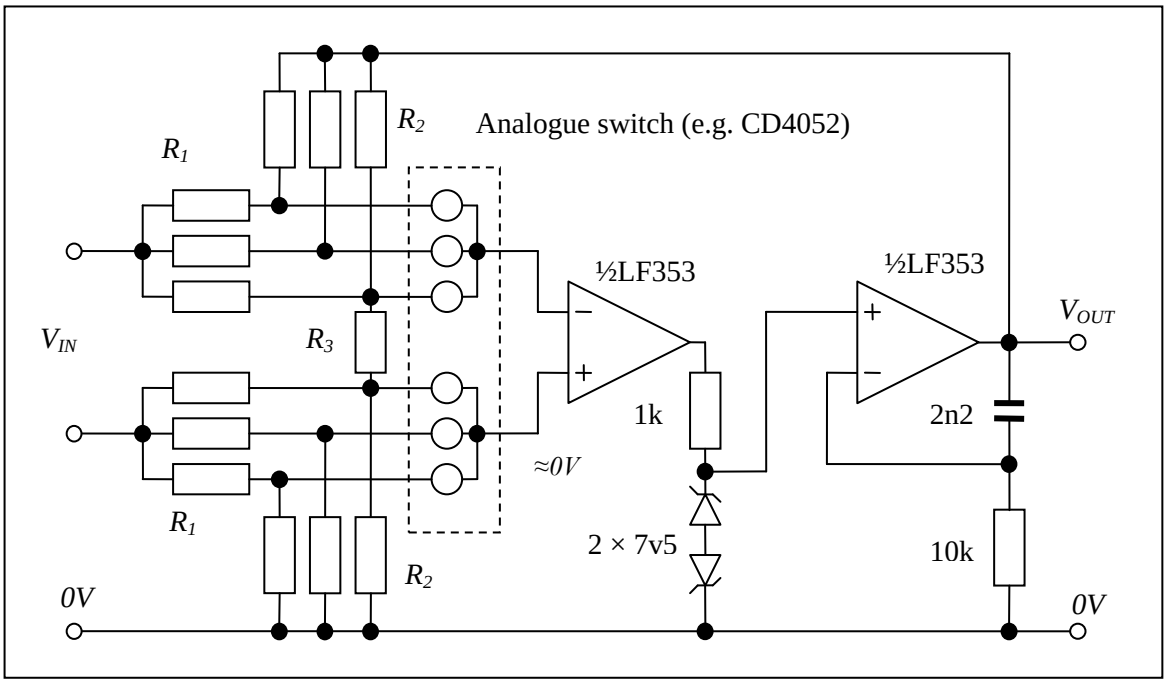


Fig. 6.3.2 A switched gain differential amplifier with two-stage HGB

The gain does not need to be accurate and 1% tolerance resistors are satisfactory: -

Gain	10	100	1000
R_1	*1kΩ*	*1kΩ*	*100Ω*
R_2	*10k*	*100kΩ*	*100kΩ*
R_3	*220Ω*	∞	∞

For the high gain settings the reciprocal of the feedback factor is very simply: $\beta = 1 + \frac{R_2}{R_1}$

The result is 101 and 1001 (gain + 1). Calculating the reciprocal feedback factor for the lowest gain setting is a little more complicated. Both inputs to the differential amplifier are connected to a low resistance source so that the equivalent circuit is: -

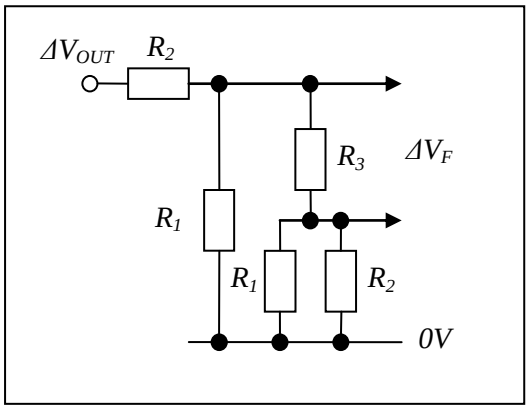


Fig. 6.3.2 The equivalent feedback network with R_3

1. Monograph: “High gain blocks”. See section 2.

The resistors have substantially different values and the feedback factor is not particularly critical so that one can rely on a very approximate calculation.

The first resistive divider R_1/R_2 is equivalent to a voltage source reduce by a factor of eleven in series with R_1 in parallel with $R_2(R_1 \parallel R_2)$. The overall circuit is, therefore, the reduced source voltage across the series combination of R_3 and two lots of parallel combinations $(R_1 \parallel R_2)$. The change in current through R_3 is, therefore:

$$\frac{R_1}{R_1+R_2}=\frac{1}{11} \quad \Rightarrow \quad \Delta I=\frac{\Delta V_{OUT}}{11}\div\left(R_1 \parallel R_2+R_3+R_1 \parallel R_2\right)$$

The parallel combinations are lower than R_1 by roughly 10% so adding R_3 (220Ω) brings the series combination back up to nominally $2R_1$. The change in voltage across R_3 is, therefore: -

$$\Delta I=\frac{\Delta V_{OUT}}{11}\div 2R_1 \;\Rightarrow\; \Delta V_F=\frac{\Delta V_{OUT}}{22R_1}R_3$$

The reciprocal feedback factor is, therefore: $\beta=\dfrac{\Delta V_{OUT}}{\Delta V_F}\approx\dfrac{22R_1}{R_3}\approx 100$

The "D" factor for a two-stage HGB is [1]: $D_2(s)=\dfrac{s+1}{\alpha\beta s^2+s+1}$

The product $\alpha\beta$ determines the stability margin. Spreadsheet modelling shows that a value between 0.2 and 2 is satisfactory [1]. The higher value results in a larger resonant peak. With a maximum reciprocal feedback factor of 1001, therefore, the value of $\alpha\approx\dfrac{2}{\beta}=\dfrac{2}{1001}$ is determined.

The integrator time constant for an LF353 with gain-bandwidth $(f_B=4MHz)$ is: $\tau_B=\dfrac{1}{2\pi f_B}=4\times10^{-8}s$

The ratio of time constants for stability is: $\alpha=\dfrac{\tau_B}{\tau_1}\approx 2\times10^{-3}$

The OPI time constant required for stable operation at the maximum gain is: $\tau_1\approx 2\times10^{-5}s$.

Convenient standard values are 2n2 and 10k: -

$$\tau_1=RC=2.2nF\times10k=2.2\times10^{-5}s \;\text{ so that }\; \alpha\beta=0.18 \;\text{ or }\; \alpha\beta=1.8$$

At low frequency the error analysis is: -

$$f<<f_1 \quad \Rightarrow \quad D_2(s)\approx 1+\alpha\beta\left(\frac{f}{f_1}\right)^2-j\alpha\beta\left(\frac{f}{f_1}\right)^3 \;\text{ with }\; f_1=\frac{1}{2\pi\tau_1}\approx 7.2kHz$$

At the higher operating frequency of 75Hz the ratio of is $\dfrac{f}{f_1}\approx 10^{-2}$ and the maximum quadrature error (gain = 1000) is about 2ppm and negligible.

1. Monograph: "High gain blocks". See section 2.

The Isotech microK "bridge"

1. Introduction

This monograph is included in part 4 because the microK "bridge" relies on a number of high accuracy active circuits – a high stability DC current source, a differential amplifier and a very accurate analogue to digital converter (ADC) [1]. The main application is resistance thermometry though the ADC can also measure the output of a thermocouple. Isotech offers five versions with resistance ratio accuracy, over the full range (0 – 1.05), from 0.5ppm to 30ppb [2]. The main advantage, compared to conventional null-balance bridges, is a major reduction in size and weight – no bulky ratio transformer. Project manager, Paul Bramley (PB), also claims higher reliability based on an inherently "drift free" and robust design – no internal adjustments or "moving parts" (e.g. "trimpots") "requiring regular service visits".

Fig.1.1 The MicroK "bridge" (picture courtesy Isothermal Technology Ltd [4])

The main disadvantage, theoretically, is the relatively poor noise performance for low source resistance (<10Ω) though PB does not agree [3], especially when one takes into account settling time when switching between input channels.

In later models (starting with the microK100 project) noise matching was improved with "a large array of (parallel) amplifiers" [4] resulting in a noise resistance of "a few hundred Ohms" [3]. A good clue is the bandwidth settings: 0.5Hz, 0.1Hz and 0.02Hz with measuring times of 2, 10 and 50s respectively. In the brochure for the microK "Gold" the stated resolution is 10nV (bandwidth not specified) [2].

Although described and marketed as a "bridge" the microK is basically a low frequency (5-10Hz) alternating DC instrument. PB insists, however: "..it really is very like a Wheatstone bridge..". You decide dear reader [5].

1. Bramley, P. and Pickering, J.: "Better Accuracy in Temperature Calibration and Measurement through a New Type of Analog-to-Digital Converter". The International Journal of Metrology (Cal. Lab. Magazine, pages 30-35, Oct 2006). (Paul Bramley is of Metrosol Ltd and John Pickering is of Metron Designs Ltd).
2. Isotech brochure: "The new microK family of precision thermometry bridges" (Edition 2: 1516), available at: www.isotech.co.uk
3. Isotech claim "low noise" but do not provide much information. See sections 6.1 and 6.14.
4. Bramley, P: "A few words from the microK100 design team" 2009, from www.metrosol-ltd.co.uk
5. See section 6.3.

2. The microK measurement system (overview)

2.1 Introduction

The measurement system is based on an ADC originally developed by (and licensed from) the National Physical Laboratory (NPL) and Metron Designs Ltd. Pickering (of Metron) and Georgakopoulos et al [1] describe it as a "delta-sigma" (Δ-Σ) converter but Bramley and Pickering [2] prefer "sigma-delta" (Σ-Δ). The "sigma" refers to "sum" (Greek letter Σ) and delta refers to difference (Greek letter Δ). This is a remarkable bit of mixed signal circuitry (analogue and digital) requiring average timing errors of less than 0.5ps (pico-second!), apparently [3].

The measuring range of the ADC is quoted as ±5V (not including the polarity reversal switch) [4]. The input voltage range (thermocouple mode) is quoted as ±125mV [3], hence the need for a differential amplifier with various gain settings.

2.2 Voltage (thermocouple) measuring mode

The connectors are gold plated tellurium copper alloy to keep thermal emfs to a minimum. Also, the polarity reversing switch has very low thermal emfs (solid state - MOSFET) mounted immediately behind the connectors [4]). Cold junction compensation (CJC) is external. Polarity reversal appears to be for thermocouple mode only – for eliminating the DC offset of the amplifier. This suggests that the ADC is inherently bipolar (i.e. the input can be positive or negative and the polarity switch is not used to achieve this [5].

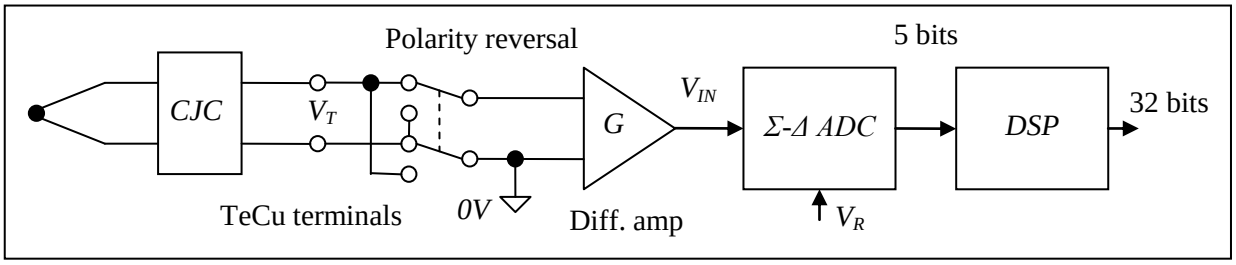


Fig. 2.2.1 Outline schematic of the measurement system (Thermocouple measuring mode)

X represents the result of a single A-D conversion and *Y* the calculated result. V_O is the amplifier offset. If one assumes that the gain/ADC, thermocouple, reference and offset voltages are stable between measurement cycles: -

$$X_1 = G\left(\frac{V_T + V_O}{V_R}\right) \qquad X_2 = G\left(\frac{-V_T + V_O}{V_R}\right) \Rightarrow \; Y = \frac{X_1 - X_2}{2} = G\frac{V_T}{V_R}$$

In thermocouple mode accuracy and noise performance are most likely limited by the reference voltage. Voltage reference devices are notoriously noisy unless one employs a very low frequency low-pass filter (with large, low leakage capacitors). A typical 5V reference (e.g. LT1021), for example, produces $3\mu V_{PK\text{-}PK}$ in the range 0.1 – 10Hz bandwidth. The microK employs "a very good buried zener reference" [6].

1. Pickering, J. R., Georgakopoulos, D., Williams, J.M. and Wright, P. S: "Effect of a PWM feedback DAC on the Noise and Linearity of a Delta-Sigma ADC", Int. Conf. on A to D and D to A Converters and Their applications, 2005.
2. Bramley, P. et al: "Better Accuracy in Temperature Calibration…." Available via Isotech website.
3. Isotech brochure: "The new microK family of precision thermometry bridges": "In order to achieve our target of <0.05ppm, we needed to be able to produce pulses whose edges have relative timing errors of 0.5ps (about the time it takes light or electrical signals to travel 0.15mm)."
4. Bramley, P and Robinson, N.: "Cost Effective Techniques used to Validate the Performance of the microK Resistance Thermometry Instrument with sub mK Uncertainty." Available via Isotech website.
5. PB confirmed. See section 6.16.
6. PB confirmed. See section 6.6.

It is also likely that one of the inputs of the differential amplifier is held at local 0V (rather than the thermocouple), to avoid an error due to gain asymmetry (c.f. common mode rejection). This can be achieved with either a direct connection or, more likely, with an active guard [1].

The microK brochure [2] quotes a DC voltage measuring accuracy of 0.25μV with a range setting of 20mV [3].

2.3 Resistance ratio measuring mode

In resistance measuring mode a DC current (0 – 10mA) is passed through the resistor and the voltage measured (relative to an internal reference: V_R). The current is then reversed and the measurement repeated. The difference is calculated in order to eliminate thermal emfs and the amplifier input offset voltage. The current is proportional to the internal reference voltage [4] so that each basic measurement is, in effect, a resistance ratio (relative to an internal reference resistor in the current generating circuit). Other ratios are then calculated from "substitution measurements" [5] ("switching at several times a second"). This requires the measuring current, differential amplifier gain and ADC to be sufficiently stable for the short time between substitutions. The differential amplifier and ADC also need to be very linear and stable between substitutions (typically <10ppb for the microK "gold").

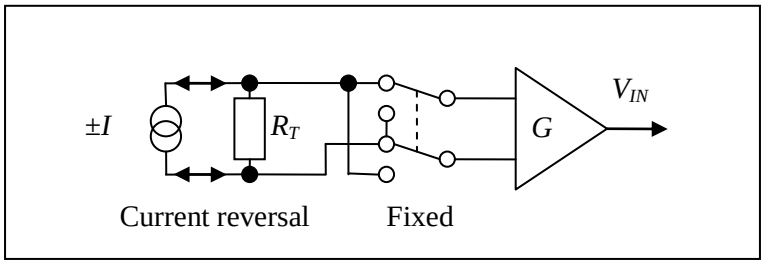


Fig. 2.3.1 The measuring current reversal scheme (resistance measuring mode)

It would seem (see next section) that exact current reversal is not practicable but this need not be a problem as long as the selected current is sufficiently stable. If one assumes that all parameters are stable between measurements: -

$$X_1 = G\left(\frac{I_1 R_T + V_O + V_{THERMAL}}{V_R}\right) \text{ and } X_2 = G\left(\frac{-I_2 R_T + V_O + V_{THERMAL}}{V_R}\right) \Rightarrow Y = X_1 - X_2 = G\frac{(I_1 + I_2)R_T}{V_R}$$

Once again the *Xs* represents the result of a single A-D conversion (probably 50 - 100ms per conversion) and *Y* the calculated result.

Fluctuations of the reference, offset and thermal emfs set a lower limit to the reversal frequency and an upper limit to each A-D conversion cycle (100ms for the original microK models [6]), equivalent to an operating frequency of 5Hz, though a more recent microK brochure quotes 6-10Hz [2]. This is a problem common to all so called "DC" instruments (e.g. the "Kusters comparator"). It is stated in [7], for example, with unintended irony: "The reversing cycle of the bridge should be as short as possible in order to be able to cope with changing thermal EMFs".

The various microK models have three channels for measuring resistance – typically one for the thermometer being calibrated and two for reference thermometers (SPRTs) or a single SPRT and an external "Wilkins" type transfer standard resistor.

1. PB confirmed. See section 6.7.
2. Isotech brochure: "The new microK family of precision thermometry bridges": (Edition 2: 1516)
3. See also: Bramley, P. and Robinson, N.: "Cost Effective Techniques used to Validate the Performance of the microK Resistance Thermometry Instrument with sub mK Uncertainty". www.isotech.co.uk
4. PB confirmed. See section 6.6.
5. Paul Bramley, P., Tavener, J. and Pickering, J.: "Using a Substitution Measurement Topology to eliminate the Effect of Common Mode Errors in Resistance Measurements used in Temperature Metrology" 2007 NCSL International Workshop and Symposium.
6. Bramley, P.: "A few words from the microK design team" 2006, from www.metrosol-ltd.co.uk
7. Kusters, N.L. and MacMartin, M. P.: "Direct-Current Comparator Bridge for Resistance thermometry". IEEE Trans. On Instrum. And Meas., vol. IM-19, No. 4, Nov 1970.

The measuring currents are proportional to a master internal reference voltage, V_{REF}, probably via a 12-bit multiplying digital to analogue converter (MDAC) [1], and inversely proportional to an internal reference resistor. The reference supplied to the ADC is also derived from the same master reference [1]: -

$$I_1 = k_1 \frac{V_{REF}}{R_5} \qquad I_2 = k_2 \frac{V_{REF}}{R_5} \text{ and } V_R = k_3 V_{REF} \Rightarrow Y = G\left(\frac{k_1 + k_2}{k_3}\right)\frac{R_T}{R_5}$$

The values k_1, k_2 and k_3 must also be sufficiently stable between A-D conversions. This makes sense – matched pairs of resistors and MDACs, with the required short term stability, are readily available and variations (low frequency noise and drift) in the voltage reference have a much reduced effect. It is also cited as the main reason for calling the microK a "bridge", which is rather strange [2].

To measure the ratio of two external resistors a further two measurement cycles (a total of four) are required – by the "substitution" method [2] (probably a total of 200 – 400ms per measurement): -

$$\frac{R_T}{R_T'} = \frac{Y}{Y'} = \frac{X_1 - X_2}{X_1' - X_2'}$$

This, again, requires sufficient stability between substitutions.

According to PB [3] the microK goes one better – taking three consecutive samples with the middle sample double weighted in the calculation. If one assumes that the total offset drifts at a constant rate (parameter α): -

$$V_O + V_{THERMAL} = V_0 + \alpha t$$

The samples are taken at uniformly timed intervals (ΔT) with $t = 0$ for the first sample. Once again the *Xs* represents the result of a single A-D conversion and *Y* the calculated result: -

$$X_1 = G\left(\frac{I_1 R_T + V_0}{V_R}\right) \qquad X_2 = G\left(\frac{-I_2 R_T + V_0 + \alpha\Delta T}{V_R}\right) \quad X_3 = G\left(\frac{I_1 R_T + V_0 + 2\alpha\Delta T}{V_R}\right)$$

$$\Rightarrow \quad Y = X_1 - 2X_2 + X_3 = 2G\frac{(I_1 + I_2)R_T}{V_R}$$

Using the substitution method, therefore, the ratio of two external resistors is: -

$$\frac{R_T}{R_T'} = \frac{Y}{Y'} = \frac{X_1 - 2X_2 + X_3}{X_1' - 2X_2' + X_3'}$$

This requires a minimum of six A-D conversion cycles (300 – 600ms per measurement). The process is then repeated a sufficient number of times and the average calculated, in order to achieve the required resolution and accuracy.

Clearly, higher order correction for drifting offsets is possible with more data. A useful software facility would be the ability to monitor and display a graph of the actual offsets.

1. PB confirmed. See sections 6.6 and 6.10.
2. Bramley, P. et al: "Using a Substitution Measurement Topology…." In section 4: "…the internal zener reference is used by both the current source and the ADC means that it is in fact a bridge system". See section 6.3 for more on this issue.
3. PB comment. See section 6.8.

3. The voltage controlled (ultra-constant) current source

This is an ingenious two-stage design, probably due to John Pickering of Metron Designs Ltd. It is described in Bramley et al [1] though not in much detail. The first stage is an "improved Howland current pump" based on an op-amp with positive and negative feedback [2]: -

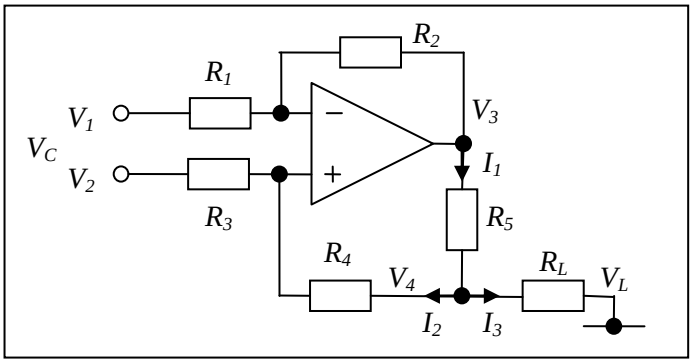


Fig. 3.1 The first stage current source

If one assumes an ideal op-amp then the action of feedback is to make both inputs the same voltage. According to Ohm's law and Kirchhoff's laws: -

$$V_- = V_1 + (V_3 - V_1)\frac{R_1}{R_1 + R_2} = V_+ = V_2 + (V_4 - V_2)\frac{R_3}{R_3 + R_4} \quad \Rightarrow \quad \frac{V_1 R_2 + V_3 R_1}{R_1 + R_2} = \frac{V_2 R_4 + V_4 R_3}{R_3 + R_4}$$

Also, the output current is: $$I_3 = I_1 - I_2 = \frac{V_3 - V_4}{R_5} - \frac{(V_4 - V_2)}{R_3 + R_4}$$

For a constant current source the output resistance is infinitely high (zero output conductance). One can assume, therefore, with V_1 and V_2 constant: -

$$\frac{\partial I_3}{\partial V_4} = \frac{\partial V_3}{\partial V_4}\frac{1}{R_5} - \frac{1}{R_5} - \frac{1}{R_3 + R_4} = 0 \quad \Rightarrow \quad \frac{\partial V_3}{\partial V_4} = 1 + \frac{R_5}{R_3 + R_4}$$

From above one also finds: $$\frac{V_1 R_2 + V_3 R_1}{R_1 + R_2} = \frac{V_2 R_4 + V_4 R_3}{R_3 + R_4} \quad \Rightarrow \quad \frac{\partial V_3}{\partial V_4}\frac{R_1}{R_1 + R_2} = \frac{R_3}{R_3 + R_4}$$

Both conditions are satisfied if: $$\frac{\partial V_3}{\partial V_4} = 1 + \frac{R_5}{R_3 + R_4} = \frac{R_3}{R_1}\left(\frac{R_1 + R_2}{R_3 + R_4}\right) \quad \Rightarrow \quad \frac{R_3}{R_4 + R_5} = \frac{R_1}{R_2}$$

The result has an elegant symmetry. A reasonable choice could be, for example: -

$$R_1 = R_2 \quad \Rightarrow \quad R_3 = R_4 + R_5$$

From the inverting side: $$R_1 = R_2 \qquad \Rightarrow \qquad V_- = V_1 + \frac{V_3 - V_1}{2} = \frac{V_1 + V_3}{2}$$

The action of feedback is to make the inputs of the ideal op-amp the same: $V_+ = V_- = \frac{V_1 + V_3}{2}$

1. Bramley, P. et al: "Using a Substitution Measurement Topology…"
2. Sheingold D. H.: "Impedance & Admittance Transformations using Operational Amplifiers". The Lightning Empiricist, Vol. 12, No. 1. (Jan 1964)

$$V_+ = V_2 + I_2R_3 = \frac{V_1+V_3}{2} \quad \text{and} \quad V_+ = V_3 - I_1R_5 - I_2R_4 = \frac{V_1+V_3}{2}$$

Add the last two equations: $V_2 + I_2R_3 + V_3 - I_1R_5 - I_2R_4 = V_1 + V_3$

From above and with a little algebra: $R_3 = R_4 + R_5 \quad \Rightarrow \quad I_3 = I_1 - I_2 = \frac{V_2 - V_1}{R_5} = \frac{V_C}{R_5}$

The first stage has a truly differential input – a good idea for any high accuracy circuit. It is probably the case that V_1 is connected to the control voltage source local 0*V*.

Clearly R_5 is a critical component, though the circuit performance (output resistance) also relies on accurate matching of the other resistors. Bramley et al estimate an output resistance of about 9MΩ [1], due to component tolerance (not specified), which they deemed insufficient and, therefore, a second "cascode" stage was added: -

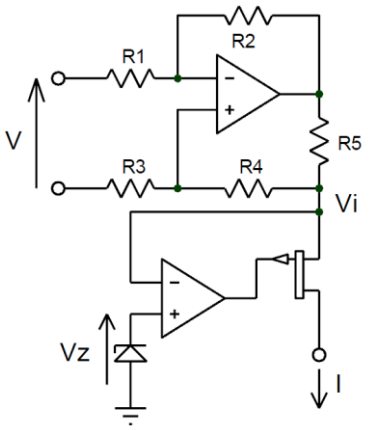


Figure 8: Current source with cascode stage.

Fig.3.2 The two-stage current source (courtesy Bramley et al [1])

This is a strange choice of name. The term "cascode" [2] usually refers to a two-stage high frequency amplifier [3]. It can be described as a voltage follower with a P-channel JFET source follower output stage (i.e. inside the loop). It is not clear what V_Z represents; A reasonable guess is that it is connected to V_2 (see fig. 3.3) so that the current flowing through R_3 and R_4 is zero. The load resistance, as "seen" by the first stage current regulator, then becomes the drain-source resistance of the JFET plus the actual load resistance in series.

According to the MicroK brochure [4] there are three range settings for measuring current: 0-100μA, 0.1-1mA and 1-10mA. A reasonable estimate for the range of the control voltage is 0 ± 5V. From the analysis above this can be achieved by selecting R_5 (50kΩ, 5kΩ and 500Ω, respectively). The switches do not need to be fast but "on" resistance and "off" leakage current are critical. For more detail see [5].

1. Bramley, P. et al: "Using a Substitution Measurement Topology…"
2. PB comment: See section 6.12.
3. According to Wikipedia "cascode" is a conflation of cascade and pentode, originally coined for a dual triode high frequency amplifier – grounded cathode followed by grounded grid, designed to eliminate the "Miller" capacitance effect. The name applies to similar transistor circuits, including the widely used dual gate MOSFET.
4. Brochure: "The new microK family of precision thermometry bridges" (Edition 2: 1516)
5. Part 4, monograph 6: "A high accuracy current source".

The actual circuit is likely to be, therefore: -

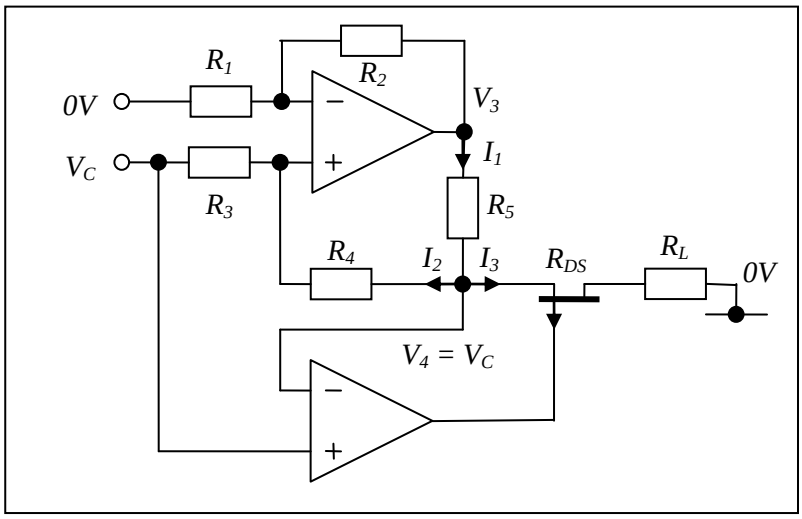


Fig. 3.3 A two-stage current regulator (including the "cascode")

The effect on the circuit is rather interesting: -

$$V_4 = V_C \;\Rightarrow\; I_2 = 0 \;\Rightarrow\; V_+ = V_- = V_C \;\Rightarrow\; V_3 = 2V_C \;\Rightarrow\; R_{DS} + R_L = R_5$$

The drain-source resistance of the P-channel JFET automatically adjusts so that the total load resistance $(R_{DS} + R_L)$ is precisely the same as the reference resistor R_5. The actual load resistance, R_L, may vary, within limits, but the current remains "ultra-constant" [1].

The P-type JFET works only for positive control voltage. To reverse the current there is probably some means of selecting an N-type JFET [2]. Also, one can think of the first stage current source as a very high voltage source, V_{HI}, with a very high source resistance, R_{HI}. The equivalent circuits, drawn in a familiar form are: -

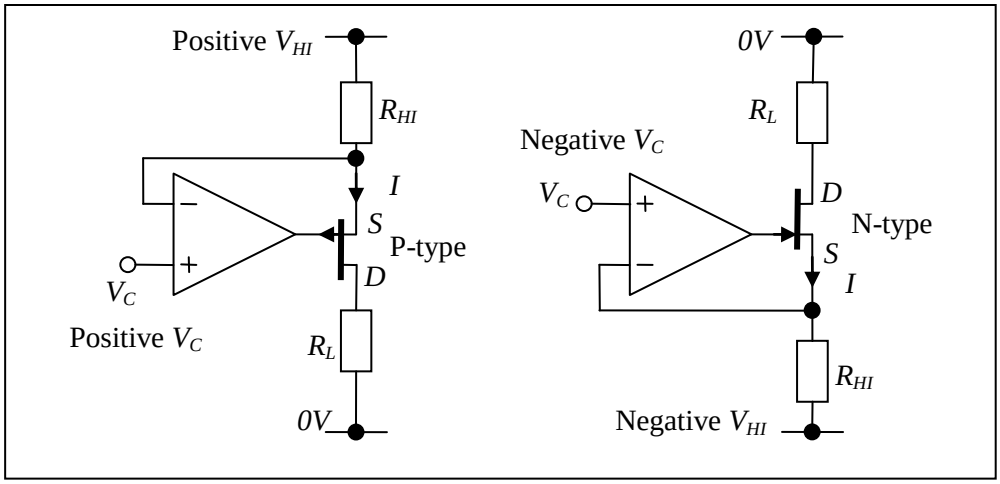


Fig. 3.4 The second stage is a voltage follower with a JFET source follower in the loop

There appears to be a problem at zero current – no loop gain. The circuit probably works down to 10% of each range setting (i.e. minimum current ±10μA).

1. Bramley, P. et al: "Cost Effective Techniques …". The Metron product (I-REF2) is quoted to have an output resistance of >10GΩ.
2. PB confirmed: See section 6.11

The specification of the top-of-the-range (MicroK "Gold") is 30ppb ratio accuracy over the whole measuring range (ratio of 0 - 1.05) [1]. A reasonable target for the current source is a contribution of no more than 10ppb. The timescale between substitutions is of the order of only 100ms but this is still a tall order. The noise level (well within the range of 1/f noise of most semiconductors) is also a major factor and probably contributes to the instrument's overall noise performance [2].

To achieve this level of performance it would be necessary to employ a very good (low noise) reference voltage source. The reference resistor, R_5, is also critical – it is probably necessary to employ a two-terminal-pair technique, especially for the lowest resistance value (500Ω). The voltage sensing connections (to resistors R_2 and R_4) need to be as close as possible to the component R_5 [3]. Each connection would need its own switch: a total of 12 for three values of resistor R_5.

According to Bramley et al "Although the above analysis shows that the inclusion of a cascode stage reduces the errors caused by the current source to insignificant proportions, the measurement system in the microK is also actively guarded to further improve performance" [4]. I presume this refers to a virtual earth: The resistor being measured is the feedback element of an op-amp (or a high gain block?) [5]: -

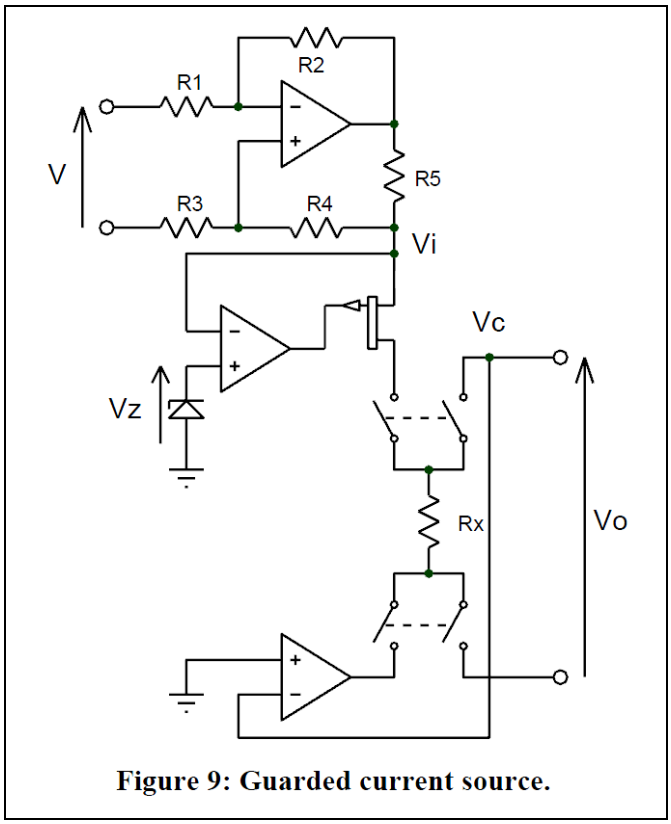


Figure 9: Guarded current source.

Fig. 3.5 The substitution method (courtesy Bramley et al [4])

The load resistance of the first stage current regulator is now the JFET drain-source resistance plus the resistance of the current carrying lead (plus the switch resistance) of the two-terminal-pair resistance, R_X, being measured. The latter is usually quite small $(R_L << R_5)$ so that $R_{DS} \approx R_5$ (approximately).

The guard amplifier ensures that the voltage sensing lead of R_X is accurately 0V all along its length (negligible current flows in the voltage sensing lead) hence eliminating common mode at the differential amplifier input.

1. Brochure: "The new microK family of precision thermometry bridges" (Edition 2: 1516)
2. I could find nothing in the published literature regarding the noise contribution due to the current source.
3. Part 1, monograph 1: "High accuracy resistors" See section 3.
4. Bramley, P. et al: "Using a Substitution Measurement Topology…". See section 3.3.2.
5. Part 4 monograph 1: "High gain blocks". Even better would be a two-stage high gain block (HGB type 1). See section 4.1.

The four switches for selecting R_X (fig. 3.5) allow the substitution method: The MicroK has three sets of terminals for external resistors (typically two resistance thermometers and a transfer standard resistor). Speed and "on" resistance are not critical but thermal EMFs and leakage current must be low or stable between substitutions.

The JFET adjusts to compensate for any change in R_L thus ensuring a very accurate constant current, at least in the short term, regardless of which resistor is selected. Fig. 3.5 is reproduced in a more convenient form: -

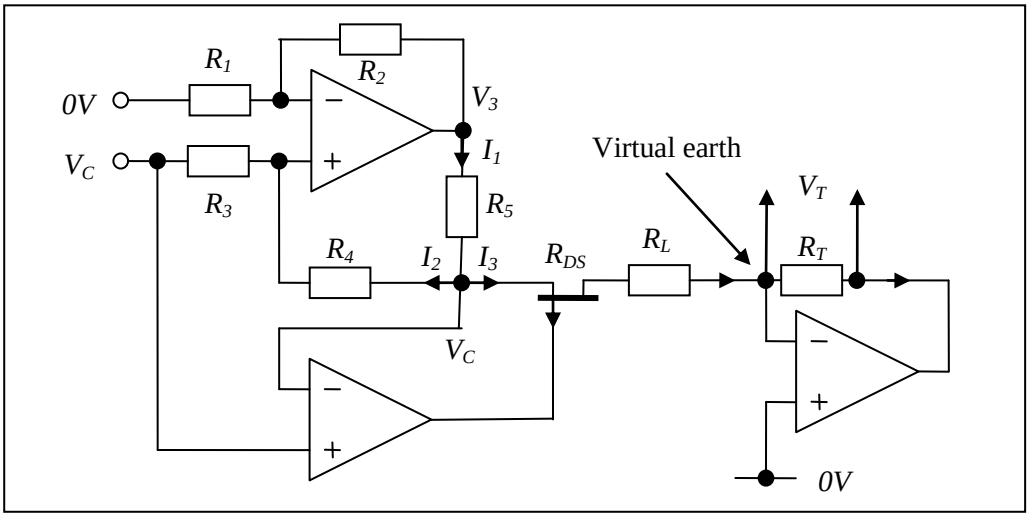


Fig. 3.6 Current regulator into a virtual earth (active guard)

The virtual earth is much better than a direct connection to earth [1].

If you are now scratching your head, dear reader, you are not the only one. **The following is an alternative approach.** The reference resistor, R_5, is more clearly a two-terminal-pair device: -

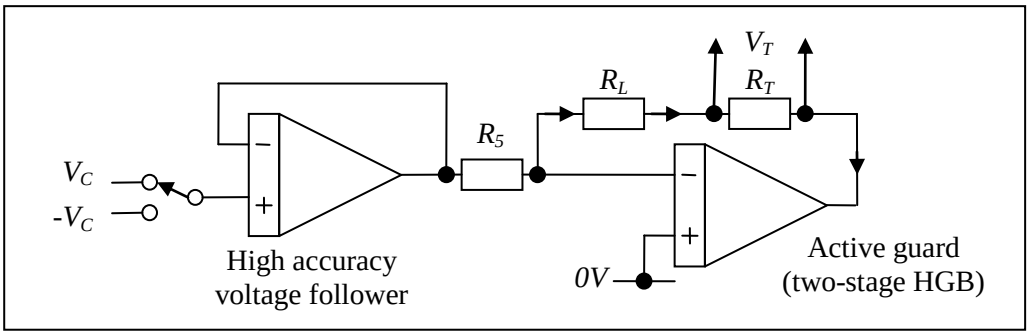


Fig. 3.7 An alternative current source

The lead resistance plus switch resistance, R_L, is shown explicitly to emphasise the fact that one of the potential leads of R_T is no longer held accurately at 0V. PB points out, correctly, that this would introduce common mode at the input of the differential amplifier [2]. Later, however, he confirms that the differential amplifier is "floating" [3], at least for more accurate models, by which I presume he means the power supply is bootstrapped.

It would seem to be a key feature of the microK design that common mode must be avoided – hence the "ultra-constant" current source. PB: "It's unavoidable as far as I can see" [2].

A bootstrapped power supply with high accuracy voltage follower offers an alternative….

1. Part 3, monograph 6: "An F17 type ratio transformer bridge". See section 1.2.
2. PB feedback: See section 6.13.
3. PB comment: See sections 6.7 and 6.15.

4. The differential amplifier

Very little information could be found on the microK differential amplifier hence the following includes **more speculation**. In later models better noise matching (lower noise voltage/ resistance) is achieved with "a large array" of non-inverting amplifiers in parallel, apparently [1]: -

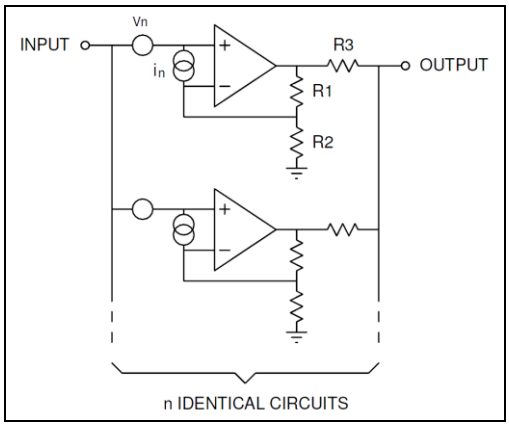


Fig. 4.1 Noise matching by amplifiers in parallel (Courtesy PB [1])

This is rather strange as both inputs need to present a very high input resistance. The basic idea is to reduce voltage noise at the expense of increasing current noise (reduced noise resistance). The number and type of amplifiers and noise resistance are not specified.

PB feedback confirms that the actual design is "considerately more complicated than disclosed" [2]. A likely candidate is a fully differential instrumentation amplifier: -

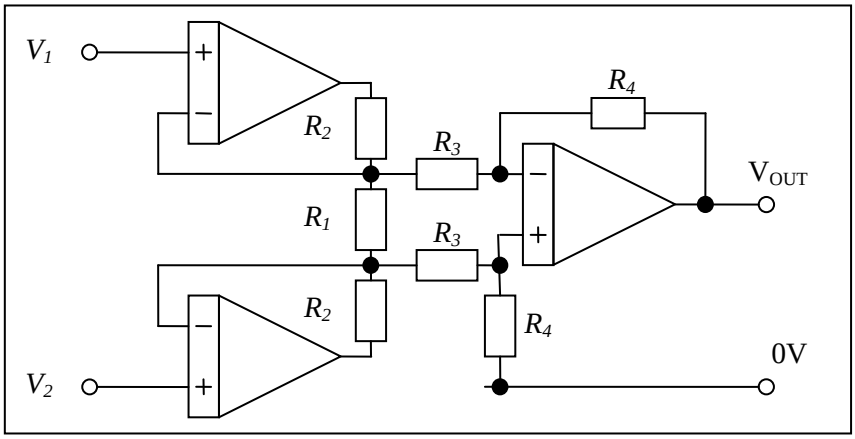


Fig. 4.2 A basic instrumentation amplifier configuration

Gain stability (short term) is critical and so the resistors would have to be very good ones (e.g. ultra precision bulk metal film types in a matched network) [3]. Linearity is also critical so that the op-amps would have to be very high performance or, more likely, composite op-amps or high gain blocks [4] [5] with a low noise front end [6].

The gain needs to be variable (switch selected feedback factors) which is also not trivial [5].

1. Bramley, P: "A few words from the microK100 design team" 2009. www.metrosol-ltd.co.uk
2. PB feedback: See section 6.14.
3. Part 1, monograph 1: "High accuracy resistors".
4. Part 4, monograph 1: "High gain blocks".
5. Part 6, monograph 3: "High accuracy amplifiers, integrators and differentiators". See, for example, a variable gain differential amplifier, section 6.3.
6. Part 5, monograph 2: "Low noise BJT preamps".

Common mode rejection is less critical as one of the inputs (probably V_2 - on the non-inverting side) is held, accurately, at local 0V by the active guard. See previous section fig. 3.6.

An alternative approach: -

Noise matching with a transformer is not possible and so a parallel input stage is the only practicable way of noise matching to a low source resistance. This can be achieved with a number of low noise matched pair BJTs in parallel, each with its own constant current source. These replace a single matched "long tail pair" in a composite op-amp circuit. The second stage differential amplifier could be based on a two-stage HGB. For details on low noise and high accuracy design see the relevant monographs [1 and 2].

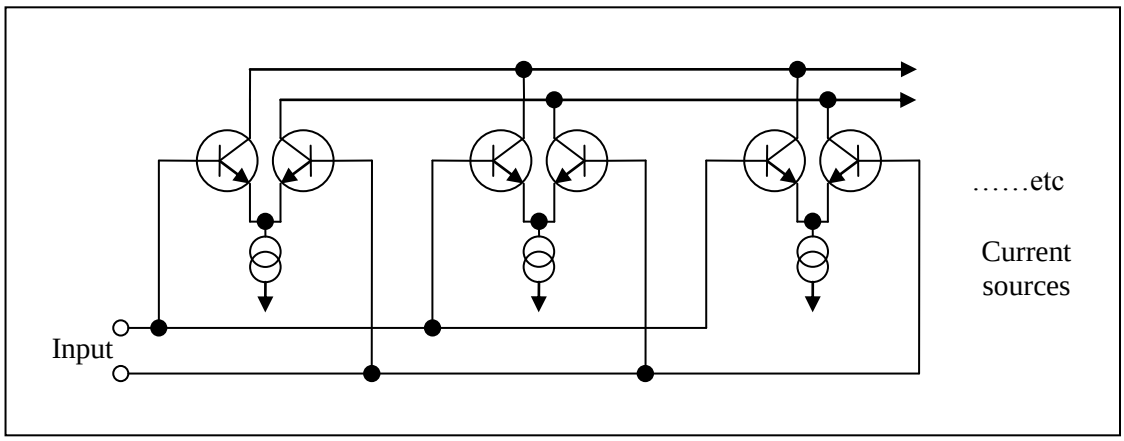


Fig. 4.3 Parallel connected pairs to reduce noise voltage

Common mode rejection depends on the output resistance of the current sources as well as the matching of the transistor pairs [1].

Another alternative is to connect a number of instrumentation amplifiers in parallel using a technique employed by Analog Devices [4]: -

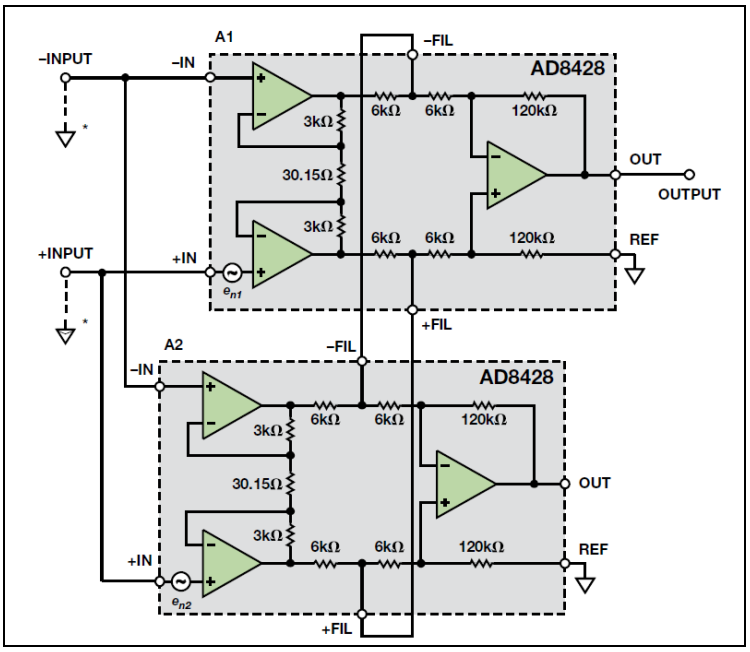


Fig. 4.4 Instrumentation amplifiers in parallel

1. Part 5, monograph 2: "Low noise BJT pre-amps"
2. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators"
3. Website: www.interfet.com
4. Gerstenhaber, M., Johnson, R. and Hunt, S.: Analog Dialogue 49-05, May 2015

A less likely option is a large area dual matched JFET. The IF3602 by Interfet Corporation [3], for example, boasts a low noise voltage of $0.3nV/\sqrt{Hz}$ at 100Hz but I have not been able to obtain information on current noise or low frequency performance.

As noted above common mode rejection is less critical as one input to the differential amplifier (probably V_2) is held, accurately, at local 0V: -

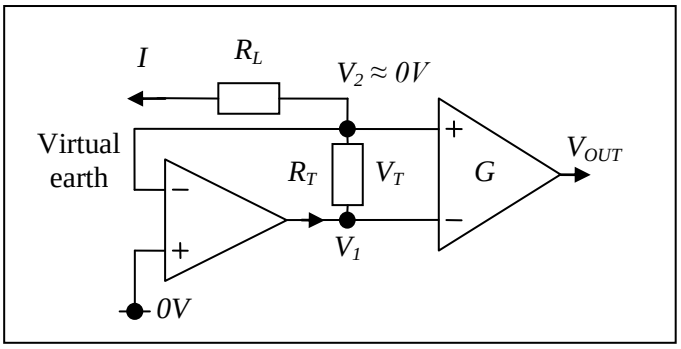


Fig. 4.4 The active guard circuit and differential amplifier

The main disadvantage with this method is that the (variable) resistance, R_L, appears in series with the constant current source, hence the need for an ultra-high output impedance ("dual stage") design. The required common mode rejection of the differential amplifier, on the other hand, becomes readily achievable.

Limited common mode rejection can be thought of as an asymmetry of the input. The gain from the non-inverting input to the output is slightly different to that from the inverting input so that, for example (referring to fig. 4.4): -

$$V_{OUT} = (G + \Delta G)V_2 - GV_1 = G(V_2 - V_1) + \Delta G V_2$$

The term $\Delta G V_2$ represents the error at the output. The virtual earth voltage, V_2, is determined by the open loop gain of the guard amplifier (typically $>10^5$ for a single low cost op-amp). Common mode rejection ratio is defined as $|\Delta G|/G$ and, in data sheets, is often quoted in dB (>80dB or 10^4 is readily achieved with standard components). The error, expressed as a ratio (compared to the output signal) is, therefore: -

$$\left|\frac{V_2}{V_1}\right| < 10^{-5} \text{ and } \frac{|\Delta G|}{G} < 10^{-4} \Rightarrow \left|\frac{\Delta G V_2}{G(V_2 - V_1)}\right| < 10^{-9}$$

In the previous section it was suggested that the two-stage constant current source could be replaced by a voltage follower. The resistance, R_L, consists of the current carrying lead and switch and sets up a common mode, V_2, at the input which is now independent of V_1 and much more problematic: -

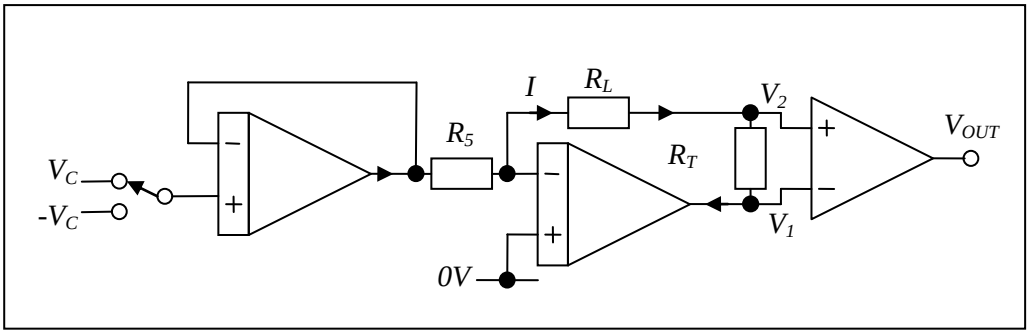


Fig. 4.5 An alternative approach

In an extreme example R_L could be 10Ω with a measuring current of 10mA hence $V_2 \approx 100mV$. A reasonable requirement is that this should result in an error at the output of the differential amplifier equivalent to no more than the noise limit $(\approx 1nV)$. If one assumes a zero differential signal $(R_T = 0)$: -

$$V_2 - V_1 = 0 \qquad \Rightarrow \qquad V_{OUT} = \Delta G V_2$$

Referred to the input: $$\frac{V_{OUT}}{G} = \frac{\Delta G}{G} V_2 < 1nV \quad \Rightarrow \quad \frac{\Delta G}{G} < 10^{-8} \qquad (CMRR > 140dB)$$

This is a very tall order for a conventional design (virtually impossible). Fortunately there is a simple solution: a "bootstrapped" floating power supply. The power supply to the differential amplifier and ADC is driven, very accurately, to follow the common mode voltage. The differential amplifier/ADC stages "see" no common mode. See fig. 4.6.

The problem is moved to the digital interface where the odd 100mV of common mode should not be a problem, especially if optical or differential signalling is employed – a good idea anyway (for speed and rejection of interference).

The microK employs an FPGA to perform all the high speed logic functions (ADC control and digital filtering) and it should be possible to incorporate, if necessary, a (less speed critical) serial interface to the main processor with a pair of opto-couplers.

The principle of the inside-out follower [1] is quite simple: the input signal is in series with the negative feedback connection so that the 0V(B) of the bootstrapped supply "follows" V_2. The supply to the high gain block (HGB) is also bootstrapped so that it, also, does not "see" any common mode. The open loop gain, at low frequency, of the HGB is so high that following accuracy can be better than 1ppm (two-stage HGB) [2].

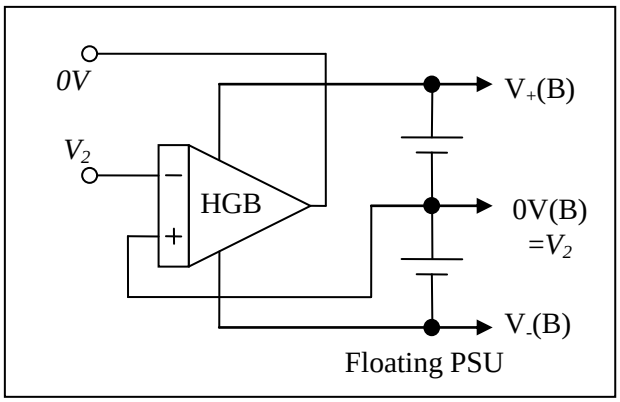


Fig. 4.6 The inside-out follower and bootstrapped power supply

In a revealing comment Bramley et al conclude that the differential amplifier contributes significantly to a quadratic non-linearity of the microK measurement system [3]: -

"A good example of this [quadratic error] is the power-coefficient of the resistors used to set the amplifier gain. Typically, resistors have a linear temperature coefficient of resistance. However, since the power dissipated in the resistor is proportional to the square of the voltage across it, this leads to a variation in resistance that is quadratic with applied voltage."

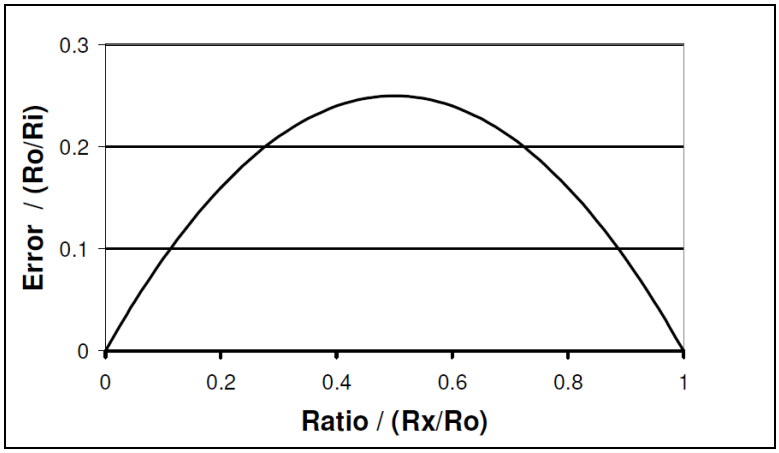


Fig. 4.7 The quadratic error due to the current source (courtesy Bramley et al [4])

Elsewhere the quadratic error is also attributed to the output resistance of the current source [4].

1. Part 4, monograph 2: "High accuracy voltage followers".
2. Part 4, monograph 1: "High Gain Blocks".
3. Bramley, P. et al: "Cost Effective Techniques …".
4. Bramley, P. et al: "Using a Substitution Measurement Topology…" See section 3.3.1

5. The sigma-delta ADC

5.1 Overview

The analog to digital converter is described, in outline, in Bramley et al [1]: -

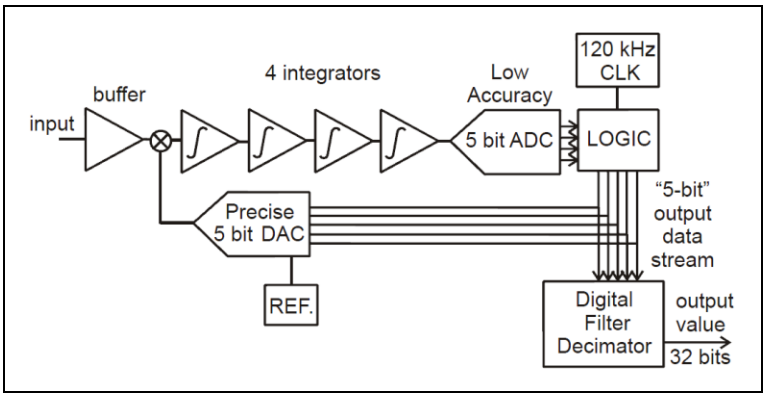


Fig. 5.1.1 The sigma-delta ADC in outline (courtesy Bramley et al [1])

The following analysis is based on the above (Fig. 5.1.1) and a few other clues (i.e. more speculation): -

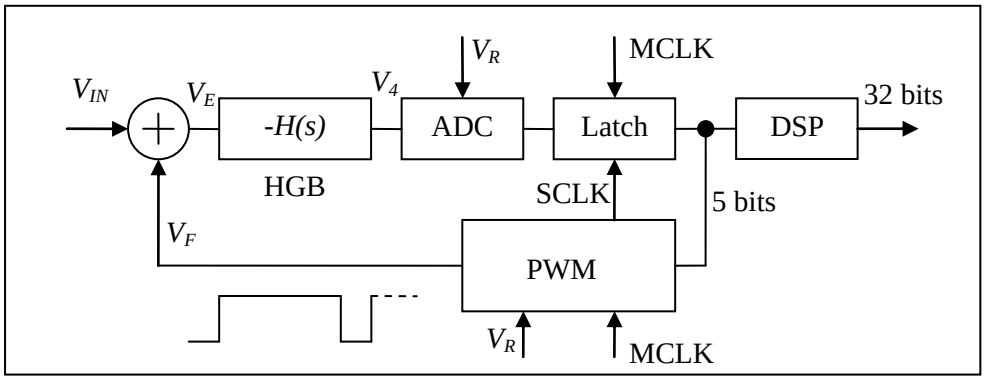


Fig. 5.1.2 A functional analysis

Unlike a conventional bridge the microK ADC is designed to oscillate between states either side of null balance.

The main elements are a summing junction, a high gain block (HGB with dynamic characteristic *H(s)*) a high speed analogue to digital to converter (5 bit: probably a flash type ADC), a high speed latch and a very accurate pulse width modulator (PWM). The output is a rapidly changing 5 bit code which is then passed through a low-pass digital filter (digital signal processor: DSP). At its very simplest this could be a simple accumulator that adds together a large number of samples. The microK DSP produces an output of up to 32 bits, apparently.

A summing junction was chosen, rather than a differential input stage, probably because the common mode rejection required would otherwise be impractical (see the next section).

The high gain block (HGB) is, in effect, a very high gain amplifier (at low frequency) with a dynamic characteristic designed to ensure closed loop stability (apart from the high frequency oscillation) and high accuracy. Bramley et al describe this as "four integrators" [1]. This may be the case at low frequency but it would be better to use a single integrator followed by three "one-plus-integrators" (see later) [2]. This author refers to such a circuit as a "four-stage HGB". Four stages does seem a bit excessive – it is shown in [3], for example, that three stages should be sufficient for the level of accuracy required.

1. Bramley, P. et al: "Better Accuracy in Temperature Calibration…"
2. PB insists on "four integrators". See section 6.16.
3. Part 4, monograph 1: "High gain blocks". See section 3.2

The flash ADC sampling frequency and PWM cycle period is set by the sampling clock (SCLK). In the microK literature this is variously specified as 5μs [1] and, more recently, 10μs. In the latest brochure, for example, (section: Performance by Design): "In order to achieve our target of <0.05ppm, we needed to be able to produce pulses whose edges have relative timing errors of <0.5ps" [2]. This is consistent with a PWM period of 10μs: -

$$\frac{0.5ps}{10\mu s} = 5\times10^{-8} \qquad (0.05\text{ppm})$$

For the MicroK Gold the timing accuracy is even greater (±0.1ps).

The accuracy of the pulse width modulator is paramount. The PWM is, in effect, a multiplying digital to analogue converter (MDAC): the average value of the output, over a single cycle, is the reference voltage multiplied by the 5-bit binary input code (see below). This is the key to the ADC's astonishing performance: -

The low resolution of the flash ADC/PWM pair means that the loop is very unlikely to achieve an exact (stable) null balance. The loop will usually oscillate, therefore, switching between states either side of null balance.

Apart from this instability one can think of the whole ADC as a linear feedback circuit, albeit with an A-D-A conversion process inside the loop (flash A-D followed by PWM D-A). See fig. 5.1.3.

As long as the PWM conversion is accurate the average value of its output, $\langle V_F \rangle$, is proportional to the average value of the 5-bit data stream and the master voltage reference. If one assumes that the A-D-A conversion process has a gain of one and the input, V_{IN}, is a low frequency sinewave then, in the complex representation $(s = j\omega)$, on average, referring to fig. 5.1.2: -

$$\langle V_F \rangle = (1+\delta)\langle V_4 \rangle = -(1+\delta)H(s)(\langle V_F \rangle + V_{IN}) \;\Rightarrow\; \frac{\langle V_F \rangle}{V_{IN}} = \frac{-(1+\delta)H(s)}{1+(1+\delta)H(s)}$$

The symbol $\langle V_F \rangle$ represents the average value of the PWM output (as seen with a low-pass filter) and δ a small error (systematic or random) due to the A-D-A process. With very high $|H(s)|$ to a very good approximation: -

$$|(1+\delta)| \approx 1 \text{ and } |H(s)| >> 1 \;\Rightarrow\; \frac{\langle V_F \rangle}{V_{IN}} \approx \frac{1}{H(s)} - 1$$

The circuit (fig. 5.1.2) behaves like a high accuracy inverting amplifier (gain = -1): $\langle V_F \rangle \approx -V_{IN}$

I.e. If one viewed the PWM output signal, V_F, through a 1kHz low-pass filter, for example, it would appear to be an inverted version of V_4.

Since $\langle V_F \rangle$ is accurately proportional to the average of the 5-bit data stream the latter is accurately proportional to the input voltage.

The brochure also specifies a "measurement time (per channel) of < 2s (1s with computer interface)" representing 200,000 (100,000) 5-bit samples [2]. It is likely that many more samples are required for 10ppb resolution. If one assumes 10^6 PWM cycles (10s average time) and the master clock jitter is purely random then the timing accuracy reduces to 10^3 times that required for a single cycle (i.e. 0.1ns). For more see [3].

1. Bramley, P. et al: "Better Accuracy in Temperature Calibration…"
2. Brochure: "The new microK family of precision thermometry bridges". Edition 2 (1516).
3. Part 4, monograph 7: "High accuracy DACs and ADCs.

It is clear that the feedback must be negative. A signal inversion could be implemented in the A-D-A circuitry but it is probably simpler to incorporate it into the high gain block. In the analysis of op-amp circuits, for example, feedback is applied to the inverting input (representing 180 degree phase shift) and the loop gain is usually specified from the non-inverting input, hence the choice of sign convention (*-H(s)* rather than *H(s)*).

The input is bipolar (range from -5V to +5V) [1]. This could be implemented in the PWM design but it is probably much simpler to keep the PWM unipolar (0 or 10V) and add an offset into the summing junction. A reasonably good low frequency model is, therefore, a high accuracy inverting/summing amplifier, based on a high gain block: -

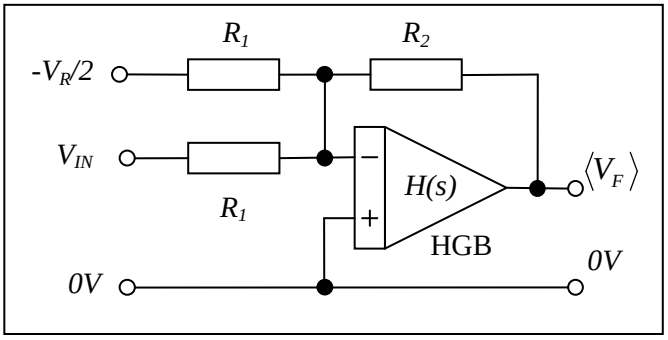


Fig. 5.1.3 A low frequency model of the sigma-delta ADC (unipolar V_F)

For an input range of ±5V, therefore, the local reference voltage needs to be V_R = 10V and the PWM output range is V_F = 0 or 10V. The average output of the PWM is accurately related to the input: -

$$\langle V_F \rangle = -\frac{R_2}{R_1}\left(V_{IN} - \frac{V_R}{2}\right)$$

The AC analysis of this circuit results in the transfer function: $\dfrac{\langle V_F \rangle}{V_{IN}} = -\dfrac{R_2}{R_1}\left(\dfrac{H(s)}{1+H(s)}\right)$

The factor in brackets (the "*D*" factor) embodies the dynamic response. It also describes the difference between the actual output and the ideal. At low frequency the gain is very high so that, to a very good approximation: -

$$|H(s)| >> 1 \;\Rightarrow\; \frac{\langle V_F \rangle}{V_{IN}} \approx -\frac{R_2}{R_1}\left(1 - \frac{1}{H(s)}\right)$$

According to PB, however, the PWM output is bipolar [2] (probably -5V or 5V) and the offset is not required. The analysis, however, remains the same: -

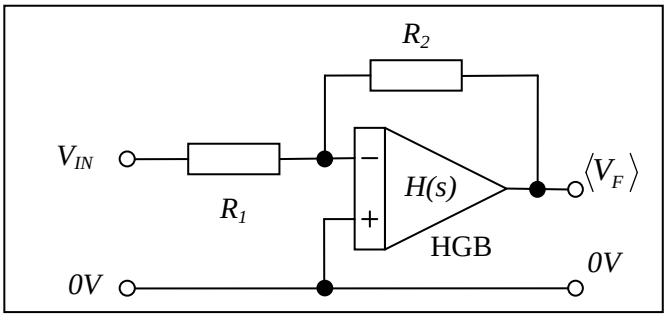


Fig. 5.1.4 A better low frequency model (bipolar V_F)

It would be reasonable to assume that R_1 and R_2 are a pair of accurately matched resistors (ideally: $R_1 = R_2$).

1. Bramley et al: "Cost Effective Techniques…"
2. PB comment: "Our PWM is actually bipolar. The symmetry helps ensure linearity…" See section 6.16

5.2 The summing junction and high gain block

To avoid problems with common mode the summing junction is almost certainly a virtual earth – adding currents. In the paper by Bramley et al, for example, a diagram shows resistors connected to the non-inverting input of an op-amp with nothing but a capacitor providing the feedback to the inverting input [1]. This is clearly a mistake but a sufficient clue as to the intention [2]: -

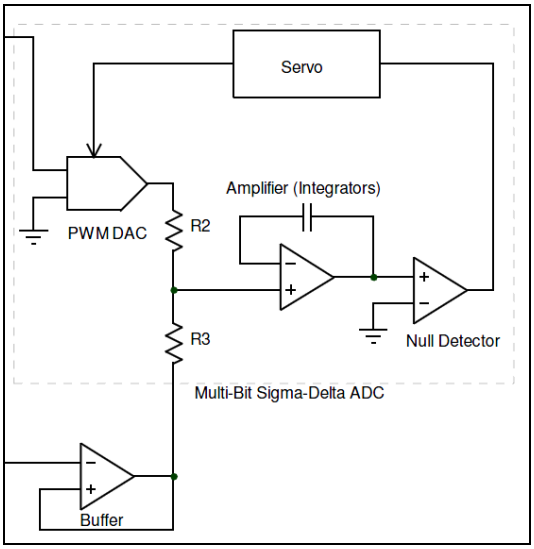


Fig. 5.2.1 Part of the diagram by Bramley et al [1]

Much more likely is the following circuit – combining a summing junction and integrator: -

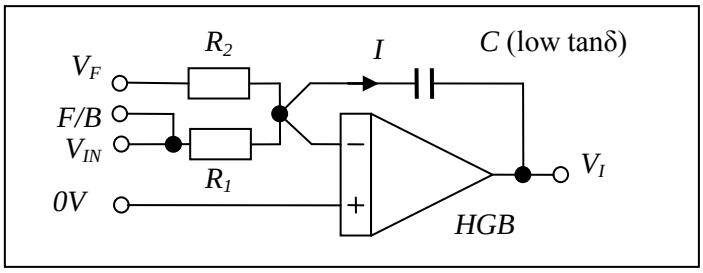


Fig. 5.2.2 The first stage (high accuracy) adder/integrator

Resistors R_1 and R_2 are critical components. Ideally $R_1 = R_2$ at least for the time between substitutions. **Here is another problem: R_2 includes the output resistance of the PWM D-A converter.** I shall, however, continue with the analysis (more anon).

The result is a falling output, with a constant slope for the first part of the PWM cycle (V_F = +5V) and an increasing output for the second part (V_F = -5V: see the appendix for detail). The output of the adder/integrator at the end of a PWM cycle (≈10μs) is usually higher or lower than at the start: -

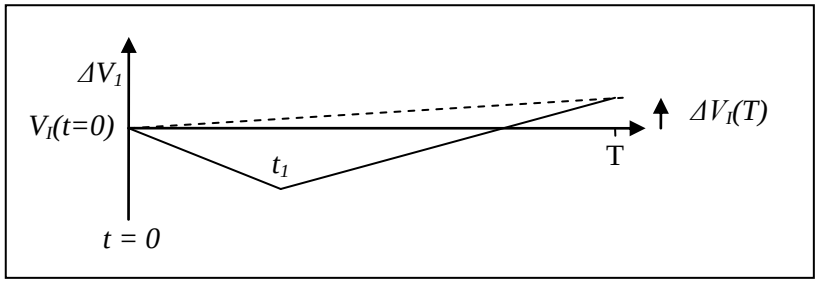


Fig. 5.2.3 Output of the integrator (single PWM cycle, zero noise)

1. Bramley, P. et al: "Using a Substitution Measurement Topology…" See section 3.3.2.
2. PB comment: "You're absolutely right…." See section 6.16.

The integrator is also a critical component, hence the need for a low noise, high performance HGB and a good quality feedback capacitor. The extra gain, at low frequency, can be provided with two or three (non-critical) one-plus-integrator stages: -

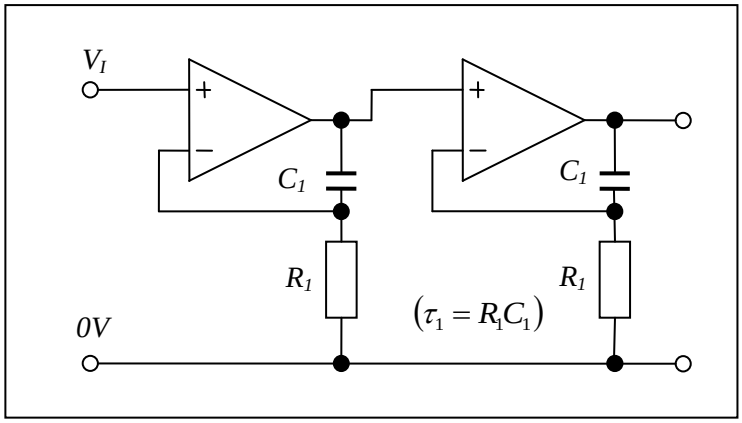


Fig. 5.2.4 Two one-plus-integrator stages

The combined frequency response for a four-stage HGB is as follows: -

- At very low frequency the gain is huge, limited only by the DC gains of the op-amps. With a low cost op-amp this is at least 10^{20} (10^5 per op-amp). This is much higher than required.
- As frequency increases the HGB behaves like four integrators in series with the gain falling at a rate of four decades per decade of frequency.
- At a frequency, chosen by design (typically 16kHz), the gain of the one-plus-integrators level off to a gain of one leaving a single integrator characteristic as the gain passes through 0dB (magnitude of one and phase lag of approximately 135 degrees at about 100kHz). The resulting 45 deg phase margin provides sufficient closed loop stability.
- At very high frequency (about 5MHz) the op-amps run out of steam (with "extra poles") and the gain starts to drop further, with increasing phase shift.
- The closed loop response has very high accuracy at low frequency but a slight resonant peak at high frequency. See figs. 5.2.6, 5.2.7 and 5.2.8.

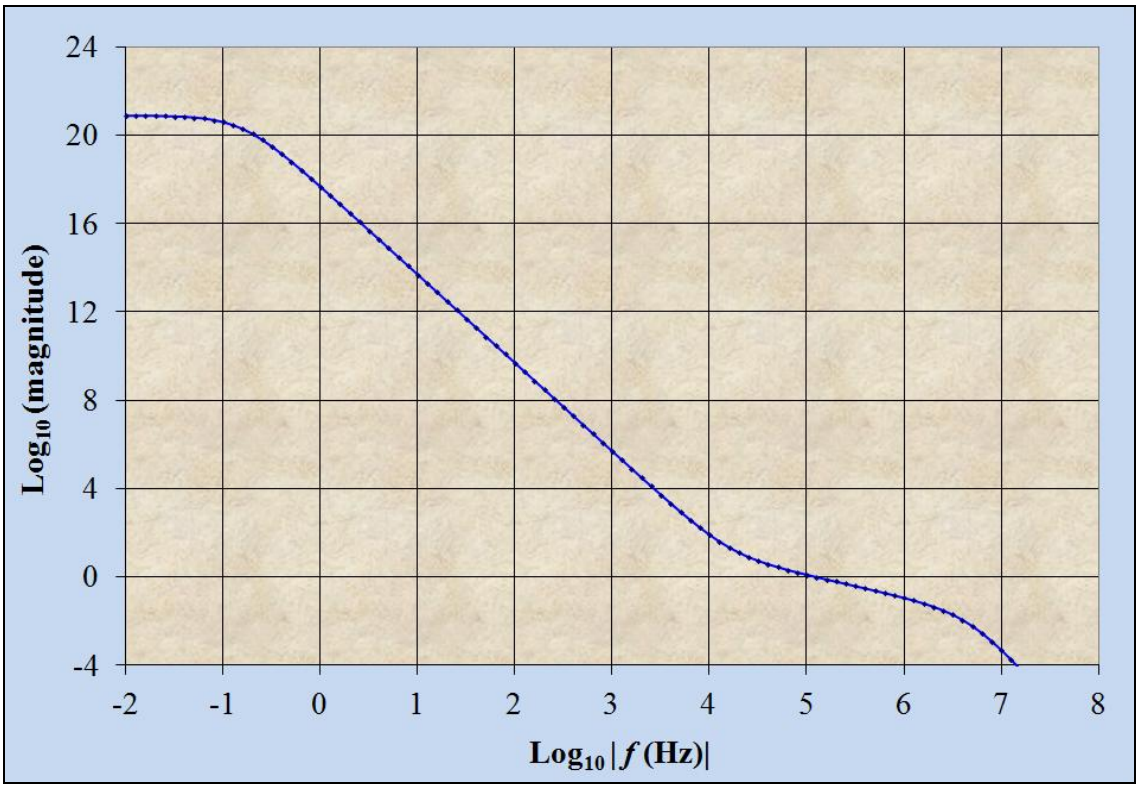


Fig. 5.2.5 Open loop frequency response of a four-stage HGB

A four-stage HGB seems excessive. In AC bridge designs three-stage HGBs provide more than sufficient open loop gain for the target accuracy [1]. I shall assume, therefore, a three-stage HGB, including the flash A-D-A process. The open loop transfer function is, in the usual complex representation $(s = j\omega)$: -

$$H_3(s) = \left(1 + \frac{1}{\tau_1 s}\right)^2 \frac{1}{\tau_2 s}$$

In a form normalised to the time constant of the one-plus-integrators $(s = j\omega\tau_1)$: -

$$H_3(s) = \frac{(1+s)^2}{\alpha s^3} \quad \text{with} \quad \alpha = \frac{\tau_2}{\tau_1}$$

The low frequency model of the sigma-delta ADC is, therefore: $\frac{\langle V_F \rangle}{V_{IN}} = -\frac{R_2}{R_1}\left(\frac{H_3(s)}{1+H_3(s)}\right) = -\frac{R_2}{R_1} D_3(s)$

The "*D*" factor for a three-stage HGB is, in normalised form $(s = j\omega\tau_1)$ [1]: $D_3(s) = \frac{s^2 + 2s + 1}{\alpha\beta s^3 + s^2 + 2s + 1}$

β is the reciprocal of the feedback factor. If both resistors (see fig. 5.1.4) have the same value then $\beta = 2$

$\alpha = \frac{\tau_2}{\tau_1}$ is the ratio of the time constants. The value of $\alpha\beta$ determines the phase margin and the size of the resonant peak and the real and imaginary errors at low frequency for a sinusoidal input. Analysis and modelling (and experience) indicate that a good compromise is [2]: -

$$\alpha\beta = 0.4 \;\Rightarrow\; D_3(f) \approx 1 + j0.4\left(\frac{f}{f_1}\right)^3 + 0.8\left(\frac{f}{f_1}\right)^4 \quad \text{with} \quad f_1 = \frac{1}{2\pi\tau_1} \approx 16kHz$$

The imaginary component represents a phase error (quadrature) and the real part the in-phase error. The frequency $f_1 \approx 16kHz$ represents the highest that is practical for readily available (low cost) op-amps [1].

Typical component values are: resistors and capacitors for the one-plus-integrators $(\tau_1 \approx 10^{-5} s)$ are 10kΩ and 1nF respectively. For the integrator: -

$$\alpha\beta = 0.4 \;\Rightarrow \alpha = 0.2 \;\Rightarrow\; \tau_2 \approx 2\times10^{-6} s \;\Rightarrow\; 10\text{k}\Omega \text{ and } 200\text{pF} \quad (f_2 \approx 80kHz)$$

One could argue for a higher value for the resistors because of the output resistance of the PWM (appears in series) with a correspondingly lower value for the feedback capacitor. This is probably unwise due to parallel (stray) capacitance injecting "spikes" into the integrator capacitor. It would probably be a good idea to pass the resistor through a grounded screen anyway, to minimise this problem. For more detailed discussion see [3].

The slight resonant peak (about +4dB) at high frequency also means that, with a step change at the input, the ADC output overshoots (see fig. 5.2.6). The transient is quite large (≈ 30%) but converges rapidly to an accurate result (see figs. 5.2.7 and 5.2.8). Such an event would happen at the start of each reversal cycle unless there is some form of "preset" of the ADC state by the control logic.

1. Part 4, monograph 1: "High gain blocks". See sections 2 and 4.1.
2. Part 4, monograph 2: "High accuracy voltage followers". See section 2.4.
3. Part 4, monograph 7: "High accuracy DACs and ADCs"

The dynamic response (frequency domain) for the transfer function is as follows: -

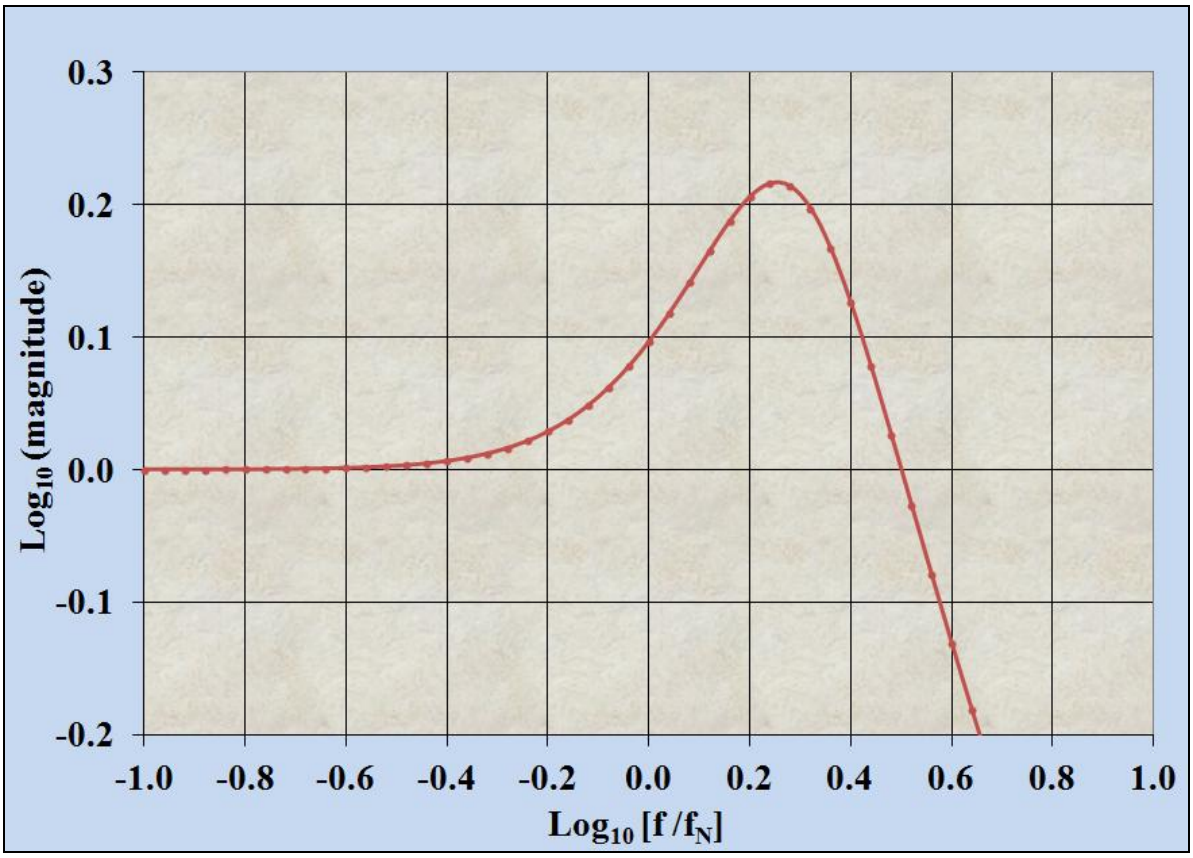


Fig. 5.2.6 Frequency response [1]

$$D_3(s) = \frac{s^2 + 2s + 1}{\alpha\beta s^3 + s^2 + 2s + 1} \quad \text{with} \quad (s = j\omega\tau_1) \quad \text{and} \quad \alpha\beta = 0.4$$

The natural" frequency is: $f_N = \frac{1}{2\pi\tau_1} \approx 16kHz$

1. Excel spreadsheet: "MicroK freq response"

The step response (time domain) is as follows [1]: -

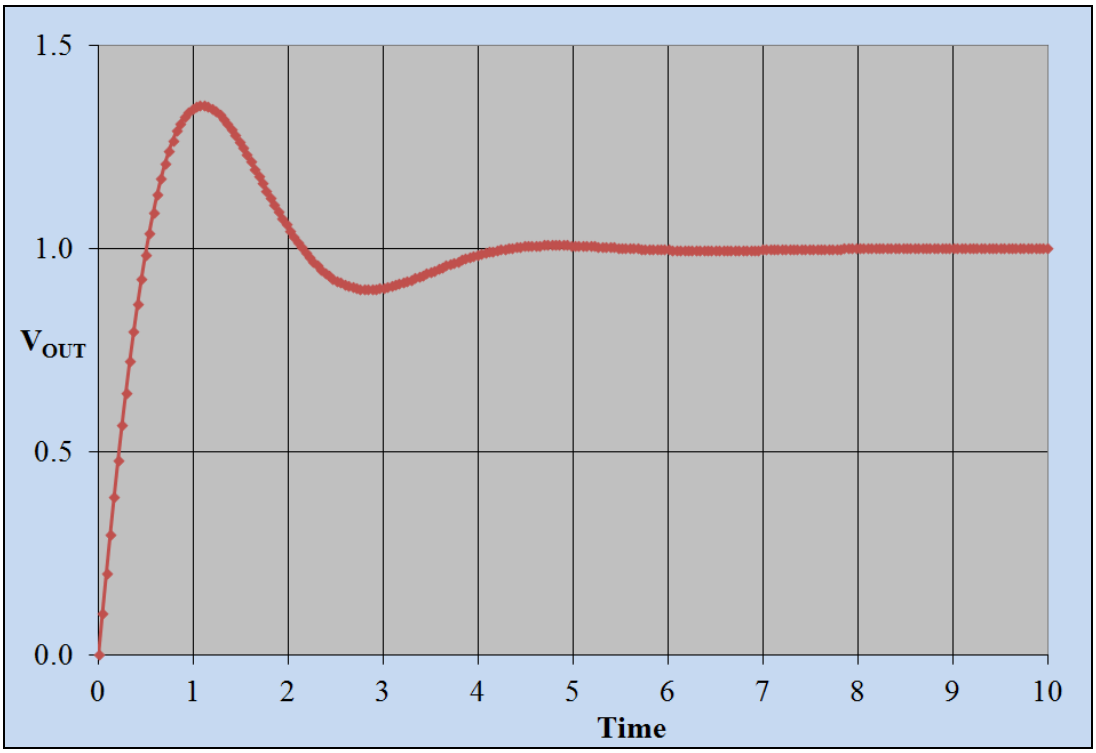


Fig. 5.2.7 Unit step response (units of τ_1 time constants: 10µs)

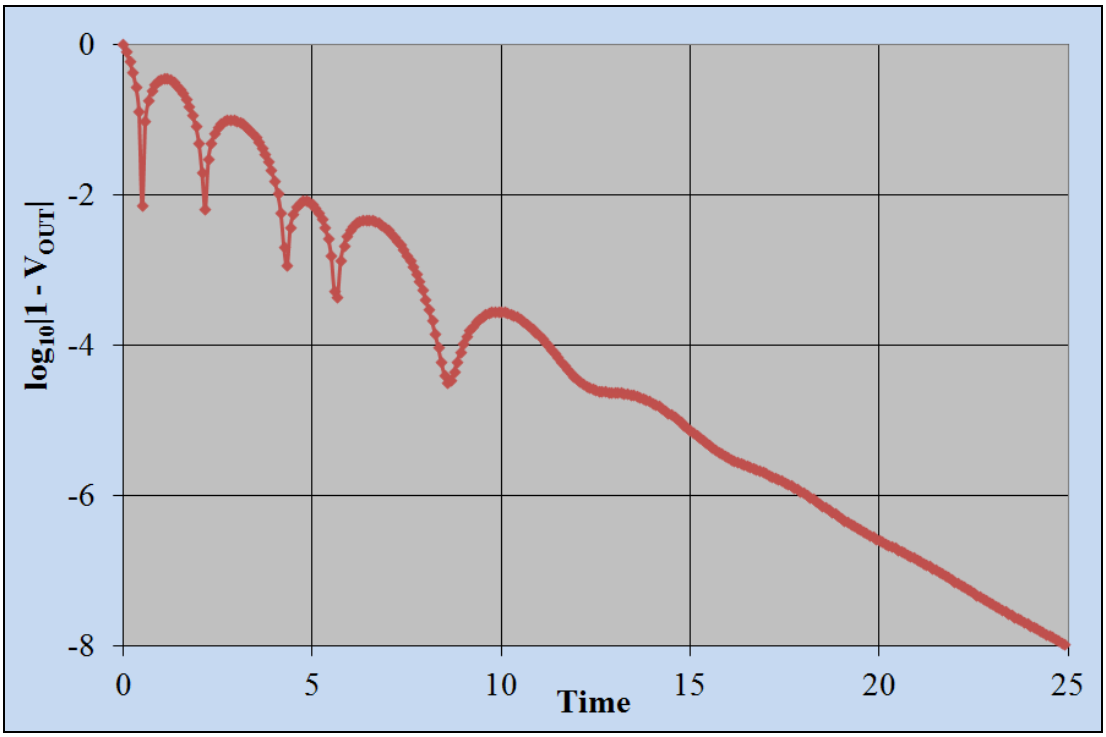


Fig. 5.2.8 Error analysis (units of τ_1 time constants)

I.e. error settles to <10ppb in 25 time constants (≈ 250µs)

N.B. Settling time could be significantly reduced if the ADC is preset to a suitable state. For more detail see [2].

1. Excel spreadsheet: "MicroK step response"
2. Part 4, monograph 7: "High accuracy DACs and ADCs"

5.3 The flash ADC and pulse width modulator

One of the most critical elements of the microK measurement system is the pulse width modulator (PWM). Whereas the microK PWM has a resolution of only 5 bits it needs to be very accurate – for the microK gold to >25 bits (ratio error < 30ppb)! According to PB the PWM is bipolar [1] (probably switching between -5V and +5V): -

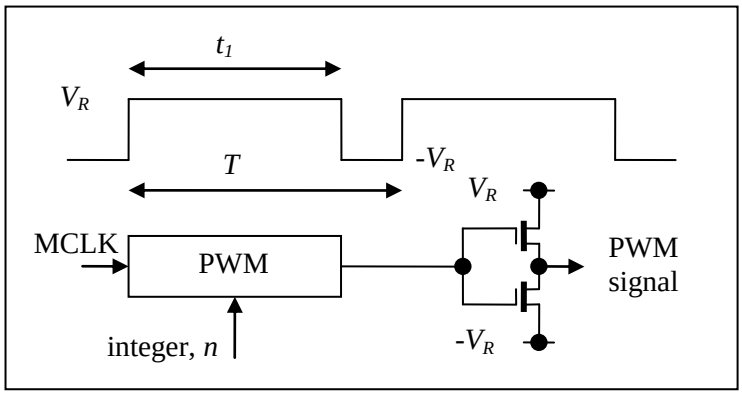


Fig. 5.3.1 The principle of a bipolar PWM based MDAC

For the ideal PWM the average output voltage, over a single PWM cycle, is the reference voltage multiplied by the ratio of two integers: -

$$V_{AVE} = 2V_R \frac{t_1}{T} - V_R = V_R\left(\frac{2n}{N} - 1\right) \qquad n = 0,1,2,.....N \quad \text{with} \quad N = 31 \ \text{(11111 binary)}$$

I.e. There are 31 time slots and the PWM signal varies from 0 to 31 slots occupied by contiguous "ones".

In a diagram by Bramley et al it is suggested that the input to the PWM is a (5 bit) binary code [2] (see fig. 5.1.1). This is not entirely surprising as it is likely that a field programmable gate array (FPGA with 150,000 gates), on which the microK logic is based, usually includes standard building blocks (including IP or "intellectual property" blocks). One of these is likely to be a flash ADC with 5-bit binary decoder included. The PWM is also (probably) constructed with standard building blocks: a synchronous counter (5-bit binary), comparator and parallel latch: -

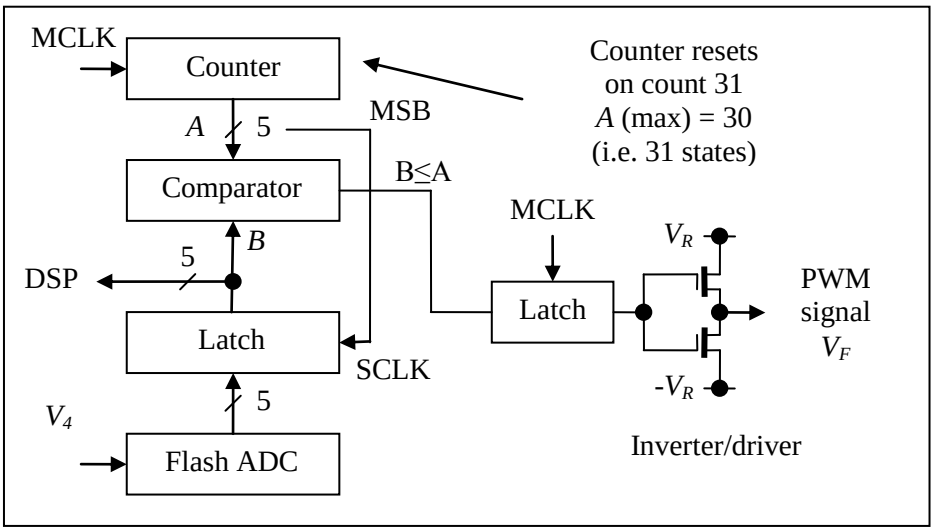


Fig. 5.3.2 A conventional PWM design

Examples: -
ADC outputs all "zeros" (00000): B≤A is always valid and the comparator output remains one (PWM output = 0).
ADC outputs code 5 (00101): See fig. 5.3.3. (PWM output = 1 for five contiguous "slots").
ADC outputs all "ones" (11111): The counter never reaches 31 and B≤A is never valid (PWM = 1 for 31 slots).

1. PB comment: See section 6.16.
2. Bramley, P. and Pickering, J.: "Better Accuracy in Temperature Calibration…"

It is not clear, in any of the published literature, how the astonishingly accurate timing is achieved. The figure quoted (0.5ps) for 50ppb [1], however, is not an absolute figure – it is the voltage profile, within each PWM cycle, that must be accurately related to each 5-bit code. Timing accuracy is only part of the picture. Nothing of the inner workings of the logic is published and the following is, therefore, **highly speculative**: -

Some design considerations for the logic circuitry: -
a). The exact frequency of the master clock (MCLK) is not important – as long as it is sufficiently stable over each PWM cycle. For a PWM cycle of 10μs the master clock is 3.1MHz.
b). Random fluctuation of the master clock frequency introduces noise into the measurement. The effect averages out over a large number of cycles but may extend the time required to achieve the target resolution. A rather special (very low noise) clock is probably used.
c). Systematic variations in frequency over each PWM cycle must be avoided. It is possible to imagine, for example, some interference (an impulse) from the sampling clock (SCLK) getting into the master clock circuit causing the frequency to jump at the start of each PWM cycle which then begins to settle down to the steady state frequency. The result would be a non-linearity of the PWM output relative to *n* (PB confirmed [2]).

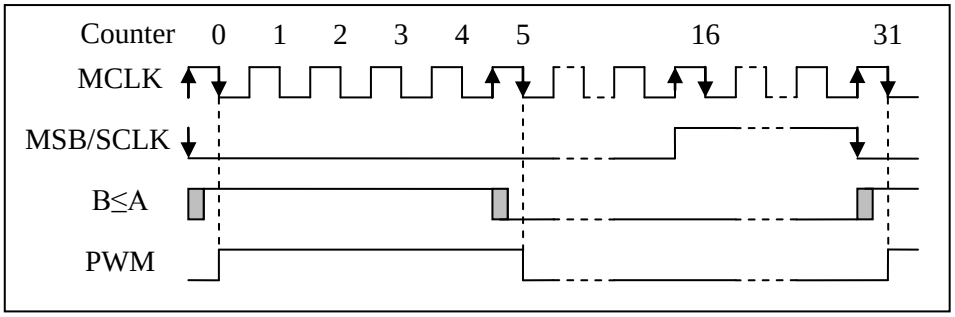


Fig. 5.3.3 Timing diagram of a synchronous PWM circuit (B = 5)

d). The design needs to be synchronous and "glitch free". Most high speed FPGAs have standard building blocks that do this, plus methods of distributing a master clock to minimise clock "skew" (i.e. clock transitions reaching different parts of the chip at different times). In addition to a master (power-up) reset for all flip-flops a master timing/sequence generator is required. Also, high speed flash converters often employ a track/hold circuit to hold the input constant during sampling, to minimise glitches (in this case known as "sparkles").

The timing diagram (fig. 5.3.3) provides an example of a synchronous PWM circuit (see fig. 5.3.2): -

- Counting occurs on the rising edge of MCLK. On a count of *N* (31) the 5-bit counter self-resets. I.e. it never gets to 31 (11110 changes to 00000). This is necessary for full bipolar operation (PWM output range from $-V_R$ to $+V_R$).
- The most significant bit (MSB) of the counter can be used as a sampling clock (SCLK). A falling edge indicates the start of a PWM cycle and is used to latch the flash ADC output. The rising edge of SCLK could be used to indicate that the code B is valid and can be "pushed" on a stack or into a FIFO memory.
- After a short delay both inputs to the comparator settle to valid states (*A* and *B* both valid).
- After a further delay (variable, depending on the inputs) the output of the comparator (B≤A) becomes valid.
- As long as the total delay is less than half a MCLK cycle (≈160ns for a 10μs PWM cycle) the high speed PWM latch reliably synchronises the PWM signal to the falling edge of the MCLK. The approximate PWM signal (error caused by the variable comparator delay) becomes a very accurate PWM signal - the on/off ratio is accurately related to the binary code B.
- The flash A-D-A process need not be accurate – the feedback loop finds a balance eventually. The data stream to the DSP, however, must be the same as that which generates the PWM signal. It is best, therefore, to take the data stream from after the ADC latch.
- With a PWM cycle of 10μs/sample it should be possible to employ a fast microcontroller and an interrupt driven subroutine (fast "push" and "pop" stack operation) to acquire the data stream.

1. Brochure: "The new microK family of precision thermometry bridges". Edition 2 (1516).
2. PB feedback: "Timing jitter needs to be good, but more importantly it can't have any significant correlation with the conversion or measurement cycles." See section 6.16.

e). **The PWM output resistance is critical: -**

- The "on" resistance of the MOSFETs appears in series with the high precision resistor (R_2 see fig. 5.2.2) and needs to be sufficiently small, symmetrical and stable over the time between substitutions.
- The summing resistors must be "lowish" or else contribute Johnson noise. If $R_2 \approx 10k\Omega$ then $\Delta R \approx 0.1m\Omega$.
- A very low output resistance ($\approx 20m\Omega$) is possible with large area MOSFET devices but the problem is then large gate-source/drain capacitance and large charge injection.

f). It is not clear how the high level PWM signal (10Vpk-pk) is generated – all high speed logic families and FPGAs operate on much lower power supplies. Perhaps there is a high level MCLK and a level shifting MOSFET driver before the PWM signal is synchronised.

g). It should be possible to dispense with the binary encoder in the flash ADC. The result is a faster conversion: -

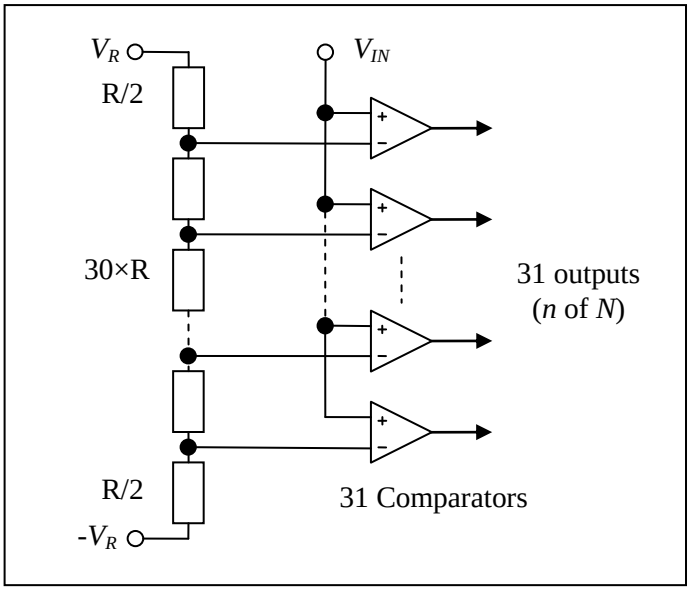


Fig. 5.3.4 A basic flash ADC (bipolar without binary decoder [1])

The n of N code scheme is easily converted to a PWM signal with a pre-settable shift register: -

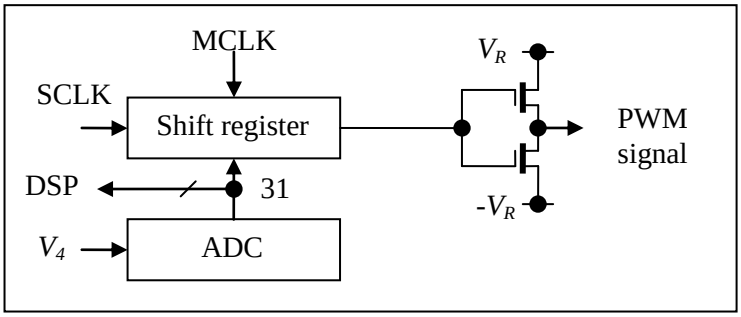


Fig. 5.3.5 A simplified high speed PWM

N.B. The n of N code is often referred to as a "thermometer" code, by analogy with the level of mercury in a simple thermometer (example: 1111100000000….etc represents n = 5). One advantage is that it is less prone to large errors due to "sparkles" (a fast changing ADC input and one relatively slow comparator can produce a false 0 (example: 1110100000….etc for n = 5). This would represent an error of 1/31 for the PWM cycle and would be easily corrected by the feedback loop. A binary decoder, on the other hand, could introduce a very large error, upsetting the loop equilibrium, requiring time to recover. A track/hold circuit would help.

h). However the performance is achieved it is clearly very ingenious!

1. It is usual for flash converter comparator thresholds to be set half way between binary states, hence the two R/2 resistors.

6. Feedback and comment from Paul Bramley

6.1 Noise performance

"I don't really agree with this statement. If set to comparable settings i.e. ASL bridge bandwidth and number of samples per reading set to give ~same response time, then the two are very similar. Users report that the microK gives consistently same standard deviation and F18/F900s give more variable SD when making same measurement. I concede that "on a good day" an F900 can give slightly lower SD as it should due to differences in the way it works, but competent users prefer microK."

6.2 Performance compared to conventional null balance bridges

"Other and very significant differences are: -
1. Drift free. ASL/MI bridges require regular service visits to maintain alignment and therefore performance.
2. Much higher reliability due to use of more modern manufacturing techniques, elimination of contacts (connectors, potentiometers) and robust design (comprehensive input protection)."

6.3 Is it a bridge?

"Yes... I used to think of this as a reversing DC potentiometric instrument. However, when you look at the measurement architecture it looks more like a Wheatstone bridge. One arm is formed by the DUT and the current sense resistor on the current source (these carry the same current, of course). The other arm is formed by the current steering resistors in the ΣΔ ADC. The null detector is the one in the ΣΔ ADC, which works by balancing its input voltage against the delta modulated bit stream it generates. The measurement is, of course, dependent on the values of these 3 resistors so we overcome this problem by using a substitution technique. Anyway, it really is very like a Wheatstone bridge plus some active circuitry (which you would use in any modern Wheatstone bridge anyway). More detail is in the paper attached *- see section 4) [1]. To be honest, I'm less concerned whether people think of it as a bridge, it's only its performance that matters."

*The section to which this refers includes the following diagram (and the mistake mentioned above – see section 5.2): -

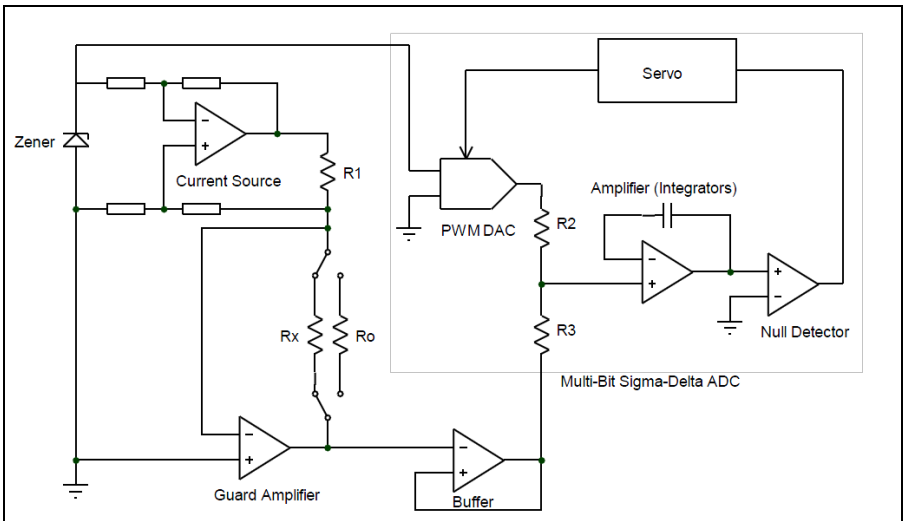


Figure 10: microK bridge measurement system.

Fig. 6.3.1 *From section 4 (courtesy Bramley et al [1])

CID: Does it look like a Wheatstone bridge? Can you see any mistakes?

1. Bramley, P. et al: "Using a Substitution Measurement Topology…"

6.4 Timing accuracy

"Yes, timing accuracy is critical, but only relative timing of edges as PWM is used as the feedback DAC in the ΣΔ converter."

6.5 Reversal frequency and 1/f noise

"The reversal also eliminates the 1/f noise from amplifiers provided reversal rate is above the corner frequency. Input reversed "behind" connectors for DC voltage measurements. Current reversed for resistance measurements."

6.6 Master reference voltage device

"We use a very good buried zener reference. As you say this is one of the ultimate limits for voltage measurement. Not such a problem with current as the reference is shared between ADC & current source so it's inherently a ratio measurement with ΣΔ ADC providing the balance mechanism."

6.7 The differential amplifier

"The diff-amp is "floating" but we use active guarding to maintain one of the inputs at an internal 0V reference. This eliminates the common-mode problem optimally by ensuring that lead resistances to the 4-terminal resistance do not lead to a common-mode voltage."

6.8 Elimination of amplifier offset and thermal emfs

"Actually, we make more than 2 measurements and weight them. For example, using 3 with the middle one reversed but weighted by x2 eliminates any effect of offsets that are changing linearly with time... a useful tip I picked up from my former brilliant colleague, John Yewen. But the principle of reversal is, of course, involved (but only eliminates "static" offsets)."

6.9 The need for certain components to have very good short term stability

"The + & - currents are very well matched. However, the current reversal means that the stimulus is the difference between the two current polarities and this difference is measured with the ADC. Both of these change linearly with Vref so I don't think it is just the substitution that eliminate sensitivity to imbalance between the + & - current."

Also: "Yes, with the "fast" reversals the stability of gain etc. is not a problem, but you are right... this is a pre-requisite."

6.10 The measuring currents are proportional to a master internal reference voltage, V_{REF}, probably via a 12-bit multiplying digital to analogue converter

"Yes, it is a 12-bit DAC. You can lose "probably" if you wish".

6.11 The current regulator

PB feedback confirmed that section 3 analysis is not far from the truth "…the principle is correct" but later "…we also don't wish to give away our IPR!"

"We actually buffer (voltage follower) V4 connecting back to R4 which changes your calculations (simplifies them) but the principle is correct. We tend to publish only the principle and leave out non-essential detail."

"We couldn't use one JFET for both directions (gate channel would become forward biased). We therefore use both a P & N JFET where the "Unused" one is isolated by suitable biasing.
"Again, I think regarding the JFET as the active component in a servo system rather than a voltage controlled resistance helps clarify operation. Of course we use JFETs so that there is no current injected into the current source circuits."

6.12 Is the current regulator second stage a "cascode" stage?

"I think of Cascode as a common base/gate stage following an output stage to provide isolation. This is commonly used to reduce the miller effect to improve bandwidth but it is equally good at isolating the output from the load to increase output impedance. I think the term is legitimate."

Later, also: "Just reading this now after commenting above. Wikipedia does mention the increased output impedance as well as the more usual benefit sought of higher bandwidth."

Referring to fig. 3.3 (two stage design): -

|
|
|
Detailed feedback redacted to protect IPR.
|
|
|

Referring to estimated component values: -

"Values are wrong (wouldn't wish to disclose precise details obviously) but the principle is correct."

6.13 Suggested alternative design (Fig. 3.7): -

"Not as good as above when you consider the effect of current flowing in lead (& track resistances). You either end up with common mode voltage or sense current being affected by these resistances. It's unavoidable as far as I can see."

6.14 The differential amplifier

"As you seem to have deduced, the amplifier is considerably more complicated than disclosed but the principle is correct. We use an array of op-amp based amplifier stages to amplify the signal. These are then added/averaged. The signal sums linearly and the noise (uncorrelated) sums as the square root. You therefore get a root N (N = number of amplifiers) improvement in voltage noise at the cost of current noise. It's exactly the same noise impedance matching using a transformer in F18... but it's harder with DC because you need a lot of amplifiers rather than just optimising the turns ratio!
For commercial reasons we don't provide much more detail... it wasn't easy to get it working. But it does work, giving true differential, low voltage noise (noise impedance is a few hundred ohms) highly linear (ppb) amplifier. We did look at using matched transistor pairs such as the LM194 but chose to use a different approach using multiple op-amps (in parallel) to lower the noise and in series (to increase gain & ensure adequate linearity). We don't really need such a high CMMR to achieve our performance. By connecting the sense point of the guard amplifier to the diff-amp input and using the cascode isolated current source there is virtually no effect of lead resistance on the current source OR the CM of the diff-amp."

6.15 Bootstrapped PSU

"Yes... a good technique and used only on the more accurate microK products."

"Yes, one of the factors we had to consider in the design of the ultra-linear amplifier is the power coefficient of the resistors used to set the gain. As you appreciate, these need to be lowish value (to avoid Johnson noise) but then their non-linear resistance with voltage can create gain non-linearity."

6.16 The ADC

“All correct though the appearance of PWM might be a little misleading for readers. It is a 5 bit PWM feedback DAC that is fed with a sigma-delta modulated feedback code. Not sure it needs elaboration, but showing just a PWM stage might lead people to think that this is a PWM DAC being used directly as an ADC.”

“A simple accumulator works, but it isn't optimal of course. We use various filters (including simple accumulation) depending on the target specification.”

“The 2s (timescale) includes the reversals (several of them) and the substitution. The accuracy is 30ppb but a single reading at 2s has a high uncertainty due to noise (even the Johnson noise is more than the 30ppb). To get the noise low enough to be able to "use" the accuracy requires multiple samples per reading (using the settings) in the same way that the F18/F900 need to reduce the bandwidth in order to exploit their underlying linearity.”

“More stages gives us more gain and lowers the noise, we do use 4 stages in microK as design simulations showed this to be worthwhile. The details are fading from memory as this was all done in 2005!”

“Spot on, nice description of how it works. Our PWM is actually bipolar. The symmetry helps ensure linearity (I think!)”

Error in outline diagram (fig. 5.2.1): -

“You're absolutely right! Can't recall why I used the poor representation for the integrator (may be a mistake or may have just been trying to indicate the intention of the sub-circuit). Not quite as fig 5.2.2 as it really is fully bipolar ΣΔ ADC, but the principal is spot on.”

“We find that 3 integrators is okay for a normal ΣΔ ADC and more only leads to instability as you indicate. Have a 5-bit feedback reduces the quantisation noise and simultaneously allows us to add another integrator. When modelled the quantisation noise reduces not by a factor of 32 (5 bit) but by about 600. This makes the QN negligible (the pre-amp noise then dominates the input referred noise.”

“Absolutely right though the ADC dead time is very small. The photoMOS relays have on/off times of a few ms so these are more significant (though worth it for their solid state reliability). The lost time does reduce S/N ratio but not by much.”

“Yes, the relative timing accuracy is critical as it carries the full accuracy burden of the measurement. We have swapped an analogue accuracy (for a conventional 5-bit DAC) for a timing requirement. This is generally easier thought it took us many months to develop the technology to achieve this and is part of the core IP (all hidden in the FPGA and surrounding circuitry, of course).”

“Timing jitter needs to be good, but more importantly it can't have any significant correlation with the conversion or measurement cycles. It wasn't too hard.”

“It's properly bipolar so this is not a problem.”

Critical components in the flash ADC?

“Yes, we don't need anything "special" and rely on the design being very tolerant rather than high stability components.”

“The design of the ADC internal processing is such that it very rapidly converges after a step change. There is very little latency. We lose only a few cycles and this is built into the FPGA logic.”

“Hi Chris, thanks for sharing a truly impressive overview (better than anyone else I have come across). I think you have done a great job in understanding the guts of the microK. Some of the conclusions are not correct and I have tried to share information where I can. Obviously a lot is rather valuable IP and regrettably I can't really go further. I have learnt a few lessons from reading your work here and will put them to good use in future iterations of the microK family.”

6.17 Further comments from PB

The main aspects of the microK design that were intended to be improvements over IVD instruments were:
1. All solid state design to improve reliability. There are no relays (the reed relays in F18 were a source of regular failures).
2. Minimise connections and use gold finish were unavoidable (tin plated contacts in F18 often failed after a few years, also many silver plated open trim potentiometers used in F18 used, the wipers of which would fail… no potentiometers used in microK).
3. Inherent stability using substitution topology. The IVD solution argues that it is inherently stable because a transformer cannot lose or gain turns (I used this feature whilst at ASL). However, not all the flux in the primary links the secondary (even with the magnetisation windings working) and the shortfall is made up by using a tweak winding to compensate for this. Unfortunately the circuit that performs this function can itself drift with time Also. After digit 4 the ratio is realised using a MDAC, this is an old part with limited stability compared with more modern equivalents.
4. Robustness: the microK has been design to survive 240V AC as well as full 5kV transient discharges directly into the input terminals. I have tried this and it works although it did eventually fail due to a flash-over on the PCB caused (I believe) by a large transient on the mains supply. We don't advertise this design feature as we don't want customers trying it out! The F18 range were easily damaged as the front end goes straight into the matched transistor pair & other electronics. Service repairs of F18 were common compared with precisely no field failures for microK in 12 years and with a larger installed base.
5. EMC: the front end of microK has been developed to provide immunity to radiated and conducted emissions to the EU standards required for CE marking. F18 used the get out clause in the standard that allows you to cap sensitive inputs for EMC testing (i.e. the instrument isn't really tested for EMC). Users report that microK gives consistent noise levels (inherent to the bridge) where competitor (ASL, Guildline & MI) exhibit varying noise levels depending on what other equipment is in use. In theory, the IVD technology should provide better signal to noise (if input referred amplifier noise = detector noise) because it is constantly balancing rather than switching between measurement of the DUT & reference. I have been thinking about this over the weekend and have come up with a new topology that overcomes this problem. When we work on the next iteration of the microK range I'll see if this can be included in the design. Many thanks for prompting me to look again at this issue, it's been a useful exercise in trying to see microK versus IVD bridges from an outside perspective. Since it's launch the microK has steadily established itself as the instrument of choice for primary temperature metrology. Two recent comments made to me at the last TEMPMEKO conference illustrate this:
1. Greg Strouse of (a long time advocate of ASL bridge technology and a good customer when I was with ASL) chided me (tongue in cheek) because he had been telling people for many years that if they wanted the best measurement instrument that had to buy and ASL F-bridge because AAC bridges were simply better than any DC based instrument but now he was having to eat his words and recommend the DC based microK. Greg used to buy only ASL AC bridges but NIST are now on a rolling programme to replace all their F-bridges with microK.
2. Wes Tew approached me to say that he had been telling people that DC bridges would never compete with AC bridge because no one could make a DC current source that was stable enough or an ADC that was sufficiently linear and low noise but he said "I was wrong". I really appreciated his comments as (like you) he understood the magnitude of the challenge in getting a reversing DC instrument to provide performance only previously obtainable with an AC resistance bridge.

An independent assessment by NIST & NRC of a range of instruments from ASL, MI, Guildline & Isotech (Metrosol) showed the microK to exceed specification with the two microK bridges tested come out best of the ~15 tested (outperforming even the F900). I can't seem to find the report but below is a presentation that I think was based on this work. You will see that the microK 100 came in at about half its specification limit whereas the F18 bridge was twice it's limit. My experience is that F18/F900 can just about achieve their specification but require annual realignment to maintain this. On the other hand, microK is entirely consistent and needs no maintenance (there are no pots to adjust anyway). Whilst customers can send their microK back for checking (and as you observe it is included in some packages) very few do, but when they do there are absolutely no changes in performance (linearity checks are unchanged within the measurement uncertainty).

https://www.isotech.co.uk/assets/uploads/Technical%20Articles/performance-assesementbridges.pdf

Appendix: More analysis of the microK sigma-delta ADC

The summing junction and an integrator are most likely combined in a single stage, reproduced for convenience: -

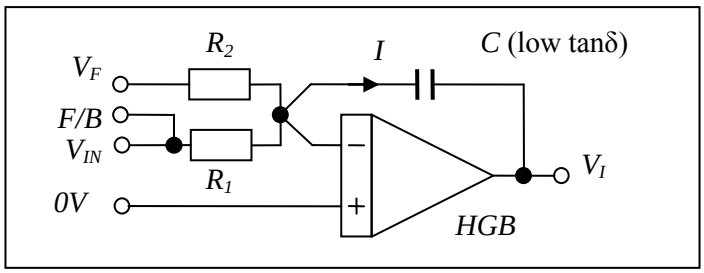


Fig. 1 The summing junction and integrator

If one assumes an ideal HGB, equal resistors, Ohm's law and the basic property of a capacitor, the current is: -

$$I = \frac{V_{IN}}{R} + \frac{V_F}{R} = \frac{V_E}{R} = -C\frac{dV_I}{dt}$$

The equivalent error voltage is V_E (see overview block diagram, fig. 5.1.2). If one also assumes that the PWM is bipolar (i.e. switches between $-V_R$ and V_R): -

First part of the cycle: $V_F = V_R \;\Rightarrow\; V_E(t) = V_1 = V_{IN} + V_R \quad \text{for} \quad 0 \le t \le t_1$

Second part of the cycle: $V_F = -V_R \;\Rightarrow\; V_E(t) = V_2 = V_{IN} - V_R \quad \text{for} \quad t_1 \le t \le T$

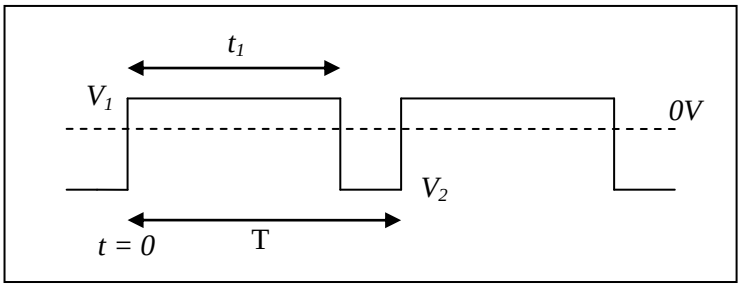


Fig 2. The error voltage as a function of time

T represents the period of a single PWM cycle (probably 10μs). Divide by C and integrate, retaining the negative sign on the LHS for convenience: -

$$-\frac{dV_I}{dt} = \frac{V_E}{RC} \;\Rightarrow\; -V_I(t) = \frac{1}{\tau}\int_0^t V_E\,d\xi \quad \text{with} \quad \tau = RC \qquad \text{(The time constant of the integrator)}$$

If one assumes that the input and reference voltages are constant.

First part of the cycle: $-V_I(t) = \frac{1}{\tau}\int_0^t V_1\,d\xi = V_1\frac{t}{\tau} + V_I(t=0) \quad \text{for} \quad 0 \le t \le t_1$

Second part of the cycle: $-V_I(t) = \frac{1}{\tau}\int_{t_1}^t V_2\,d\xi = V_2\frac{(t-t_1)}{\tau} + V_1\frac{t_1}{\tau} + V_I(t=0) \;\text{for} \quad t_1 \le t \le T$

$V_I(t=0)$ is the initial output of the integrator (at the start of the PWM cycle). We are primarily interested in the change in output: subtract $V_I(t=0)$: -

First part of the cycle: $-\Delta V_I(t) = V_1 \frac{t}{\tau}$ for $0 \le t \le t_1$

Second part of the cycle: $-\Delta V_I(t) = V_2 \frac{(t-t_1)}{\tau} + V_1 \frac{t_1}{\tau}$ for $t_1 \le t \le T$

The result, assuming no noise, is a falling output, with a constant slope for the first part of the cycle and an increasing output for the second part: -

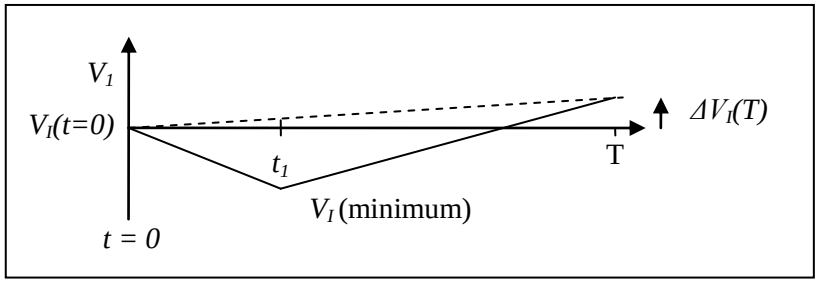


Fig. 3. A typical output of the integrator

The change at the end of the cycle is: $-\Delta V_I(T) = V_2 \frac{(T-t_1)}{\tau} + V_1 \frac{t_1}{\tau} = V_{IN} \frac{T}{\tau} + V_R \frac{(2t_1 - T)}{\tau}$

It is highly unlikely that there is no change after just one cycle – the input voltage, including noise, must be exactly equal to the average output of the PWM (one cycle). It is useful, however, to check this (special) condition: -

$$\Delta V_I(T) = 0 \quad \Rightarrow \quad V_{IN} = V_R\left(1 - \frac{2t_1}{T}\right)$$

The ratio of times is the ratio of the ADC output (n) and the total count of the PWM (N). For this special case: -

$$\frac{t_1}{T} = \frac{n}{N} \quad \Rightarrow \quad V_{IN} = V_R\left(1 - \frac{2n}{N}\right)$$

This is consistent with the aim: $V_{IN} = +5V \rightarrow -5V$ for $n = 0, 1, 2 \ldots N$ and $V_R = 5V$

It is also worth checking how the circuit would operate in practice. The minimum voltage at the output of the integrator (value at time t_1): -

$$-\Delta V_I(t_1) = V_1 \frac{t_1}{\tau} = (V_{IN} + V_R)\frac{T}{\tau}\frac{n}{N}$$

If one assumes that n is close to the correct value for the given V_{IN} (most of the time in normal operation): -

$$V_{IN} \approx V_R\left(1 - \frac{2n}{N}\right) \Rightarrow -\Delta V_I(t_1) \approx V_R\left(1 - \frac{2n}{N} + 1\right)\frac{T}{\tau}\frac{n}{N} \Rightarrow \Delta V_I(t_1) \approx 2V_R\left(\frac{n^2}{N^2} - \frac{n}{N}\right)\frac{T}{\tau}$$

This is a function of the form: $y(x) = x^2 - x$ which has a minimum at $x = 0.5$

I.e. When $\frac{n}{N} \approx 0.5$ corresponding, very approximately, to mid-range input (n = 15 or 16 and $V_{IN} \approx 0$V).

It is also reasonable to assume (see later) that the integrator time constant is about the same as the PWM cycle period (at least within a factor of ten): -

$$\frac{n}{N} \approx 0.5 \quad \text{and} \quad V_I(t=0) = 0 \quad \text{and} \quad \tau \approx T \quad \Rightarrow \quad V_I(\min) \approx -\frac{V_R}{2}\frac{T}{\tau} \approx -2.5V$$

This seems reasonable though it is clear that τ must not be much smaller than T. A larger value of τ corresponds to a lower open loop gain and reduced closed loop accuracy - a major trade-off (speed and accuracy versus overload).

The worst case happens, however, when the input changes from one extreme to the other (polarity reversal: say -5V to +5V or vice versa) at the worst possible time (at the start of a PWM cycle). The output of the PWM is $+V_R$ (+5V) for the whole cycle (to balance the previous -5V input) so that the equivalent error voltage is +10V. The slope is, from above, (assuming: $\tau = T = 10\mu s$): -

$$\frac{dV_I}{dt} = -\frac{V_E}{RC} \quad \text{with} \quad V_E = V_{IN} \pm V_R = \pm 10V$$

The worst case slope is, therefore: $$\frac{dV_I}{dt} = \frac{\pm 10V}{\tau} = \frac{\pm 10V}{10^{-5}s} = \pm 1V/\mu s$$

This is well within the capabilities of a reasonably good op-amp. The worst case output after 10µs is $\pm 10V$. This is also within the capabilities of the first stage (with ±15V PSU) but it will almost certainly overload the second and third stage (one-plus-integrators) and flash ADC. The loop must be designed to recover from such overload conditions, in a controlled way, and the timing/control/data collection algorithm should allow for the extra delay: possibly more than the response of the ideal (linear) loop response.

More generally the input voltage can be anywhere within the range. One can describe the difference between the input and the PWM state with a parameter α_k. At the start of a measurement cycle (initial state n_0): -

$$V_{IN} = V_R\left(1 - \frac{2(n_0 + \alpha_0)}{N}\right)$$

If the input voltage remains constant then: $$n_0 + \alpha_0 = n_k + \alpha_k \quad \text{with} \quad k = 0, 1, 2, \ldots$$

The index *k* represents the number of PWM cycles. After the first PWM cycle, from above: -

$$-\Delta V_I(T) = V_{IN}\frac{T}{\tau} + V_R\left(\frac{2t_1 - T}{\tau}\right) \quad \Rightarrow \quad -\Delta V_I(T) = V_R\frac{T}{\tau}\left(1 - \frac{2(n_0 + \alpha_0)}{N} + \frac{2t_1}{T} - 1\right)$$

This simplifies nicely: $$\frac{t_1}{T} = \frac{n_0}{N} \qquad \Rightarrow \qquad \Delta V_I(T) = 2V_R\frac{\alpha_0}{N}\frac{T}{\tau}$$

The (fast) change in integrator output passes straight through the one-plus-integrators to the flash ADC input.

The change at the flash ADC input, in terms of the average voltage interval (one bit) for adjacent states, at the end of the first PWM cycle is, therefore: -

$$\Delta n_1 = n_1 - n_0 \approx \Delta V_I(T) \div \frac{2V_R}{N} \approx \alpha_0\frac{T}{\tau}$$

More generally: $\Delta n_k = n_k - n_{k-1} \approx \alpha_{k-1}\dfrac{T}{\tau}$

a). The approximation is due to the fact that the intervals are unlikely to be exactly the same – it depends on the accuracy of the flash ADC. An accuracy of 1% should be good enough and is easily achieved [1].

b). If α_0 is numerically positive then the input voltage corresponds to a state above n_0. The output of the integrator increases taking the input to the ADC in the right direction (increasing n).

c). The size of the change depends on the integrator time constant, relative to the PWM cycle period. If one chooses them to be about the same $(\tau \approx T)$ the change at the ADC input corresponds to very nearly the same as the initial error. The first PWM cycle should be sufficient, therefore, to reduce the error to less than one (least significant) bit of the PWM $(-1 < \alpha_1 < +1)$. The advantage would seem to be speed – rapid convergence to approximate null balance. The main disadvantage is instability – once the error has reduced to less than one bit the loop oscillates erratically, successively overshooting and/or undershooting state thresholds. This problem is worst at PWM threshold boundaries, even for longer time constants.

d) If one chooses a longer time constant $\tau >> T$ and a high resolution ADC/PWM pair the result is a quasi-exponential convergence [2]. An exponential converges to within 10ppb in around 20 time constants.

e) Large transients (polarity reversals and substitutions) could be avoided with a smart controller – loading the PWM at the start of each "*X*" (50ms) measurement period with the previously remembered value.

f) Once the error α reduces to less than one interval the loop oscillates to produce a lower frequency triangle wave superimposed on the higher frequency triangle wave (at the integrator output). If the frequency of oscillation is sufficiently high it passes through the one-plus-integrators largely unaffected and, at the output of a track/hold circuit, the signal would more closely resemble a step-wise sawtooth (see fig. 5). The shape and frequency depend on the input voltage relative to the PWM states.

g). With a sufficiently large number of PWM cycles the extremely high gain (at low frequency) before the flash ADC ensures that the average output of the PWM is the same as the average input voltage over the same time period. With a three stage high gain block (total gain of $>10^{15}$ at <10Hz), for example, an error of 5V at the flash input corresponds to 5×10^{-15}V referred to the input! The loop in effect "counts" the total number of $+V_R$ and $-V_R$ quanta.

h). Later models probably have a minimum ADC "*X*" measurement period of 50ms (i.e. 5,000 PWM cycles of 10μs). Settling time is probably of the order 10ms so that 4,000 samples are available for calculating the average. The current is then reversed and the measurement repeated. The second resistor is then substituted and, with a current reversal, the minimum ratio measurement time is 200ms. The process is then repeated ten times for the minimum measurement time quoted of 2s "per channel" (five times for 1s with computer interface) [3]. For higher resolution and accuracy the total measurement time is quoted as up to 50s (a total of 5 million samples of which 4 million (80%) are probably useable).

1. PB confirmed. See section 6.16.
2. A quasi-linear system?
In the limit $N \to \infty$ the variables n and α become continuous. The change after each PWM cycle is: -

$$n \to n + \Delta n \;\Rightarrow\; \alpha \to \alpha - \Delta\alpha \;\Rightarrow\; \Delta\alpha = -\alpha\frac{T}{\tau}$$

If one also considers a large number of PWM cycles with a relatively large integrator time constant (a timescale long compared to the PWM cycle period) and approach the infinitesimal limit: -

$$\tau >> T \text{ and } T \to dt \;\Rightarrow\; \Delta\alpha \to d\alpha \;\Rightarrow \frac{d\alpha}{\alpha} = -\frac{dt}{\tau} \;\Rightarrow\; \alpha(t) = \alpha(0)\exp\left(-\frac{t}{\tau}\right)$$

We have, therefore, a reasonably accurate model of a quasi-continuous feedback system, driving the output of the ADC, with each PWM cycle, to a null balance with an exponentially decaying error.
3. Isotech brochure: "The new microK family of precision thermometry bridges". Edition 2 (1516).

The lower frequency triangle wave starts from below the *n*-state threshold until it reaches above the upper threshold. The PWM state then changes to *n*+1, the error becomes $1-\alpha$ (i.e. negative) and the integrator ramps down until it falls below the lower threshold and the state reverts to *n*. The cycle then repeats. If $\alpha \approx 0.5$ (about half way between PWM states) with a long integrator time constant (say $\tau = 10T$) the sawtooth is roughly symmetrical and there is little overshoot or undershoot (see fig. 4).

With $\alpha \approx 0.5$ and $\tau = 10T$ the input to the flash ADC would look something like the following. The dashed line corresponds to the actual input voltage (half way between thresholds *n* and *n* + 1): -

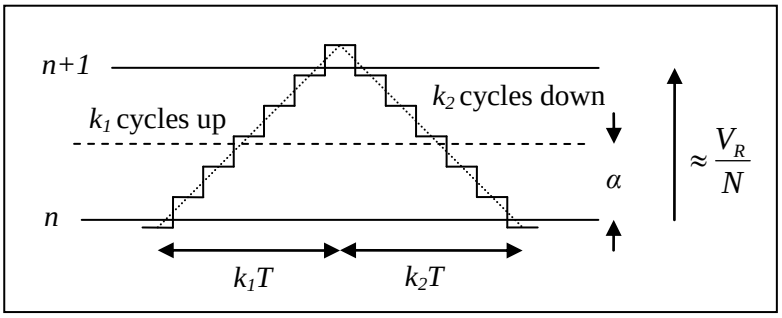


Fig. 4 Oscillation between states at the output of a track/hold circuit

If α is not too small it is possible to estimate the period and frequency of the oscillation: -

The number of PWM cycles required for each part of the sawtooth is: $k \approx \frac{2V_R}{N} \div \Delta V_I$

Referring to fig. 5, therefore, the numbers of cycles for the up ramp and down ramp are, respectively: -

$$k_1 \approx \frac{\tau}{\alpha T} \text{ and } k_2 \approx \frac{\tau}{(1-\alpha)T}$$

The period of oscillation is: $k_1 T + k_2 T = \frac{\tau}{\alpha} + \frac{\tau}{1-\alpha} = \frac{\tau}{\alpha(1-\alpha)}$

The average value from the data stream is, as expected: $\langle n \rangle = \frac{nk_1T + (n+1)k_2T}{k_1T + k_2T} = n + \frac{k_2}{k_1 + k_2}$

$$\text{but: } \frac{k_2}{k_1 + k_2} \approx \alpha \Rightarrow \langle n \rangle \approx n + \alpha$$

The average state is close to the correct value within each cycle of oscillation. This looks promising as a large number of such cycles provide an even closer approximation to the required average value.

Unfortunately, with a long integrator time constant, the maximum frequency is below the maximum practicable one-plus-integrator frequency (typically 16kHz [1]) and the integrator signal does not pass through but undergoes two more integrations, reaching the upper and lower thresholds much more quickly. Fig. 5 is no longer valid and the analysis is far more complicated [2]. For example: -

The maximum frequency is when $\alpha = 0.5$ so that with $\tau = 10T = 100\mu s$: $f_{MAX} \approx \frac{0.25}{10^{-4} s} = 2.5kHz$

1. A higher frequency increases loop gain and improves closed loop speed/accuracy. See section 5.2.
2. PB insists “four integrators” relying on computer simulation, apparently. See section 6.16

With smaller and large values of α the frequency is even lower. **This would suggest that a lower value of integrator time constant is preferable**. If α is small (the input is close to a PWM state), for example, it takes a large number of PWM cycles to ramp up to the next state but only one or two to ramp back down again. The resulting frequency is higher (except for values of α close to 0 or 1). The sawtooth is highly asymmetric and there is a strong possibility that the down ramp overshoots (see spreadsheet simulation fig. 5). The frequency and shape of the signal is more complicated and difficult to predict but, at least, the oscillation is mostly at high frequency and the low frequency model of the loop is largely valid. In practice low frequency noise is also present (greatly amplified) which the loop attempts to correct (equivalent to a varying input signal). Mid and high frequency noise probably cause occurrences of PWM states far away from the states either side but the loop should be able to cope with this – unless the ADC input exceeds its range (overload).

It is surprisingly easy to model the (noiseless) oscillation with a spreadsheet.

The first row contains initial conditions (state = 1 (up ramp), $V = 0$)

Column A: 0-100 iterations for horizontal axis

Column B records the current state (1 for ramp up or 0 for ramp down), which is copied from the row above (previous state) column D.

Column C: Calculate new voltage: =IF(B6=1,C5+$G$4/$H$4,C5-(1-$G$4)/$H$4)

Column D: Calculate new state: =IF(B5=1,IF(C5>1,0,1),IF(C5<0,1,0))

Cells G4 and H4 hold values for α and τ/T taken from slider controls.

In the following example $\alpha = 0.1$ and $\tau/T = 1$. Note the large undershoot on the down ramp.

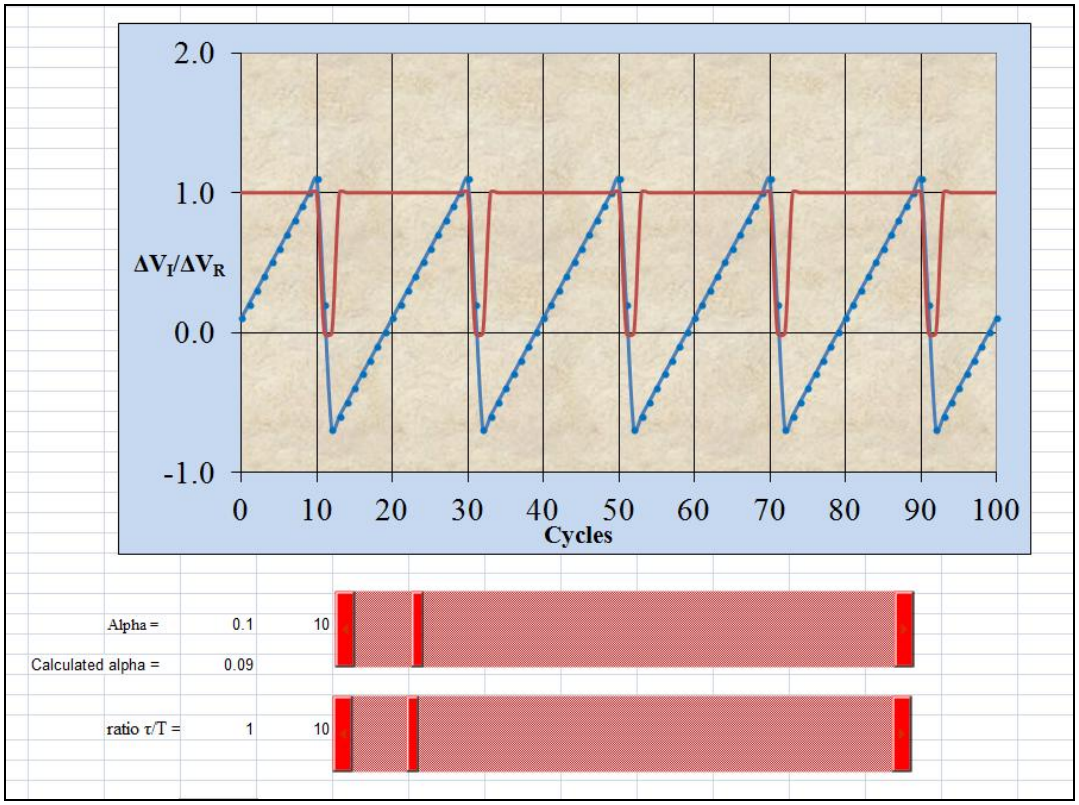


Fig. 5 A spreadsheet simulation

The calculated average value of alpha (100 PWM cycles) corresponds reasonably well with input alpha.

JFET Theory

1. Introduction

An N-type JFET is basically a thin conduction channel of N-type semiconductor material embedded in a P-type substrate forming a high quality PN junction. At each end of the conduction channel connections are brought to the surface with N-type material (drain and source). A single connection is made to the substrate (the gate). Whereas some devices are precisely symmetrical with regard to drain and source others are designed to minimise on-chip capacitance between the gate and drain. At low frequency the drain and source are usually interchangeable.

An insulating depletion region is formed at the PN junction. This acts like an insulating dielectric whose thickness depends on the voltage difference between the channel and gate. When a current flows through the channel the voltage difference varies (down the channel) and the shape of the depletion region varies accordingly. The drain-source current is, therefore, a function of three voltages: drain, gate and source.

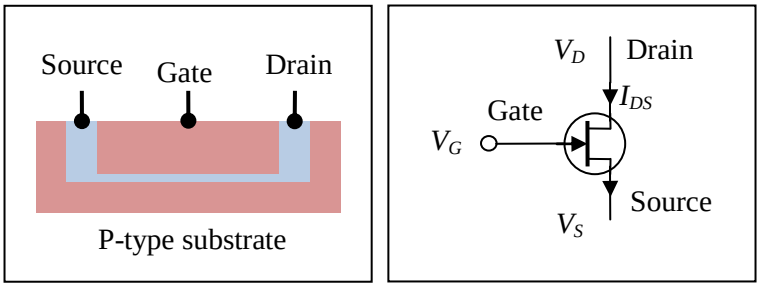


Fig. 2.1.1 A basic model for the JFET (N-type)

In most applications the PN junction is reverse biased and the current flowing in or out of the gate is negligible (a few pA at ambient temperature). The current flowing "down the drain" is almost exactly the same as that flowing "from the source" (hence the silly names).

JFETs have two modes of operation: voltage controlled resistor (VCR) and "pinched-off" mode, depending on the channel-gate voltage distribution. The latter is most frequently employed in amplifiers as voltage gain is maximised. With a low channel-gate voltage the conduction channel behaves like a resistor. As the channel-gate voltage increases, however, the conduction channel becomes very thin, the voltage-current relationship becomes non-linear and, at sufficiently high voltage, the channel is pinched off and the JFET behaves like a (gate controlled) constant current source. The channel connection labelled drain is invariably chosen to be the most positive and it is the drain end of the channel that is pinched off. The current is deemed positive when flowing from drain to source.

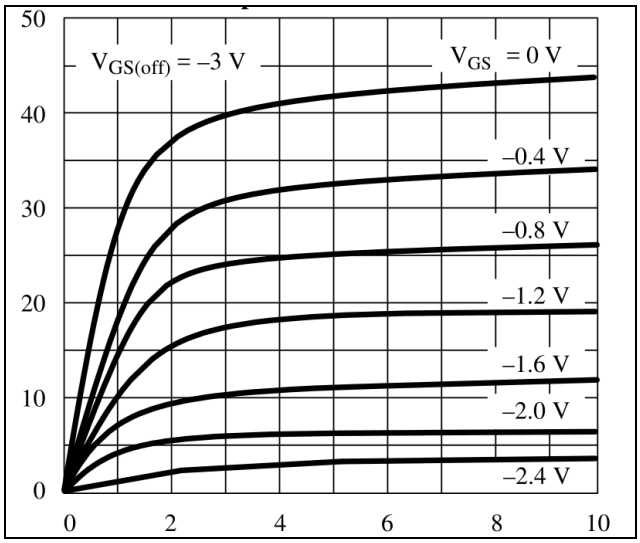


Fig. 1.1 Drain-source current (mA) versus drain-source voltage (U430)

It is remarkable that such a complex system can be described, reasonably accurately, with a fairly simple theory, based on a few plausible assumptions….

2. The theory

2.1 Basic theory of the PN junction

N-type silicon is pure crystalline silicon which is doped with a tiny proportion of impurity atoms which, in terms of potential energy, have one electron too many. These electrons prefer to wander free between the silicon atoms – able to conduct electricity. The impurity atoms become positively charged but the overall charge imbalance in each small region is still zero (except at the junction). P-type material is the opposite – the impurity atoms are hungry for an extra electron. Both types of doped silicon are fairly good electrical conductors.

At a PN junction electrons desert their N-type impurity atoms and stick to neighbouring P-type impurity atoms creating a charge imbalance – just like in a charged capacitor. Energy equilibrium is reached when the electric field created by the charge imbalance counteracts the effect by attracting electrons back to the N-type material. The result is a thin layer either side of the junction where the electrons are strongly bonded and are no longer free to conduct electrical current. If the P-type anode is made more negative than the N-type cathode the energy balance shifts and the depletion region grows thicker. The charge increases and the junction behaves like a (voltage dependent) capacitor [1]. If the anode is made more positive electrons are attracted away from the depletion region but are quickly replaced by more electrons from the N-type side of the junction resulting in a current flowing.

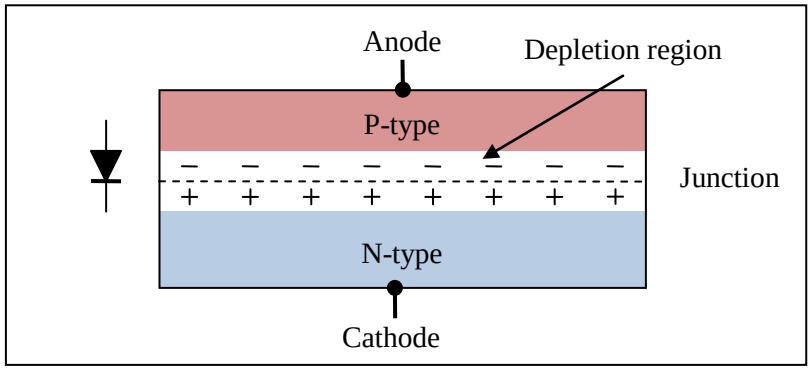


Fig. 2.1.1 The PN junction and depleted region either side (not to scale)

If the drain and source are connected together (cathode) a JFET acts like a high quality diode with very low reverse bias leakage and accurate conformity to the theoretical (thermodynamic) formula for a diode: -

$$V_{BE} = V_T \ln\left|\frac{I_B}{I_S}\right| \qquad \text{or} \qquad I_B = I_S\left\{\exp\left(\frac{V_{BE}}{V_T}\right) - 1\right\}$$

I_S = A small current depending on the doping level of the silicon.

$k = 1.38\times10^{-23}\,J/K$ is Boltzmann's constant.

T = Absolute temperature (approximately 293K at 20 ºC)

$e = 1.6\times10^{-19}C$ is the quantum of electrical charge (e.g. on a proton)

$V_T = \frac{kT}{e} \approx 25mV$ (at 20ºC) is a voltage corresponding to the thermal energy of the electrons.

1. JFET diodes perform well as voltage controlled capacitors (varactor) in certain applications.

2.2 The model

I shall assume that the N-type strip is uniformly rectangular and uniformly doped. In practice, for low noise JFETs (low resistance), the strip is much longer than it is thick and much wider than it is long. In the following diagram the vertical scale is much exaggerated: -

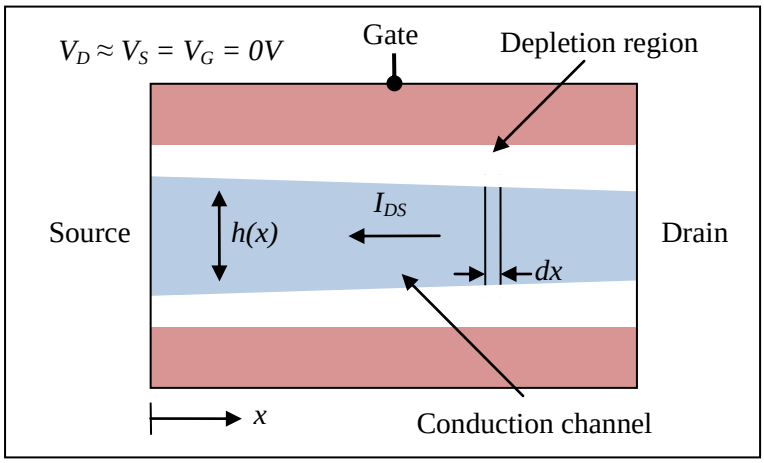


Fig.2.2.1 Simplified schematic of the conduction channel in low resistance mode

The boundary between the conducting channel and the depletion region is fuzzy. There is a gradual transition from conductor to insulator and the dimensions of the conduction channel are not precisely defined. Similarly, the connections at each end alter the effective length. Fortunately this fuzziness does not prevent one from assuming a simple geometry, with effective dimensions, as the following analysis demonstrates: -

Define: -
W = Effective width of the channel (constant)
L = Length of the channel (constant)
$h(x)$ = Effective thickness of the channel at position x (variable, depending on the channel-gate voltage)

With: $$W >> L >> h(x)$$

x = Distance from the source end
$\rho(y,z)$ = Free charge density

For N-type JFETs the free charges are electrons and the free charge density is numerically negative. The free charge density is uniform throughout the channel except at the boundary with the depletion region where it gradually drops to zero. It is also a function of temperature as the free electrons are a result of thermal equilibrium. Any electrons added to the channel are quickly absorbed as equilibrium is restored (and vice versa). From definitions and basic principles: -

The current density in the channel is: $\vec{J} = \rho\vec{v} = \sigma\vec{E}$ (Ohm's law)

Where: σ = Conductivity of the channel (numerically positive, by convention).

The conductivity and free charge density are related by the electron mobility: $\mu = \dfrac{\sigma}{\rho} \Rightarrow \sigma = \mu\rho$

The electron mobility is, therefore, numerically negative.

Usually, the drain is more positive than the source and the free electrons flow from source to drain. The current, by convention, is positive in the opposite direction.

I shall assume that the only non-zero components of the vectors are the *x* components. This seems reasonable from the simple geometry. Similarly, the free charge density is independent of *x*. The current density is, therefore: -

$$J_x = \mu\rho(y,z)\frac{dV(x)}{dx} \quad \text{with } V(x) = \text{potential in the channel and } dV = V(x+dx) - V(x)$$

The current passing through any plane (perpendicular to the channel) must be the same (independent of *x*) otherwise there would be a build up of charge in the channel. The Coulomb force is so strong that any charge imbalance, apart from that maintained by the potential gradient in the channel, would be dispersed rapidly, at least on the timescale of a low frequency model. The current density and, therefore, the potential gradient must vary down the channel.

In general, through the plane at position *x*: $$I_{DS} = \iint J_x(x,y,z)dydz = \iint \mu\rho(y,z)\frac{dV(x)}{dx}dydz$$

The channel potential gradient is a function of *x* but independent of *y* and *z* and so one can take it outside the integral. The double integral yields the free charge per unit length, which is also a function of *x*: -

$$I_{DS} = \mu\frac{dV(x)}{dx}\iint \rho(y,z)dydz = \mu\frac{dV(x)}{dx}\frac{dQ(x)}{dx}$$

Multiply both sides by *dx* and integrate from source to drain: $$\int_0^L I_{DS}dx = I_{DS}L = \mu\int_S^D \frac{dQ(x)}{dx}dV$$

The limits of the integration are from the lower to the higher voltage (usually source to drain) so that the current is positive when flowing in the direction of drain to source.

Note also, by definition: $$dV = V(x+dx) - V(x) \quad \Rightarrow \quad dV_{CG} = dV$$

Where dV_{CG} is the channel-gate voltage, which is important for the following: -

The last integral is not simple as dQ/dx depends on the potential in the channel. As the voltage in the channel increases, relative to the gate, more electrons are taken from the depletion region leaving more positively charged atoms. The electrons exit the channel via the (usually) more positive drain electrode. **In general one can model the junction as an infinite array of infinitesimal capacitors with a common connection to the gate and with the same voltage profile as the channel.** The channel becomes thinner but the free charge density remains the same (thermal equilibrium) and the total charge available for conduction reduces (becomes less negative). Each infinitesimal capacitance is also a function of the channel-gate voltage so that the change in free charge per unit length for each capacitor is: -

$$\frac{dQ}{dx} = \frac{Q_0}{L} + V_{CG}\frac{dC(V_{CG})}{dx}$$

Where Q_0 = total free charge when $V_{CG} = 0V$ all along the channel (i.e. a constant for a given device at a given temperature): -

$$\Rightarrow \quad I_{DS} = \frac{\mu}{L^2}\int_S^D \left\{Q_0 + LV\frac{dC}{dx}\right\}dV$$

The integral of the first term is simple but the second is less obvious. The dummy variable, *V*, refers to the channel-gate voltage. The general expression for drain-source current is, therefore: -

$$I_{DS} = \frac{\mu}{L^2}\left\{Q_0V_{DS} + L\int_S^D V\frac{dC}{dx}dV\right\}$$

2.3 Drain-source on resistance

One of the most useful parameters is the drain-source on resistance, measured at very low current (typically 100μA or 1mA with V_{DS} < 100mV). This is easily and quickly measured and correlates closely with other parameters – it is especially useful for selecting matching pairs from a number of single devices. The usual method for calculating resistance is based on the length, cross-sectional area and conductivity of the conductor. The subscript 0 indicates the effective value of the parameter under the test conditions: -

In the limit: $V_D = V_S = V_G = 0V$ $\qquad R_0 = \dfrac{V_{DS}}{I_{DS}} = \dfrac{L}{\sigma W h_0}$

Unfortunately σ and h_0 are not easy to calculate or analyse but, fortunately, one has the capacitor hypothesis. From above, repeated for convenience: -

$$I_{DS} = \frac{\mu}{L^2}\left\{Q_0 V_{DS} + L\int_S^D V\frac{dC}{dx}dV\right\}$$

In the limit $V_D = V_S = V_G = 0V$ the integral is zero: -

$$V_{CG} = 0 \quad \Rightarrow \quad I_{DS} = \frac{\mu Q_0}{L^2}V_{DS} \quad \Rightarrow \quad R_0 = \frac{V_{DS}}{I_{DS}} \approx \frac{L^2}{\mu Q_0}$$

In practice the resistance remains constant (the plot of current versus drain voltage is approximately linear) for a drain-source voltage of up to 1V (see fig. 2.3.1 for the case V_{GS} = 0V).

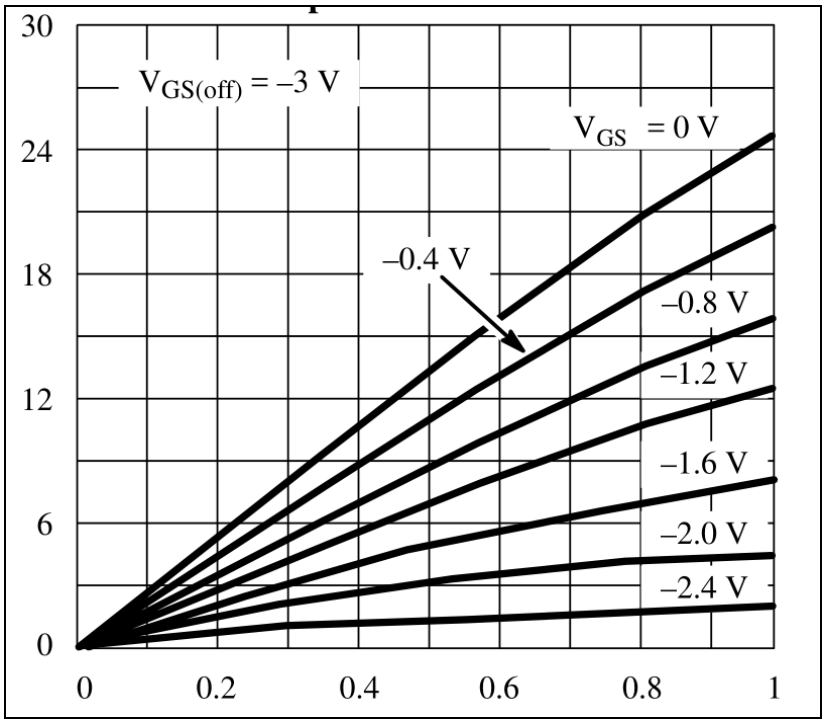


Fig. 2.3.1 Drain-source on resistance (top plot, Siliconix U430)

N.B. For an N-type JFET both Q_0 and μ are numerically negative so that the conductivity and resistance are always numerically positive.

2.4 Voltage controlled resistor mode

If the voltage on the gate is reduced, with drain and source held at about 0V, the thickness of the channel reduces but remains approximately uniform for the full length. The resistance increases and one has a voltage controlled resistor (VCR): -

$$V_D \approx V_S \approx 0V \quad \text{and} \quad V_G < 0V \quad \Rightarrow \quad R(V_{SG}) = \frac{L}{\sigma W h(V_{SG})}$$

Once again I shall employ the capacitance hypothesis. From above, repeated for convenience: -

$$I_{DS} = \frac{\mu}{L^2}\left\{ Q_0 V_{DS} + L\int_S^D V \frac{dC}{dx} dV \right\}$$

The channel-gate voltage is approximately constant down the channel and the capacitance, as well as charge, is distributed uniformly. dC/dx is constant and equal to the average capacitance per unit length (C_J/L): -

$$V_D \approx V_S \approx 0V \text{ and } V_G < 0V \quad \Rightarrow \quad L\int_S^D V \frac{dC}{dx} dV \approx L V_{SG} \frac{dC}{dx} \int_S^D dV = V_{SG} C_J V_{DS}$$

C_J is the capacitance of the junction at the specified channel-gate operating voltage. Obviously one could replace V_{SG} with V_{DG} as they are assumed to be the same. The current is, therefore: -

$$I_{DS} \approx \frac{\mu}{L^2}(Q_0 + V_{SG} C_J) V_{DS}$$

As the channel-gate voltage increases all the free electrons are removed and the current falls to zero. This happens at the "pinched-off" voltage, V_P. A reasonable model is, therefore: -

$$I_{DS} \approx \frac{\mu Q_0}{L^2}\left(1 + \frac{V_{SG}}{V_P}\right) V_{DS} \quad \text{with} \quad V_P = \frac{Q_0}{C_P} \quad \text{and} \quad C_P = C_J(V_P)$$

The resistance is, therefore:

$$R_{DS} = \frac{V_{DS}}{I_{DS}} = \left(\frac{V_P}{V_P + V_{SG}}\right) R_0 \quad \text{with} \quad R_0 = \frac{L^2}{\mu Q_0}$$

It is usual to specify the gate voltage relative to the source: -

$$V_{SG} = -V_{GS} \quad \Rightarrow \quad R_{DS} \approx \left(\frac{V_P}{V_P - V_{GS}}\right) R_0$$

As V_{GS} reduces to the (numerically negative) value of V_P the resistance increases to infinity. In practice there is not a sharp cut-off as the channel tends to leak current until the gate voltage is substantially lower than the "pinch-off" voltage (e.g. when used as an analogue switch). The resistance remains reasonably constant for a drain-source voltage in the range ±1V, depending on the type of JFET. See fig. 2.3.1 for various values of V_{GS}.

The equation for R_{DS} is only approximate as the value of V_P varies slightly with V_{GS}.

The lowest resistance is obtained with the gate-channel voltage slightly positive (i.e. the PN junction is slightly forward biased). Obviously there is a limit as the diode begins to conduct and a maximum of 200mV is typical.

A more in-depth analysis shows that the range of drain and source voltage can be extended….

2.5 The general case

As the current and drain voltage increases the drain-gate voltage approaches pinch-off and the dynamic resistance $\left(dV_{DS}/dI_{DS}\right)$ increases. The plot of current versus voltage becomes non-linear (see fig. 1.1).

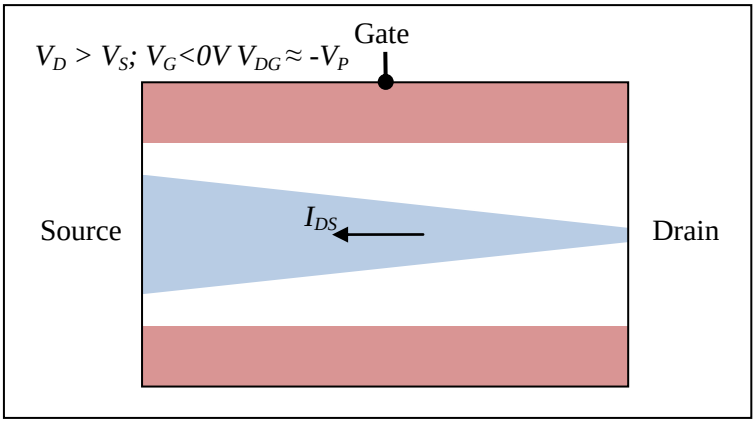


Fig. 2.5.1 The drain-gate voltage approaches pinch-off

Generally, repeated for convenience: $$I_{DS} = \frac{\mu}{L^2}\left\{Q_0 V_{DS} + L\int_S^D V\frac{dC}{dx}dV\right\}$$

In the general case the integral requires careful examination. One can employ integration by parts: -

A basic rule of differentiation is: $$d(FG) = FdG + GdF \quad \Rightarrow \quad FG = \int dGF + \int dFG$$
$$\Rightarrow \quad \int dFG = FG - \int dGF$$

In this case $F = \frac{V^2}{2}$ and $G = \frac{dC}{dx}$ so that: $$\int V\frac{dC}{dx}dV = \frac{V^2}{2}\frac{dC}{dx} - \int\frac{V^2}{2}d\left(\frac{dC}{dx}\right)$$

The first term is significant as the voltage varies down the channel. The second, on the other hand, is a small correction. **The charge distribution and the thickness of the depletion region may vary significantly down the channel but I shall assume that the capacitance gradient does not.**

$$\frac{dC}{dx} \approx \text{ constant } \Rightarrow \quad L\frac{dC}{dx} \approx C_J \quad \text{and } d\left(\frac{dC}{dx}\right) \approx 0$$

To a good approximation, therefore: $$I_{DS} \approx \frac{\mu}{L^2}\left\{Q_0 V_{DS} + C_J\frac{V_{DG}^2 - V_{SG}^2}{2}\right\}$$

Alternatively: $$I_{DS} \approx \frac{1}{R_0}\left\{V_{DS} + \frac{V_{DG}^2 - V_{SG}^2}{2V_P}\right\} \text{ with } R_0 = \frac{L^2}{\mu Q_0}$$

This equation is only valid for a range of voltages which do not pinch off the channel completely – once the current drops to zero it remains zero as the gate-channel voltage becomes more negative. A little more algebra reveals the symmetrical form (drain and source are interchangeable): -

$$I_{DS} \approx \frac{1}{2V_P R_0}\left\{\left(V_P + V_{DG}\right)^2 - \left(V_P + V_{SG}\right)^2\right\}$$

A little more algebra reveals an extended voltage controlled resistor mode: -

$$V_{DG}^2 - V_{SG}^2 = (V_{DG} + V_{SG})(V_{DG} - V_{SG}) \;\Rightarrow\; I_{DS} \approx \frac{1}{R_0}\left\{V_{DS} + \frac{(V_{DG} + V_{SG})(V_{DG} - V_{SG})}{2V_P}\right\}$$

Also: $$V_{DG} - V_{SG} = V_{DS} \;\Rightarrow\; I_{DS} \approx \frac{V_{DS}}{R_0}\left\{1 + \frac{(V_{DG} + V_{SG})}{2V_P}\right\}$$

The conductance is, therefore: -

$$\Rightarrow\; \frac{1}{R_{DS}} = \frac{I_{DS}}{V_{DS}} \approx \frac{1}{R_0}\left\{1 + \frac{(V_{DG} + V_{SG})}{2V_P}\right\}$$

Also: $$V_{DG} = V_D - V_G \;\text{ and }\; V_{SG} = V_S - V_G \;\Rightarrow \frac{1}{R_{DS}} \approx \frac{1}{R_0}\left\{1 + \frac{(V_D + V_S - 2V_G)}{2V_P}\right\}$$

In the last section it was assumed that $V_D \approx V_S \approx 0V$ so that $V_G \approx V_{GS}$ and the resulting formula is approximately the same. This is usually the case as the source is connected to 0V and the drain signal is kept small. Also, the dependence on drain voltage can be reduced by adding a proportion of it to the control signal. The most frequently cited variable attenuator circuit is, for example [1]: -

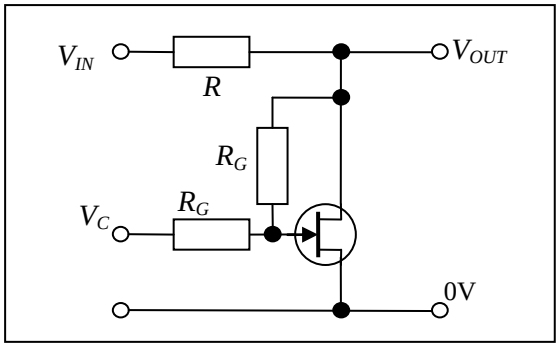


Fig. 2.5.2 VCR linearisation

The formula for conductance simplifies: -

$$V_G = V_C + \frac{V_D - V_C}{2} \;\text{ and }\; V_S = 0V \;\Rightarrow\; \frac{1}{R_{DS}} \approx \frac{1}{R_0}\left\{1 - \frac{V_C}{2V_P}\right\}$$

The circuit works best when the gate resistors are much larger than the output resistance of the attenuator. A factor of ten (at least) is advised [1]: -

$$R_G >> R \parallel R_{DS(MAX)}$$

The general case also reveals an alternative approach. If the circuit is arranged so that drain and source voltages are equal and opposite, relative to some reference level (typically local 0V), then the linear range is significantly extended (doubled – both drain and source ends of the channel can approach pinch-off). For example: -

$$V_D + V_S = 0V \;\Rightarrow\; \frac{1}{R_{DS}} \approx \frac{1}{R_0}\left\{1 - \frac{V_G}{V_P}\right\}$$

1. Siliconix Inc: “Small –Signal FET Data Book”. January 1986.
See section 7-75 “FETs as Voltage-Controlled Resistors”

2.6 Pinched-off mode

In most amplifier applications the drain-gate voltage is higher than that required for pinch-off. Part of the channel, at the drain end, becomes very narrow but the current is not completely shut off. If it were the voltage drop down that part of the channel would bring the channel-gate voltage down and the channel would expand. A stable equilibrium is achieved and the result is high transconductance and quite a high drain resistance (low drain conductance). The current is controlled almost entirely by the gate-source voltage and is largely independent of the drain-source voltage. The JFET becomes a voltage controlled current source.

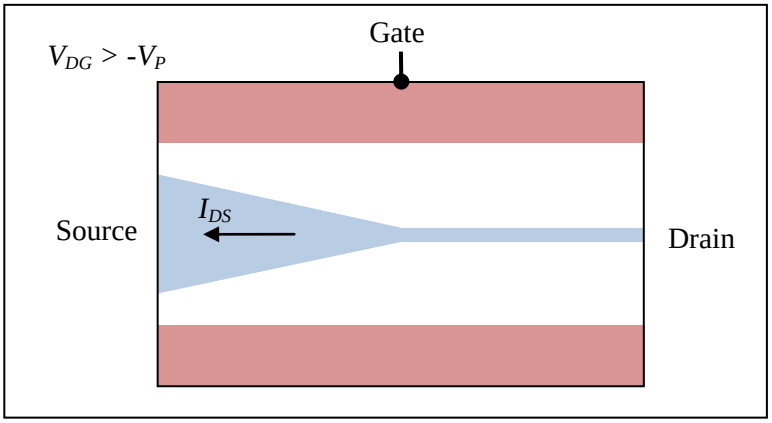


Fig. 2.6.1 Operation in pinched-off mode

From the previous section: $$I_{DS} \approx \frac{1}{2V_P R_0}\left\{ (V_P + V_{DG})^2 - (V_P + V_{SG})^2 \right\}$$

In voltage controlled resistor mode the active region is when V_{DG} and V_{SG} are both positive and each of the components in brackets are negative (V_P is negative). In pinched-off mode the contribution to the equation from the drain-gate effect falls to zero and remains so as V_{DG} increases. Remarkably, the contribution from the source-gate element remains the same.

In pinched-off mode: $$V_P + V_{DG} \geq 0 \quad \Rightarrow \quad I_{DS} \approx \frac{1}{2V_P R_0}\left\{ -(V_P + V_{SG})^2 \right\}$$

The pinch-off voltage is numerically negative and the drain-source current is, therefore, positive. The result is a square law characteristic with k and V_P approximately constant: -

$$V_P + V_{SG} \leq 0 \quad \Rightarrow \quad I_{DS} \approx k(V_{SG} + V_P)^2 \qquad \text{and} \qquad V_P + V_{SG} \geq 0 \quad \Rightarrow \quad I_{DS} = 0$$

Take the square root for a linear equation: $$\sqrt{I_{DS}} \approx \pm\sqrt{k}(V_{SG} + V_P)$$

Plots of measured data of $\sqrt{I_{DS}}$ versus V_{SG} have a negative slope (see fig. 2.6.2) and one must, therefore, take the negative square root. Equally valid, preferred by some, is a form of the equation with gate-source voltage as the chosen variable: -

$$V_{GS} \geq V_P \quad \Rightarrow \quad I_{DS} \approx k(V_{GS} - V_P)^2 \qquad \text{and} \qquad V_{GS} \leq V_P \quad \Rightarrow \quad I_{DS} = 0$$

In this form both V_{GS} and V_P are numerically negative but $V_{GS} - V_P$ is numerically positive in the operating range.

A batch of ten J112 devices provided seven with very low noise and a couple of reasonably well matched pairs: -

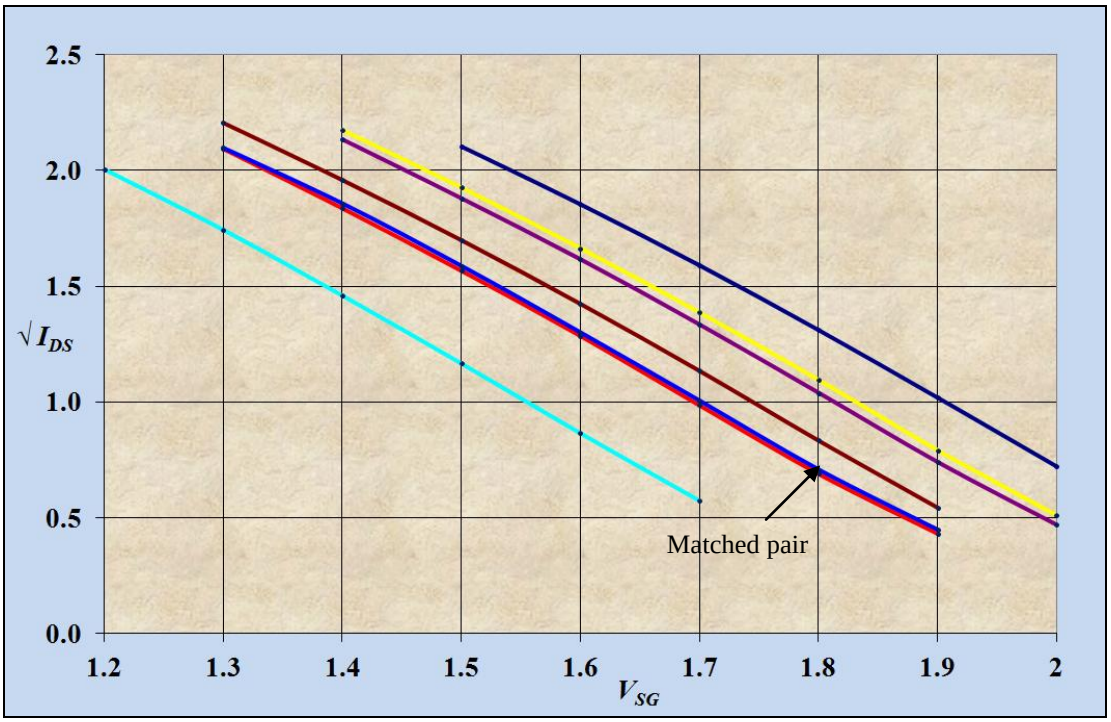


Fig. 2.6.2 Linearised plot for a batch of J112 JFETs (I_{DS} in mA)

The theoretical model has its limitations: -

$$\sqrt{I_{DS}} = -\sqrt{k}\left(V_{SG} + V_P\right)\left(1+\Delta\right) \quad \text{with} \quad \Delta << 1$$

The factor $\left(1+\Delta\right)$ represents a small correction. A measure of the accuracy of the theory is a plot of difference between measured data and the best fit straight line. The error is clearly systematic and less than ±2% over quite a wide range of current: -

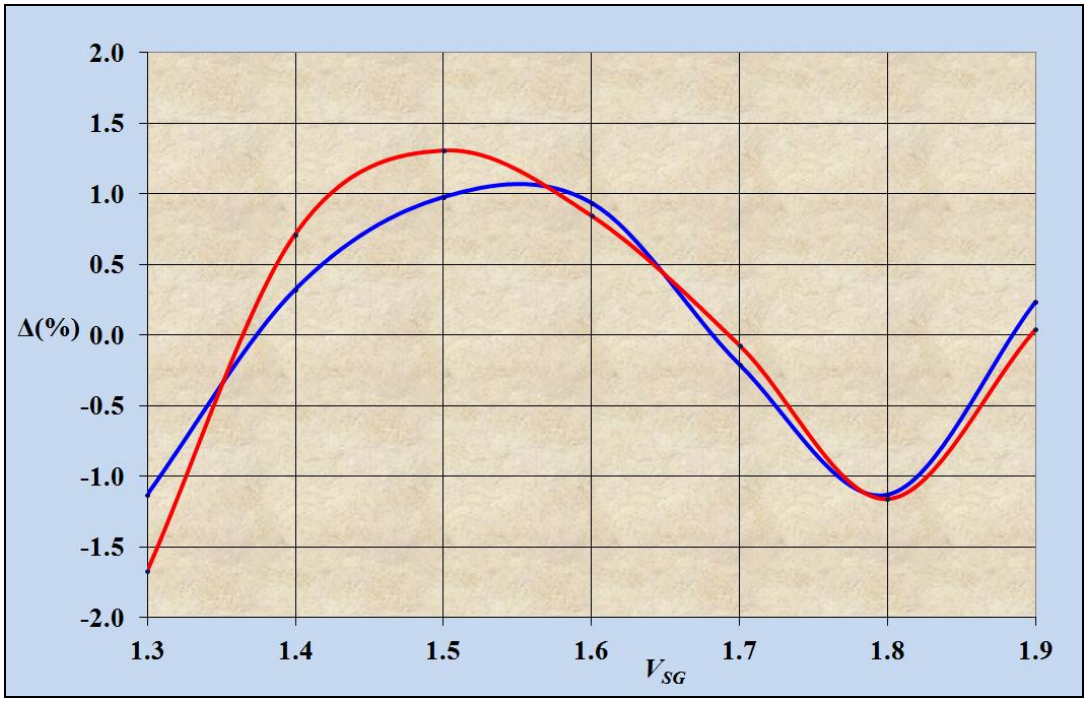


Fig.2.6.3 Deviation from best fit linear model $\sqrt{I_{DS}} \propto V_{SG}$

Most text books and application notes prefer the form: $I_{DS} \approx I_{DSS}\left(1-\frac{V_{GS}}{V_P}\right)^2$

Both V_{GS} and V_P are numerically negative with the ratio typically varying from 0 to 1. I_{DSS} is the drain-source saturation current, obtained when the gate-source voltage is zero. Manufacturers often specify minimum and maximum values for I_{DSS} and V_P despite the fact that they are notoriously variable, even for devices of the same type. This is not particularly helpful to the designer. It is easy to relate, with a little algebra, the relationship between parameters, resulting in some design rules of thumb: -

$$k=\frac{-1}{2V_P R_0} \quad I_{DSS}=kV_P^2=\frac{-kV_P}{2R_0} \quad \text{with} \quad R_0=\frac{L}{\sigma W h_0}=\frac{L^2}{\mu Q_0} \quad \text{and} \quad V_P=\frac{Q_0}{C_J(V_P)}$$

For devices of a particular type the parameter *k* is the most consistent (see fig. 2.6.2). Even though the pinch-off voltage varies by 300mV (potential offset of a pair) spreadsheet analysis shows that the slope is consistent to within ±2.5%. If the outlier is excluded the consistency of *k* is better than 0.5%.

The designer is primarily interested in the gain. This can be predicted accurately, given the operating current and drain resistance. For a single JFET with drain resistance R_D and drain output resistance R_{DO}, the effective load resistance is the two in parallel: -

$$R_{DE}=\frac{R_D R_{DO}}{R_D+R_{DO}}$$

The gain of a simple (common source) amplifier depends on the forward transconductance and effective load resistance: -

$$G=\frac{dV_D}{dV_{GS}}=-g_{FS}R_{DE} \quad \text{with} \quad g_{FS}=\frac{\partial I_{DS}}{\partial V_{GS}}$$

From above: $I_{DS} \approx k(V_{SG}+V_P)^2 \quad \Rightarrow \quad \frac{\partial I_{DS}}{\partial V_{GS}}=2k(V_{SG}+V_P)$

Also: $V_{SG}+V_P=\sqrt{\frac{I_{DS}}{k}} \quad \Rightarrow \quad g_{FS}=2\sqrt{kI_{DS}}$

In a typical design the effective load resistance is not much lower than the actual drain resistance which drops approximately 5V of DC (operating with a +15V power supply): -

$$R_{DE} \approx R_D \approx \frac{5}{I_{DS}}$$

As a ball park estimate, therefore, the gain is: -

$$G \approx g_{FS}R_D \approx -2\sqrt{kI_{DS}} \times \frac{5}{I_{DS}} = -10\sqrt{\frac{k}{I_{DS}}}$$

This is an interesting result: Higher voltage gain is obtained, in practice, at lower operating current – the opposite of what one might expect. The higher drain resistance more than compensates for the lower transconductance.

For the J112 matched pair (see fig. 2.6.2) the average value for $k=7.85 mA/V^2$ and so the expected gain at 1mA is approximately 28 (5kΩ drain resistance) and the forward transconductance is $g_{FS} \approx 5.6 mA/V$.

2.7 Deviation from the model

In practice the measured capacitance (V_{DS} = 0) appears to consist of three components: channel-gate, end electrodes and the package interconnections. As the gate voltage becomes more negative the channel-gate capacitance falls to zero quite quickly until the channel is completely shut off. The capacitance of the end electrodes then continues to fall more slowly, also due to a growing depletion region. The package and external strays contribute a fixed capacitance.

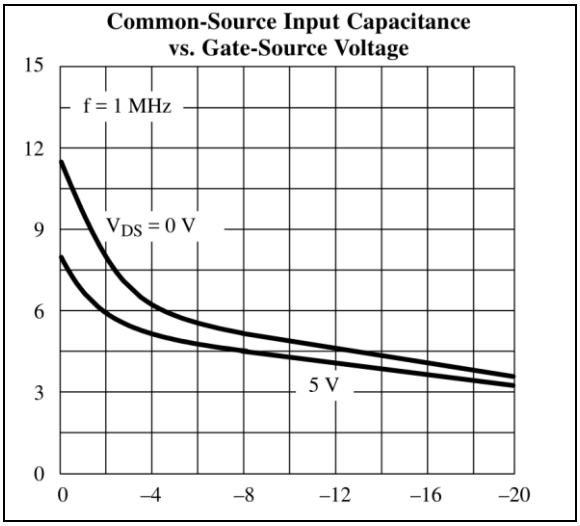


Fig. 2.7.1 Gate capacitance (pF) versus gate-channel voltage (Siliconix U430).

Clearly the hypothesis that the capacitance is uniformly distributed can only be an approximation, especially in pinched-off mode. In early work on JFETs Shockley [1] derived formulae, cited by many others, for the thickness of the conduction channel of the form: -

$$h(x) = h(0)\left(1 - \sqrt{\frac{V_{CG}(x)}{V_P}}\right)$$

This leads, by integration, directly to the result: -

$$I_{DS} = \frac{1}{R_0}\left\{V_{DS} + \frac{2}{3}\frac{\left(V_{DG}^{1.5} - V_{SG}^{1.5}\right)}{\sqrt{V_P}}\right\}$$

The model for operation in pinched-off mode needs modifying accordingly: -

$$I_{DS} = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^{\alpha}$$

Where α is between 1.5 and 2 depending on the type of device. In the case of the J112 type the best fit is obtained with a value of $\alpha \approx 2$.

1. Shockley, W. “A unipolar field-effect transistor”. Proc. IRE, vol 40, pp 1365 – 1376, 1952.

For the best matched pair the best fit values for α were 2.008 and 1.971 based on the following model: -

$$I_{DS} \approx k(V_{SG} + V_P)^{\alpha} \quad \Rightarrow \quad \log_{10}|I_{DS}| = \log_{10}|k| + \alpha \log_{10}|V_{SG} + V_P|$$

The pinch-off voltage was calculated by extrapolation using the simple square law. The results were V_P = -2.053V and -2.060V respectively).

Note that $V_{SG} + V_P$ is numerically negative in the operating region and the magnitude function is necessary.

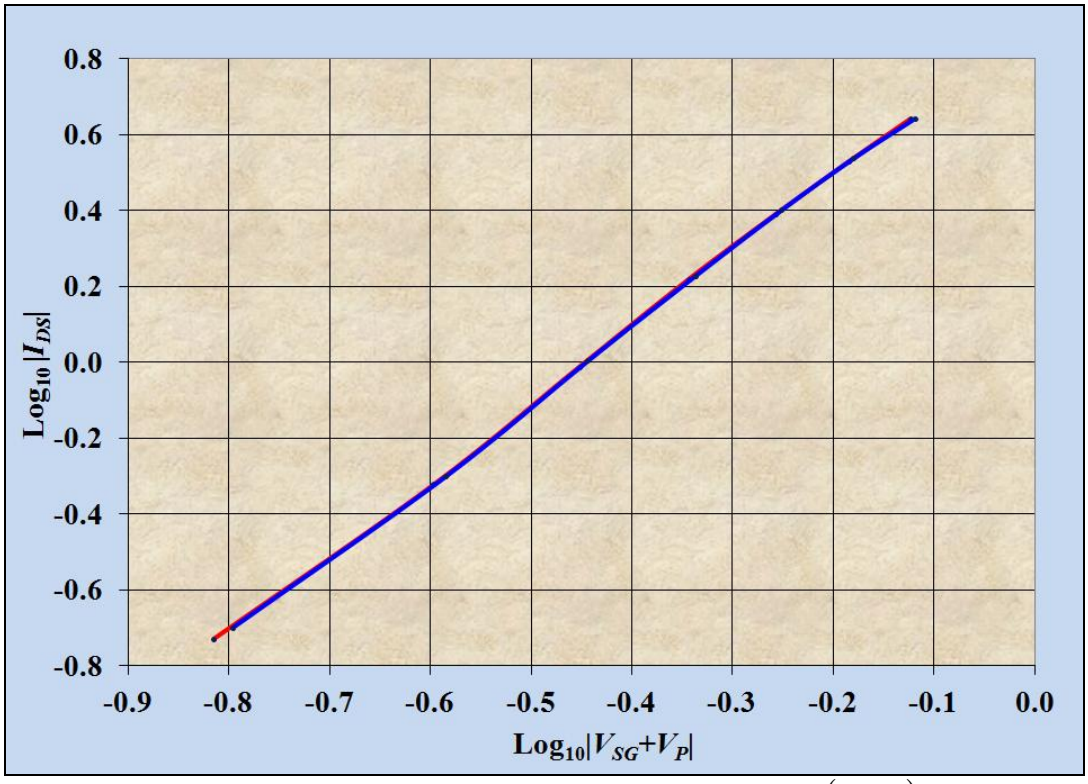


Fig 2.7.2 Logarithmic fit to calculate the average slope $(\alpha \approx 2)$

2.8 Design rules of thumb

a). Buy a batch of 10 low cost single devices (e.g. J112) and test for low noise. If at least 50% are good buy a batch of 100 – statistically this should provide a good number of low noise matched pairs.

b). Test for low noise with a prototype pre-amp circuit [1]. If you do not have access to a spectrum analyser use an audio amp with headphones and learn to recognise "white" noise versus "popcorn" noise. Reject devices with the latter.

c). Use the on resistance check to select matched pairs - with an ordinary DMM (±0.1% accuracy should suffice). Connect the gate to the source and measure the drain-source resistance.

d). Design for a fairly low operating current (compared to a similar BJT design) – usually in the range 0.1 – 1mA.

JFETs may have higher noise voltage, compared to the best BJTs, but….

1. Part 5, monograph 3: "Low noise JFET preamps"

3. JFET noise

3.1 The basic model

The basic noise model is an ideal (noise free) amplifier with a noise voltage source in series with the input and a noise current source in parallel with the source impedance (usually a source resistance). The source resistance also generates noise, V_S, depending on resistance and temperature (Johnson noise): -

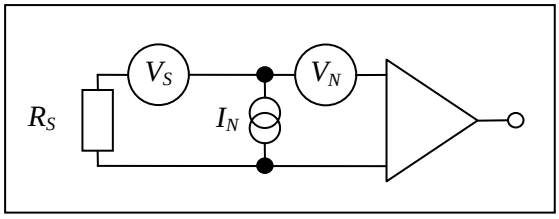


Fig. 3.1 Basic noise model for an amplifier

The noise current flows through the source impedance creating a third noise voltage source in series with V_N. The three sources are statistically independent and conform to a normal (Gaussian) probability distribution and so combine as follows: -

The total noise voltage referred to the input is: $$V_T = \sqrt{V_S^2 + (I_N R_S)^2 + V_N^2}$$

Where V_S, V_N, and I_N represent the standard deviation (root mean square -RMS) of the probability distributions. V_N and I_N are normally specified, in data sheets, in terms of spectral density (volts and amps per square root of bandwidth in Hertz) and as a (graph) function of frequency (see fig. 3.2). Johnson noise, on the other hand, can be easily calculated and the spectral density does not vary with frequency.

The random voltage, as measured across the resistor terminals (open circuit), and the random current that would flow through a short circuit can be characterised with normal distributions with RMS values: -

$$V_S = \sqrt{4kTR_S B} \quad \text{and} \quad I_S = \sqrt{\frac{4kTB}{R_S}} \quad \text{so that} \quad P_R = V_R I_R = 4kTB$$

Where: -

$k = 1.38 \times 10^{-23} J/K$ = Boltzmann's constant.

$T =$ is the absolute temperature ($T \approx 293K$ at 20ºC).

$B =$ Bandwidth of the measuring instrument, usually a spectrum analyser and often specified as 1Hz.

$P_R = 4kTB$ is the "noise power". It does not refer to any actual flow of power but it is a useful concept.

3.2 Noise performance

For an ideal JFET, at low frequency, the main source of (unavoidable) noise is the Johnson noise generated in the conduction channel, just as if it were a resistor. The voltage noise in the channel affects the width of the channel and, therefore, the drain-source current. For this reason JFET voltage noise is often modelled as a noise current source in the output circuit. This, however, translates in a complicated way, to an equivalent noise source in series with the gate, depending on the mode of operation. The most frequently used mode is the pinched-off mode and the voltage gain is sufficiently high for the gate series model (above) to be the most convenient. The only way to reduce this source of noise is to cool the device.

JFETs have other sources of noise due to the method of manufacture. The most problematic is "1/f" noise, so called as the spectral density increases inversely proportional to frequency. In some devices the lowest noise is only achieved at operating frequencies above 100Hz or even 1kHz. Similarly a mechanism known as "popcorn" noise can be a problem. The only practical solution is to splash the cash on tested/guaranteed devices or find a manufacturer who consistently produces excellent low cost (untested) JFETs. The author has found the latter policy quite realistic – typically 6 or 7 out of a pack of 10 JFETs perform well and the issue of DC matching is relatively easy to solve.

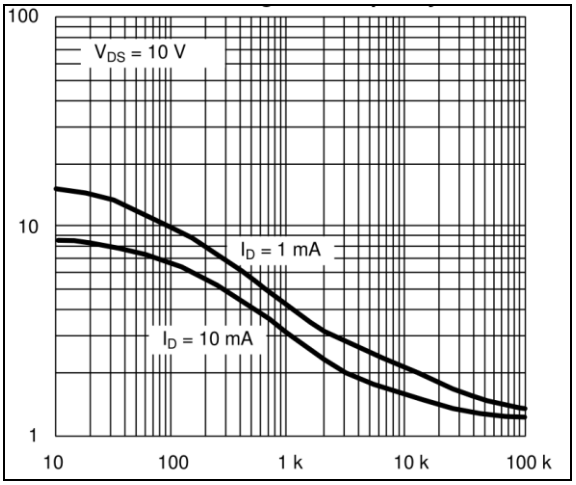


Fig. 3.2 Typical voltage noise (J112)

The other main source of noise is primarily due to the gate leakage current. It is remarkable that the quantisation of charge is directly responsible for this observable phenomenon. With a leakage current, I_L, the average number of electrons flowing through the gate in time, T, is: -

$$N = \frac{I_L T}{e}$$

Where $e = 1.60 \times 10^{-19} C$ is the quantum of electrical charge (i.e. on a proton).

These are thermally generated electrons and the rate at which they pass through the gate is highly random. According to a fundamental theory of probability the number can vary (RMS or one standard deviation) by $\sqrt{N}$.

The RMS current is, therefore: $$I_N = \frac{e\sqrt{N}}{T} = \sqrt{\frac{eI_L}{T}}$$

This is an average over a time T and a well known theory (Fourier transforms) equates this process of averaging to an equivalent ideal filter with bandwidth $2B$ Hertz.

$$\frac{1}{T} \equiv 2B \quad \Rightarrow \quad I_N = \sqrt{2eI_L B}$$

In practice data sheets usually provide noise performance in 1Hz of bandwidth (B = 1).

The leakage current depends mainly on temperature (approximately doubling every 11ºC) and the drain-gate voltage, increasing rapidly when V_{DS} > 10V. With good ventilation and a heat sink the self heating can be held to less than 10ºC above ambient and a leakage current of less than 10pA is possible. The noise current is, therefore: -

$$I_N = \sqrt{2 \times 1.6 \times 10^{-19} \times 10^{-11}} \approx 2\, fA/\sqrt{Hz}$$

A typical JFET will have a noise voltage of about $V_N \approx 5nV/\sqrt{Hz}$ and the noise resistance is: -

$$R_N = \frac{V_N}{I_N} \approx 2.5M\Omega$$

Also of interest is the noise power: $$P_N = V_N I_N \approx 10^{-23} W/Hz$$

Plus the equivalent noise temperature: $$T = \frac{P_N}{4k} \approx 0.2K$$

It is interesting to compare these results with the best BJT performance. From the datasheet of a SSM2210 with a collector current of 1mA at 25Hz: -

$$V_N \approx 1nV/\sqrt{Hz} \quad \text{and} \quad I_N \approx 1\,pA/\sqrt{Hz} \quad \Rightarrow \quad R_N \approx 1k\Omega$$

The noise resistance is much lower – more suited to resistance bridge applications.

The noise power and equivalent temperature, however, are much higher: -

$$P_N \approx 10^{-21} W/Hz \quad \text{and} \quad T \approx 18K$$

Historically it has not been practicable (or necessary) to produce a suitable noise matching transformer for low source resistance (e.g. 1Ω) to a JFET pre-amp. BJT pre-amps have been the first choice for resistance bridge applications. It may be possible, however, and the reader is invited to contact the author if interested [1].

1. Part 3, monograph 5: “Noise matching transformers”.

Null detectors – the basics

1. Introduction

According to Wikipedia (and Stig Ekelof [1]) the Wheatstone "bridge" was invented by Samuel Hunter Christie in 1833 but improved upon and popularised by Sir Charles Wheatstone ten years later. A simple version was still used in the "A" level syllabus as recently as the 1970s: a length of uniform resistance wire alongside a metre rule with copper connections and brass screw terminals mounted on a nice plank of wood. The null detector was a very simple (and delicate) moving coil galvanometer. With it we were able to establish the relationship between the length, cross sectional area, conductivity and resistance of various types of wire with approaching 0.1% accuracy.

The term "bridge" presumably refers to the way the galvanometer null detector bridges across the ratio arms: -

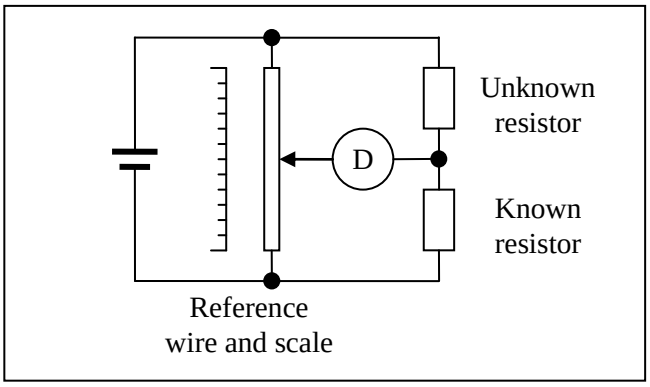


Fig. 1.1 A Wheatstone bridge

We were instructed to be careful when bridging the circuit and touch the reference wire gently with the knife-edge null detecting probe so as not to dent it.

The terms "bridge" and "null detector" have stuck – the principle is the same, though modern instruments have much more accurate dividers and more sensitive null detectors.

Whereas a simple moving coil galvanometer can be remarkably sensitive (μA DC) they are also rather fragile. With polarity reversal, amplification and filtering (first with thermionic and now semiconductor devices) the sensitivity improved substantially (sub μV DC). With low frequency AC energising of the bridge, at a frequency chosen to avoid local power supply interference and a noise resistance matching transformer sensitivity can be sub 1nV, with only a few seconds averaging time, even at room temperature.

In principle the linearity and gain of the null detector are unimportant – the main purpose is to indicate 0V (or zero current) in the presence of noise and interference. In practice, however, the out-of-balance signal can be useful to provide feedback for a control system (e.g. for a high performance servo-mechanism, based on a ratio-metric capacitive transducer and transformer bridge [2 and 3]).

1. Stig Ekelof: "The Genesis of the Wheatstone Bridge" "Engineering Science and Education Journal", volume 10, no 1, February 2001, pages 37–40.
2. Part 3, monograph 8: "A 16 bit binary differential capacitance bridge"
3. Part 1, monograph 3: "Rotary capacitive displacement transducers"

2. Low noise pre-amps

2.1 A basic pre-amp noise model

The noise performance and sensitivity of a null detector is determined mainly by the first stage – a low noise pre-amp. Principle characteristics (in data sheets) are noise voltage and noise current. These give rise to the concepts of noise power and noise resistance. The former should be as low as possible and the latter should be the same as the magnitude of the source impedance. The main aim is for the noise contribution from the pre-amp to be small compared to the noise from the source.

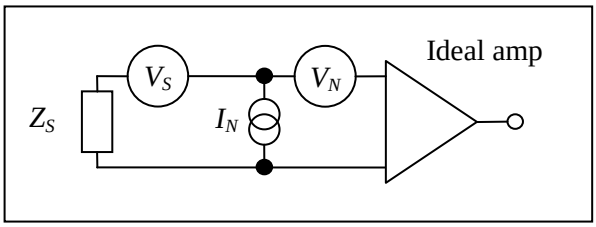


Fig. 2.1.1 A simple noise model

The symbols V_S , V_N , and I_N represent the statistical parameter (RMS) of the random voltages and current in 1Hz of bandwidth. For a refresher on noise theory see the appendix.

The noise power and noise resistance are, respectively: $P_N = V_N I_N$ and $R_N = \frac{V_N}{I_N}$

Ideally the noise resistance should be the same (within a factor of 3) as the magnitude of the source impedance: -

$$R_N = |Z_S|$$

Another useful parameter is the noise temperature – the temperature of the resistor that would produce the same noise voltage and noise current: -

$$T_N = \frac{P_N}{4k}$$

Where: $k = 1.38 \times 10^{-23} JK^{-1}$ is Boltzmann's constant.

As well as matching source and noise resistance the noise temperature of the pre-amp should be lower than the source resistance (e.g. a cryogenic resistance thermometer).

One of the easiest ways to optimise noise resistance is to employ a matched pair of bipolar junction transistors as the first stage of the pre-amp and vary the operating (collector) current. The noise resistance can vary from about 1kΩ (I_C = 1mA) to about 400kΩ (I_C = 1μA) [1]. Typical values from the data sheet for a SSM2210 (at 25Hz) are: -

Current (μA)	$V_N \left(nV/\sqrt{Hz}\right)$	$I_N \left(pA/\sqrt{Hz}\right)$	P_N (W)	R_N (kΩ)	T_N (K)
1	25	0.06	1.5×10^{-21}	417	27
10	8	0.3	2.4×10^{-21}	27	43
1000	1	1	1×10^{-21}	1.0	18

Fig. 2.1.2 Typical noise characteristics of a SSM2210 (Analog Devices Corp)

A higher operating current is not recommended as self heating can be a problem.

1. Part 5, monograph 3: "Low noise BJT pre-amps

2.2 Amplifiers in parallel

For a lower source resistance a number of identical amplifiers can be connected in parallel [1]. The output voltages are added together and divided by the number of amplifiers, *N*. The noise currents combine in parallel also according to the RMS calculation. The resulting noise voltage is lower but the noise current is higher so that the resulting noise resistance is reduced by a factor *N*: -

$$V_T = \frac{1}{N}\sqrt{V_1^2 + V_2^2 + V_3^2 + ...} = \frac{1}{\sqrt{N}}V_1 \quad \text{and} \quad I_T = \sqrt{\left(I_1^2 + I_2^2 + I_3^2 + ...\right)} = \sqrt{N}I_1 \quad \Rightarrow \quad R_N = \frac{V_T}{I_T} = \frac{1}{N}\frac{V_1}{I_1}$$

This can be achieved with a number of low noise matched pair BJTs (or dual JFETs) in parallel, each with its own constant current source. These replace a single matched "long tail pair" in a composite op-amp circuit. For practical circuits see the relevant monographs [1 and 2].

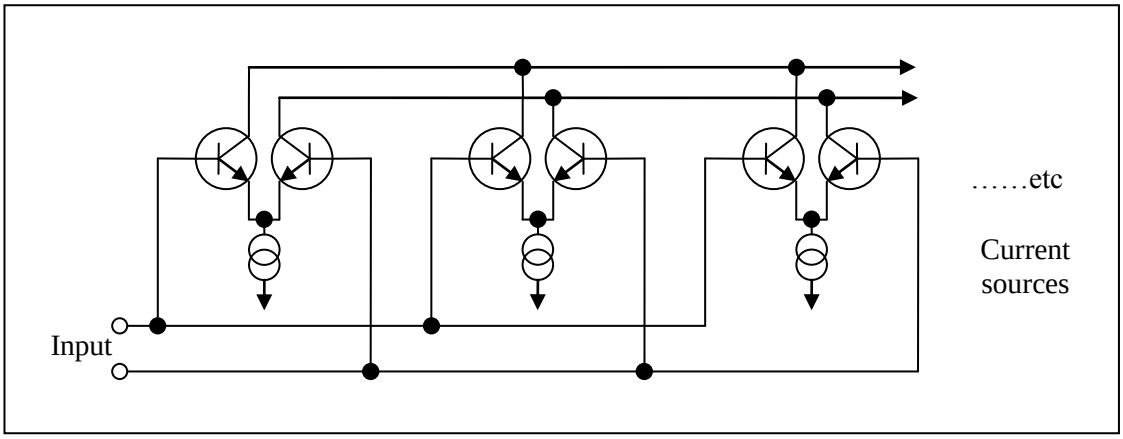


Fig. 2.2.1 Matched BJTs in parallel

2.3 Noise matching transformers

Connecting a large number of low noise pre-amps in parallel is not always practical. Some high temperature thermometers, for example, have a resistance as low as 1Ω. This would require 1000 dual BJTs in parallel. A better solution is to employ a noise matching transformer [3].

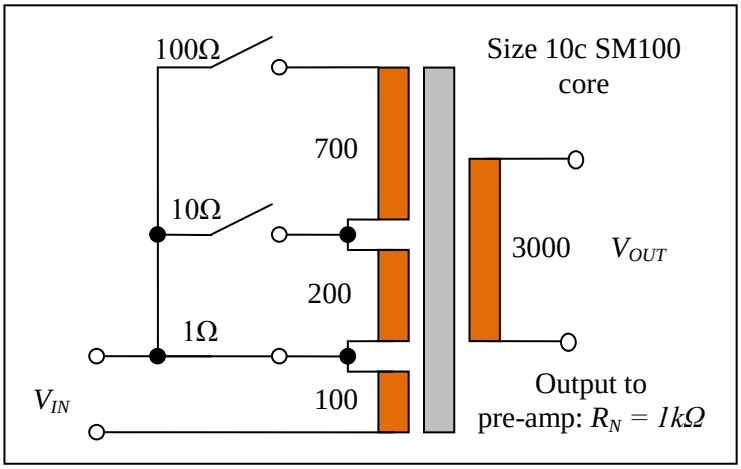


Fig. 2.3.1 A single-stage noise matching transformer

The main disadvantage is the large number of turns on the secondary – not difficult but very tedious.

1. Part 5, monograph 2: "Low noise BJT pre-amps"
2. Part 5, monograph 3: "Low noise JFET pre-amps"
3. Part 3, monograph 5: "Noise matching transformers"

2.4 Charge amplifiers

With high source impedance (e.g. a capacitance bridge: typically 1MΩ) it is better to employ a junction field effect (JFET) input stage. Remarkable results can be achieved with a low cost JFET input op-amp (e.g. LF356) though best performance is achieved with a dual matched JFET pair as part of a composite op-amp. A typical application is the charge amplifier in a capacitance bridge. It is shown elsewhere [1] that the contribution due to the noise current of the op-amp and feedback resistor is usually relatively small and a suitable model is: -

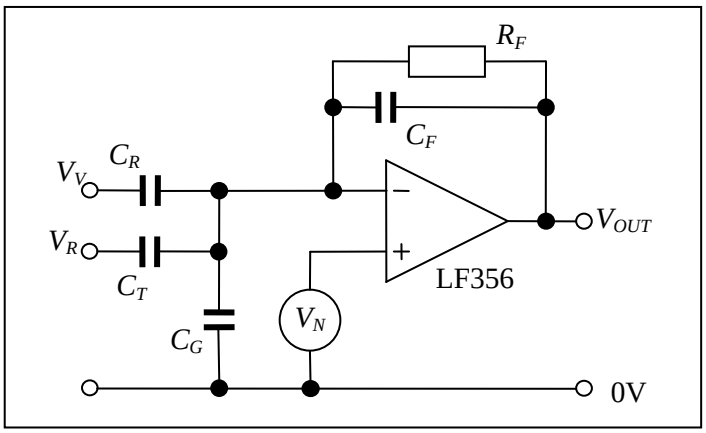


Fig. 2.4.1 A charge amplifier (outline schematic)

The symbol C_G represents ground capacitance due to (coax) connecting cables, capacitance in the transducer and input capacitance of the op-amp and should be kept to a minimum.

The action of feedback is to maintain the inverting input at local 0V ("virtual earth") and, therefore, no current flows through capacitance C_G. The circuit is an inverting amplifier so that, over a wide range of frequency: -

$$V_{OUT} = -\frac{V_V C_R + V_R C_T}{C_F}$$

In a typical bridge application the ratio of voltage is adjusted for a null balance so that the ratio of capacitance is: -

$$\frac{C_T}{C_R} = -\frac{V_V}{V_R}$$

C_T represents a transducer capacitance and C_R a reference capacitor, often incorporated into the transducer

The noise gain is that of a non-inverting amplifier (consider both inputs connected to 0V): -

$$G_N = 1 + \frac{C_R + C_T + C_G}{C_F}$$

For more detail, including noise analyses see the relevant monographs [2, 3, 4 and 5].

1. Part 5, monograph 3: "Low noise JFET pre-amps" . See section 2.
2. Part 1, monograph 3: "Rotary capacitive displacement transducers.
3. Part 1, monograph 4: "Linear capacitive displacement transducers"
4. Part 1, monograph 5: "Variable gap capacitive displacement transducers"
5. Part 5, monograph 4: "JFET theory"

3. AC gain and filtering

In low frequency applications a significant problem is low frequency (AC power supply related) interference. The low noise pre-amp, including matching transformer, usually has a fairly wide bandwidth - interference is amplified as well as the signal. A variable gain/filter stage allows the signal thorough but attenuates noise and interference sufficiently to allow a total AC amplification of up to 10^6-10^7 before the AC-DC conversion stage. A typical design for an operating frequency of 25 or 75Hz (local supply 50Hz) is as follows: -

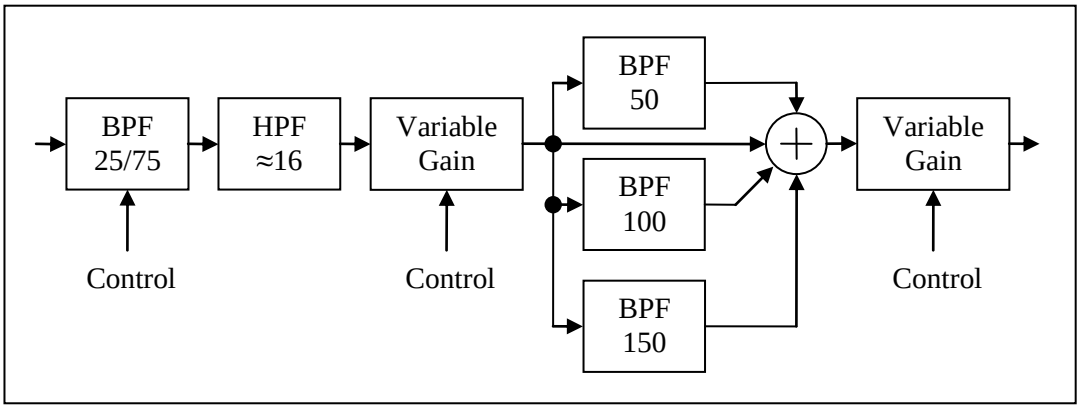


Fig. 3.1 Variable gain and filtering (outline schematic for 25Hz and 75Hz operation)

The first band-pass filter, at the selected operating frequency, reduces high frequency noise and interference sufficiently to allow further amplification. After the variable gain stage is a multi-frequency notch filter consisting of band-pass filters (single op-amp - inverting) and a summing junction. The summing function can be combined with a second variable gain stage (inverting amplifier with virtual earth).

Supply first harmonic (50Hz) can be a problem due to stray magnetic fields (from mains powered equipment), earth loops or lack of adequate screening. The second harmonic (100Hz) comes from stray magnetic fields from poorly designed power supplies. The pulses of current that charge the (large) smoothing capacitors immediately following the full-wave rectifier occurs every half cycle and the poorly routed wiring radiates magnetic flux. The lowest frequency component from this mechanism is 100Hz though all the higher harmonics are present. The third harmonic also comes from power supplies – this time from the mains transformer. With insufficient turns and/or iron a power transformer runs close to magnetic saturation. The result can be a remarkable amount of 150Hz magnetic flux spewing out of the transformer.

Stray magnetic flux passes easily through steel structures, along bench framework and, by virtue of sod's law, through your most sensitive bit of wiring. Good wiring practice helps (twisted pairs or coax cables and avoiding earth loops) but low frequency magnetic interference can be annoyingly difficult to diagnose [1].

Users of hearing aids (switched to loop operation) are frequent victims of magnetic interference due to poor wiring practice [2].

If local power supply is 60Hz the frequencies change accordingly.

1. Many audio engineers are familiar with these problems: mains "hum" and other buzzing sounds. In audio applications 100, 150Hz and higher harmonics are the most problematic - more audible than 50Hz.
2. Some years ago I was asked to investigate a mysterious "buzzing" that was bothering a number of hearing aid users in a Methodist chapel (hearing aids switched to loop), especially under a balcony. A small (3cm diameter with 50 turns) search coil, portable amplifier and headphones soon found the problem. The live cables were routed down one side of the building, via a bank of switches, but a single neutral returned along the other side. The other clues were that the problem got worse when they replaced incandescent bulbs (resistive) with more efficient fluorescent types (100Hz audible pulses) and the balcony was held up by an RSJ. They were very close to buying a new loop amplifier.

The overall effect reduces low frequency interference by at least an order of magnitude, even as the supply frequency varies (typically ±0.2% but up to ±1% with a sudden surge or drop in demand). For further reduction the whole amplifier/filter unit can be repeated two or three times.

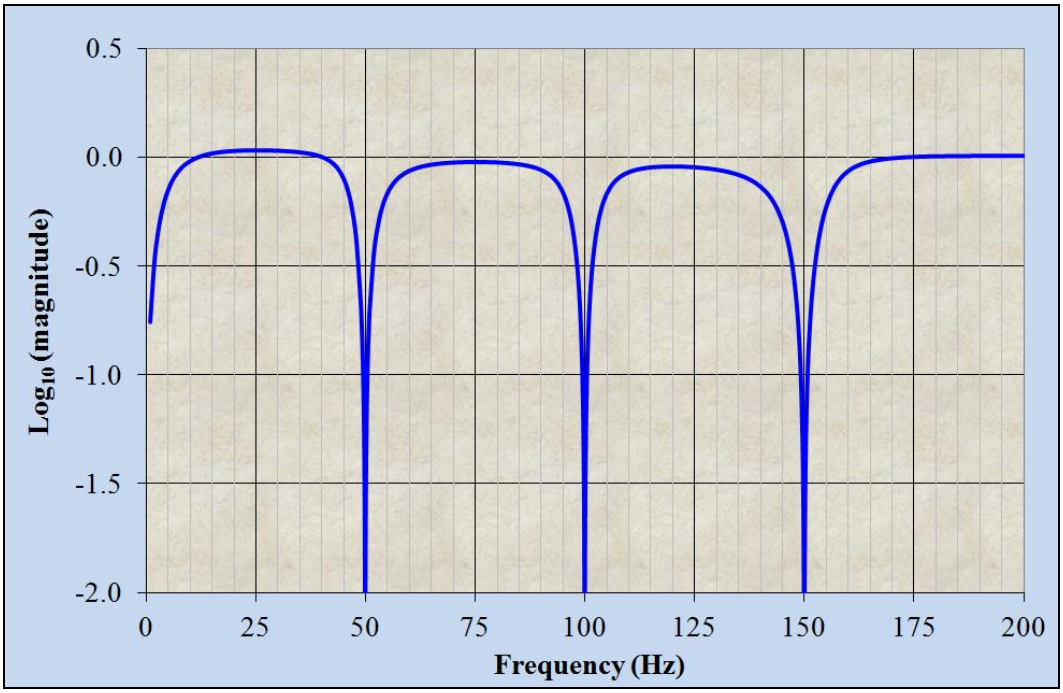


Fig. 3.2 The frequency response of the HPF and notch filter combination

The quality factors of the notch filters are chosen to ensure zero phase shift at 75Hz but this leaves a phase shift at 25Hz. The high-pass filter cut-off frequency is chosen to correct for this. The high-pass filter is most conveniently placed before the first (non-inverting, high input impedance) variable gain stage. A two-stage design is recommended as it has a very small phase shift at 75Hz [1], simplifying the design process.

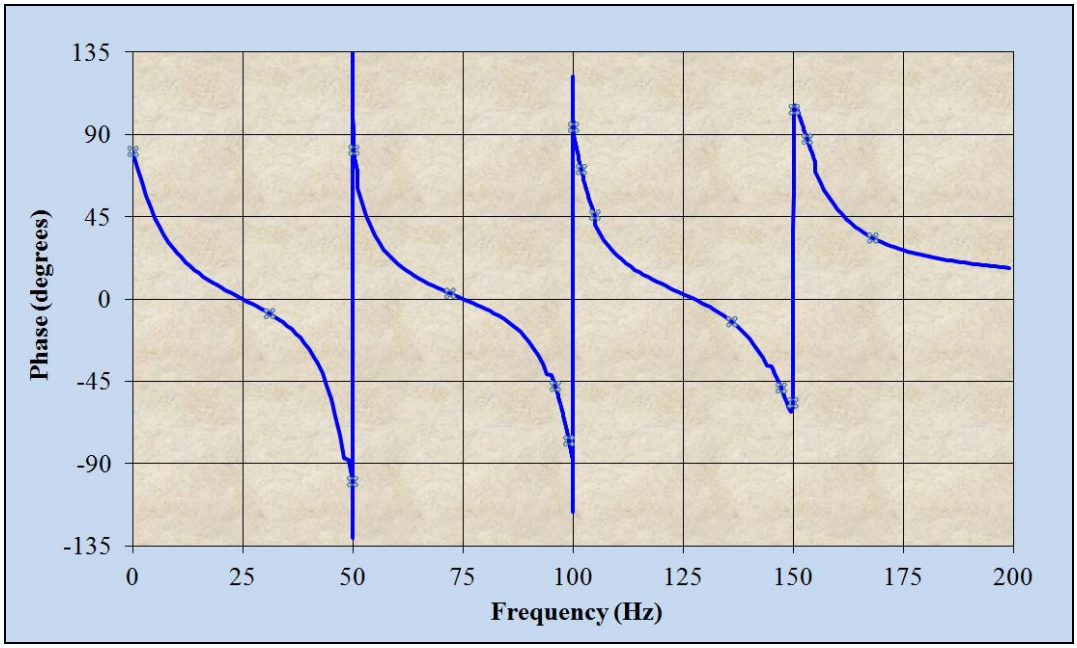


Fig. 3.3 Phase shift of the HPF and notch filter combination

In practice it is necessary for each BPF to have a variable resistor for adjusting frequency. For maximum rejection it is also advisable for the summing junction resistors to be adjustable (see appendix 2 for details).

1. Part 2, monograph 1: "Two-stage filters"

Typical values for centre frequency and quality factor are easily found with a spreadsheet simulator with sliders: -

For zero phase shift at 75Hz: $Q_{50Hz} = 5 \quad Q_{100Hz} = 10 \quad Q_{150Hz} = 10$

The phase shift at 25Hz, mainly due to the 50Hz notch, is compensated by the high-pass filter. A further slider control allows one to adjust the cut-off frequency so that the phase shift is zero at 25Hz with minimal effect at 75Hz.

The interaction between the notch BPFs also means that the centre frequency of each is not set precisely to the target value.

Spreadsheet settings: $f_{50Hz} = 49.48Hz \quad f_{100Hz} = 100.05Hz \quad f_{150Hz} = 151.45Hz$

In practice one adjusts the variable resistors with suitable test signals. There is a small amount of interaction between the settings so that the adjustments need to be repeated. Twice or three times should be sufficient.

The band-pass filters at the operating frequency are best with a lower quality factor: $Q \approx 2$ is a good compromise between selectivity and phase error due to frequency and temperature variations (mainly affecting the capacitor values – use low drift polyester types). Whereas the total phase error of the amplifier/filter sections is not critical a lower value (± a few degrees is tolerable) means less interaction between the in-phase and quadrature null-balance servos. Both components eventually reduce to zero.

Calculating practical component values can be made easier with a spreadsheet. Given the frequency and quality factor try a variety of values for R_2 in order to find a practical value for the capacitors. The resistance R_{1B} needs to be a fixed resistor plus a variable.

Freq (Hz)	Q	R_2 (kΩ)	R_{1A} (kΩ)	R_1 (Ω)	R_{1B} (Ω)	C_1=C_2 (nF)
25	2.083	39	20	2247	2540	680
50	5	200	100	2000	2041	159
75	2.022	39	20	2385	2717	220
100	10	200	100	500	503	159
150	10	200	100	500	503	106

Fig. 3.4 Practical component values

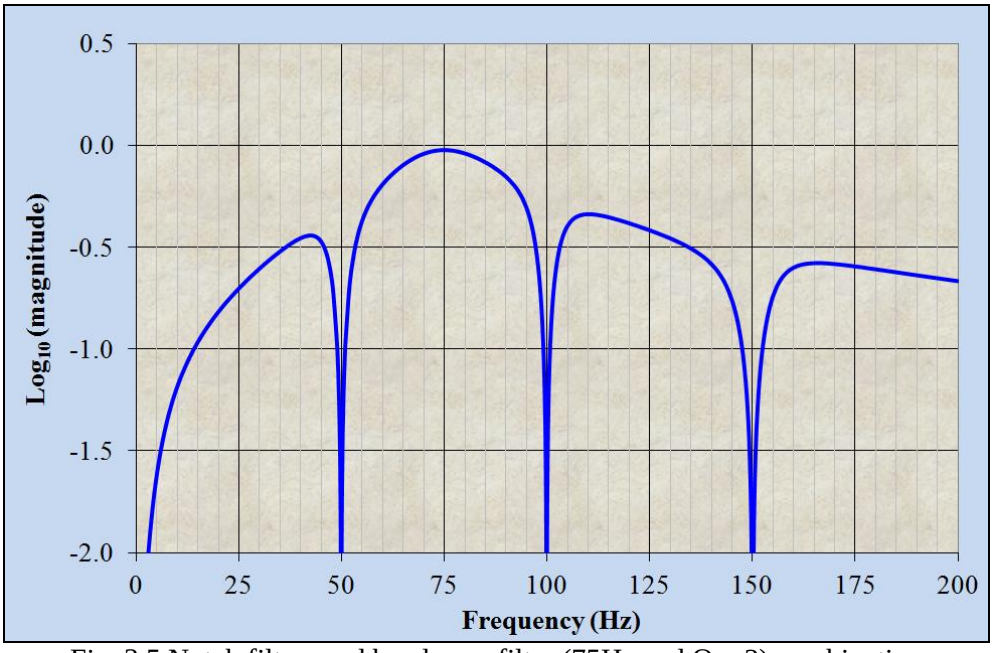


Fig. 3.5 Notch filters and band-pass filter (75Hz and Q = 2) combination

4. Synchronous rectifiers

4.1 The quad FET switch

After the amplifier/filter stages the AC signal is converted to DC. This can be achieved in a number of ways – the simplest being a quad FET switch driven with logic signals derived from the AC reference voltage. The circuit is a double-balanced bridge structure with a differential input and output – effectively multiplying the signal by ±1: -

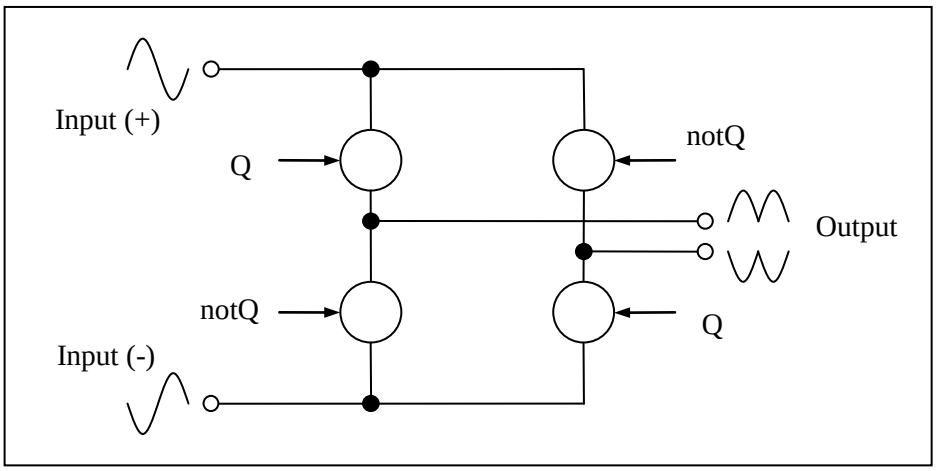


Fig. 4.1.1 The Quad FET switch synchronous rectifier (voltage mode)

If the input signal (and its inverse) is in-phase with the logic signals the result is a full-wave rectified signal (and its inverse) at the output. If the input signal is in quadrature the average output is zero. This gives rise to the alternative name: "phase sensitive detector" (PSD).

The double-balanced structure has the advantage that the charge injected from the logic signals through the FET switch gate capacitances is the same on both sides. The effect is eliminated by the following differential amplifier/low-pass filter stage. Switch (on) resistances are low and reasonably well matched so that the effect of input resistance of the following stage can be made negligible. Some designers prefer the current mode – as part of an inverting amplifier configuration. The main advantage is that the FET switches operate at 0V (both sides) and the variation of resistance (with common mode voltage) is eliminated. It is also possible to use low cost CD4066 type quad FET switches with a separate ±7.5V power supply, produced locally with a pair of zener diodes: -

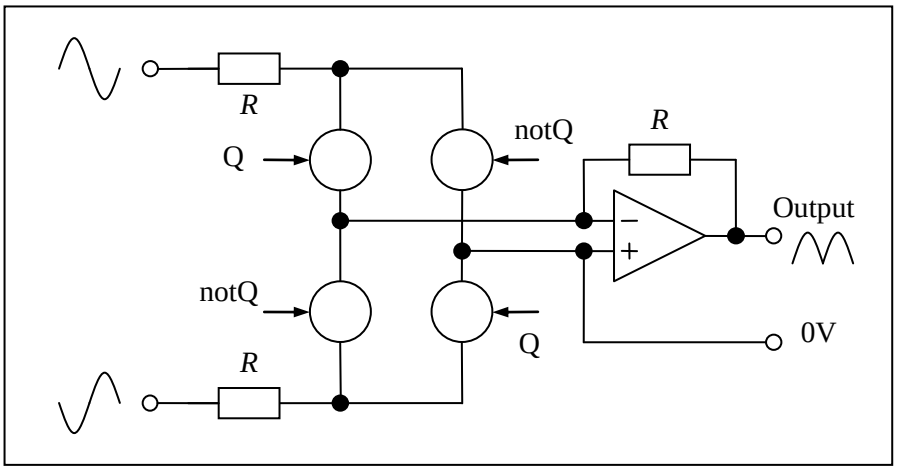


Fig. 4.1.2 The Quad FET switch synchronous rectifier (current mode)

This type of synchronous rectifier works well up to an operating frequency of 100kHz. Higher frequency (e.g. for a very high speed capacitance transducer interface) is possible with an RF type MOSFET "double balanced mixer" or "Gilbert cell multiplier" approach (sinewave drive).

4.2 Rejection of Noise and Interference

A square wave, with amplitude ±1 and frequency, ω_C, can be expressed as an infinite series of sine waves (a Fourier series). Only the odd harmonics are present (hence the sum includes only cases of k = 1, 3, 5, 7 etc). The amplitude of each harmonic reduces - inversely proportional to frequency.

$$V_{SQUAREWAVE} = \sum_{oddk}^{\infty} \frac{4}{\pi k} \sin(k\omega_C t)$$

If one assumes that the incoming signal has amplitude V_S and is at a frequency, ω with a phase (at t = 0) of θ the output of the quad FET switch is: -

$$V_{OUT} = 2V_S \sin(\omega t + \theta) \sum_{oddk}^{\infty} \frac{4}{\pi k} \sin(k\omega_C t)$$

The extra factor of two is due to the differential structure of the synchronous rectifier (voltage mode, see fig. 4.1.1). One can multiply each item on the right of Σ by sin(ωt + θ) then do the summation: -

$$V_{OUT} = 2V_S \sum_{oddk}^{\infty} \frac{4}{\pi k} \sin(\omega t + \theta) \sin(k\omega_C t)$$

The result is sum and difference frequencies: -

$$V_{OUT} = \frac{4}{\pi} V_S \sum_{oddk}^{\infty} \frac{1}{k} \{\cos([k\omega_C - \omega]t - \theta) - \cos([k\omega_C + \omega]t + \theta)\}$$

A low pass filter, after the synchronous rectifier, is used to reduce the AC components and pass the DC component, corresponding to zero frequency: -

$$\omega = k\omega_C \quad \Rightarrow \quad \omega \text{ is an odd multiple of } \omega_C$$

The synchronous rectifier only produces a DC output for odd harmonics of the carrier.

The effect of odd harmonics in the original carrier waveform is usually negligible, especially at "null balance".

For a pure carrier signal (k = 1 is the only component) the average (DC) output of the synchronous rectifier is: -

$$\omega = \omega_C \quad \Rightarrow \quad V_{OUT} = \frac{4}{\pi} V_S \cos(\theta) \quad \text{(voltage mode)}$$

If the signal is in-phase with the reference square wave $(\theta = 0)$ the output is simply proportional to the input. If the signal is in quadrature $(\theta = 90\,\text{deg})$ the average (DC) output is zero. Two synchronous rectifiers, operated at 90 degrees relative to each other, provide DC outputs proportional to the cosine and sine components (real and imaginary) respectively. With a current mode synchronous rectifier (or similar) the average (DC) output is reduced by a factor of two (factor $2/\pi$).

4.3 Output ripple

The AC components at the output are often called "ripple". The mid-higher frequency components are easily reduced by a low-pass filter after the rectifier (see next section) but the lower frequency components can be a problem. The most significant low frequency AC components are the second harmonic of the operating frequency (full wave rectified signal) and the difference between the operating frequency and the interference components.

Ripple components: -

- 25Hz operation rectified (2nd harmonic): $\omega_C + \omega \Rightarrow$ 50Hz
- 25Hz operation with 50Hz interference: $\omega_C - \omega \Rightarrow$ 50Hz and $3\omega_C - \omega \Rightarrow$ 25Hz
- 75Hz operation rectified (2nd harmonic): $\omega_C + \omega \Rightarrow$ 150Hz (no problem for the low-pass filter)
- 75Hz operation with 50Hz interference: $\omega_C - \omega \Rightarrow$ 25Hz
- 75Hz operation with 100Hz interference: $\omega_C - \omega \Rightarrow$ 25Hz
- 75Hz operation with 150Hz interference: $\omega_C - \omega \Rightarrow$ 75Hz (no problem for the low-pass filter)

Higher frequency components of interference are attenuated sufficiently by the band-pass filters before the synchronous rectifier. Clearly the main problem is the 25Hz ripple component.

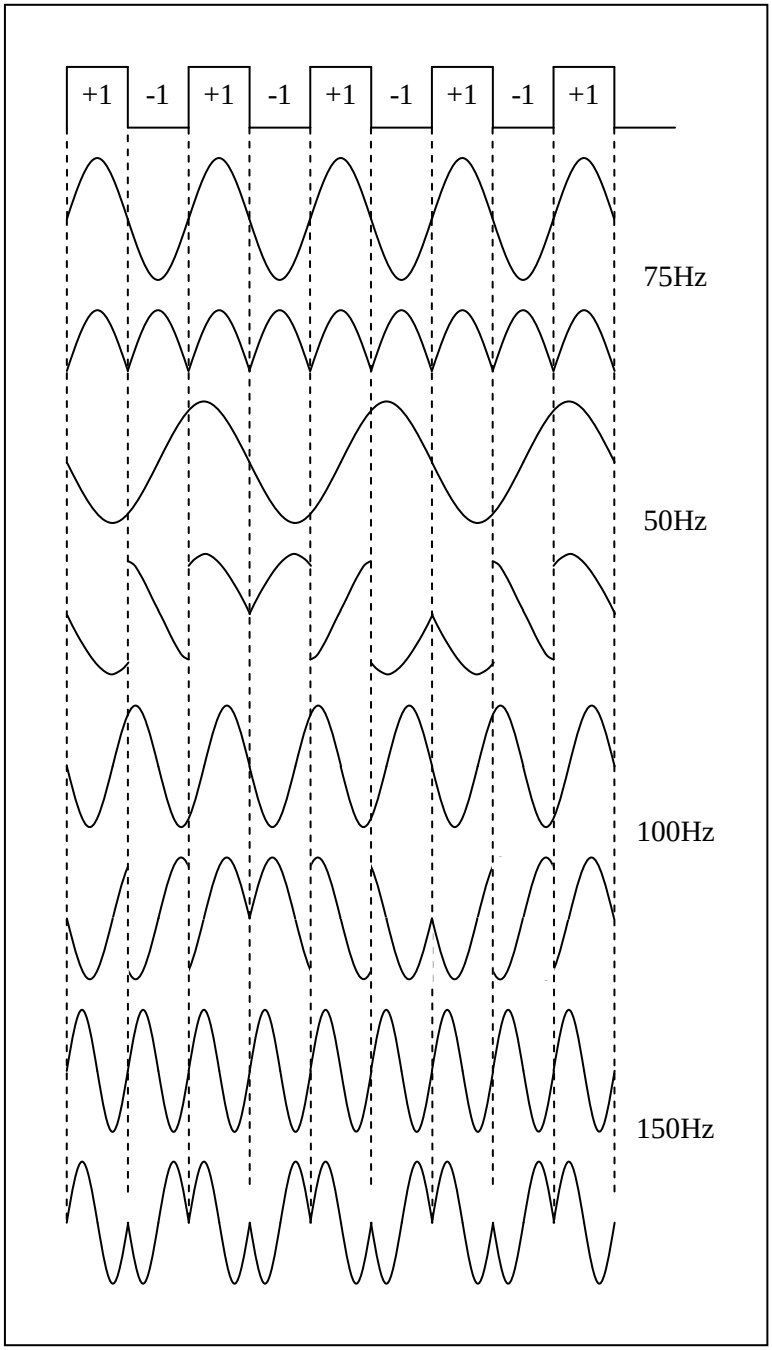


Fig. 4.2.1 Output ripple for 50Hz, 100Hz and 150Hz supply interference
(75Hz operating frequency)

5. Low-pass filters

5.1 A combined low-pass/notch filter

When operating at low frequency (25Hz and 75Hz) the main problem is the 25Hz AC component. The solution is quite simple – a combined low-pass and notch filter: A simple band-pass filter (inverting) provides a second input to a conventional second order low-pass filter: -

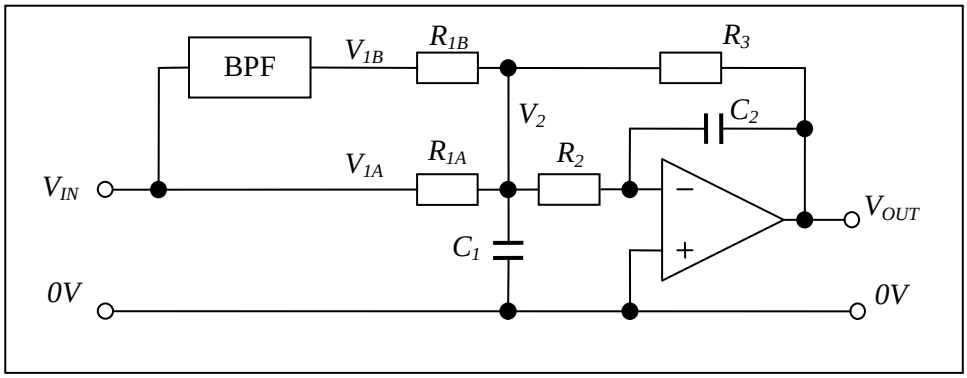


Fig. 5.1.1 A combined low-pass and notch filter.

It is shown in appendix 3 that the output of the filter is, in normalised form $\left(s = j\omega/\omega_N\right)$: -

$$V_{OUT} = \left(V_{1A}\frac{R_3}{R_{1A}} + V_{1B}\frac{R_3}{R_{1B}}\right)\left(\frac{-1}{1+2\xi s+s^2}\right)$$

The natural frequency and damping ratio are: -

$$f_N = \frac{1}{2\pi\sqrt{R_3R_2C_2C_1}} \quad \text{and} \quad \xi = \frac{1}{2}\sqrt{\frac{R_2C_2}{R_3C_1}}\left(\frac{R_3}{R_{1A}} + \frac{R_3}{R_{1B}} + \frac{R_3}{R_2}\right)$$

A typical bandwidth is 10Hz (critically damped) which provides sufficient smoothing for fast balance operation. Lower bandwidths can then be implemented by sampling the output with a 12-bit ADC plus microcontroller and employing digital filtering. A modest sampling rate of 50Hz is practicable and should be more than sufficient.

If an analogue (moving coil) meter balance indicator is required a suitable bandwidth is 0.2Hz (critically damped).

In both cases the band-pass filter (see below) has a maximum gain of 0.5 hence $R_{1B} = 0.5R_{1A}$
Practical values for the components are: -

R_{1A} (kΩ)	R_{1B} (kΩ)	R_2 (kΩ)	R_3 (kΩ)	C_1 (nF)	C_2 (nF)	Freq (Hz)	DR
150	75	39	160	398	100	10	0.90
540	270	180	24200	4400	33	0.20	1.00

Fig. 5.1.3 Practical values for low pass filters.

A very basic band-pass filter is a practical option: -

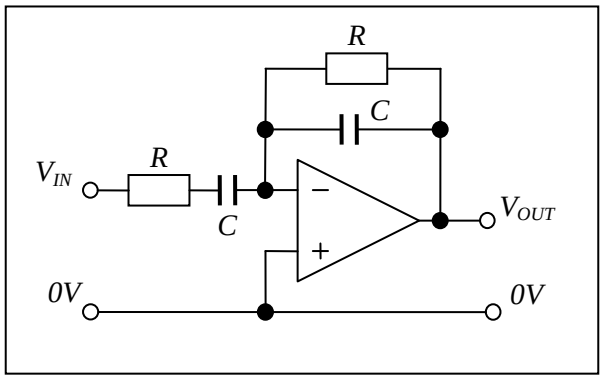


Fig. 5.1.2 A basic band-pass filter

It is shown in appendix 3 that the transfer function is, in the complex representation $(s = j\omega)$: -

$$\frac{V_{OUT}}{V_{IN}} = \frac{-sRC}{1+2sRC+s^2R^2C^2}$$

The natural frequency and damping ratio are: -

$$f_N = \frac{1}{2\pi RC} \quad \text{and} \quad \xi = 1$$

The maximum gain, at the natural frequency, is 0.5

Practical component values are $C = 470nF$ and $R = 13.5k\Omega$ (An 8k2 fixed resistor plus a 10k variable).

The combination provides a significant attenuation to the 25Hz difference component – at least -40dB even when the power supply varies by ±1%.

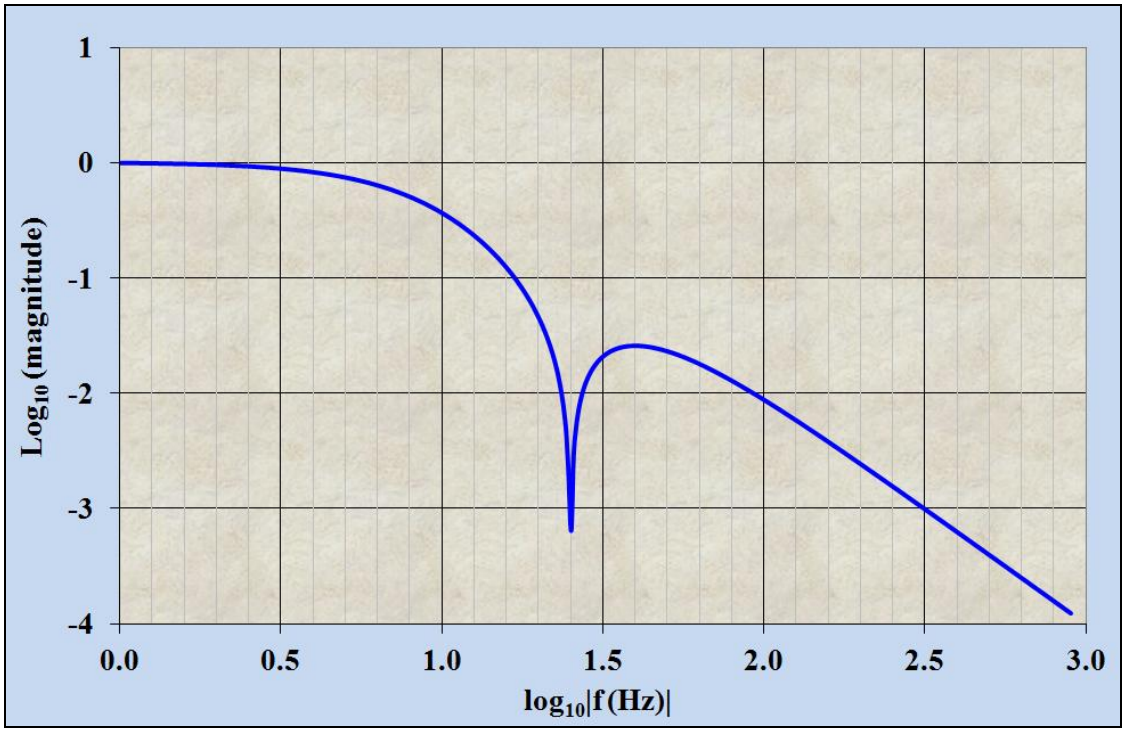


Fig. 5.1.4 Frequency response of a 10Hz low-pass plus 25Hz notch filter

5.2 A bridge as part of a control system

It is often the case that a null-balance bridge or signal conditioner is used to interface a transducer as part of a control system. Typical applications are high performance servo-mechanisms based on ratiometric variable capacitance transducers. One example is a limited angle servo-mechanism used for the positioning of a mirror (e.g. rapid scanning of a laser beam or microscope image) [1 and 2].

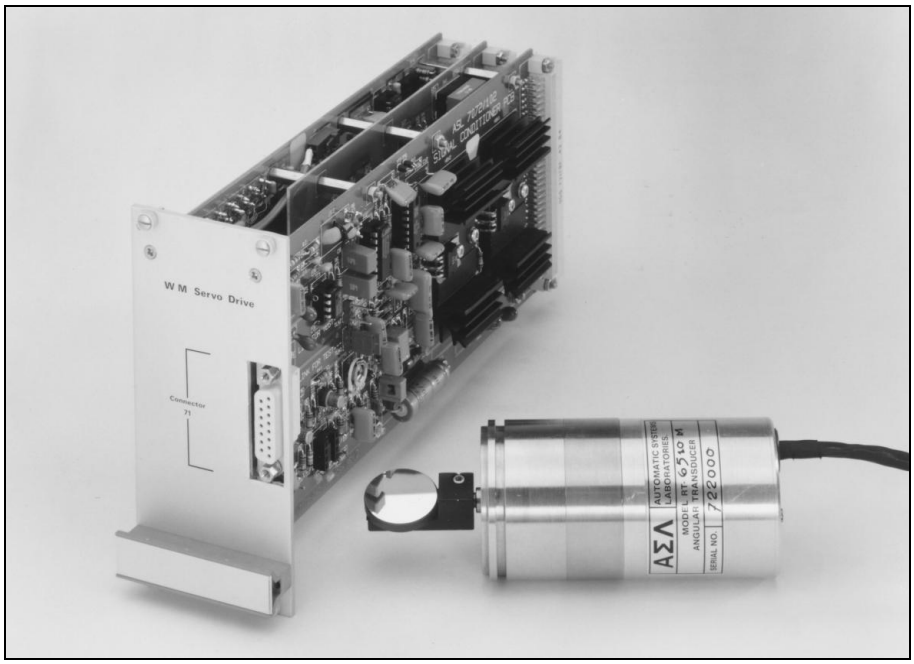


Fig. 5.2.1 A high accuracy servo system

The conventional approach is to employ a "signal conditioner": a transducer interface that produces a DC output proportional to position. The interface produces second harmonic ripple, even without interference. The low pass filter reduces the ripple considerably but does not eliminate it completely.

Ripple attenuation versus phase shift, due to the filter, then becomes a major trade-off. Ripple can be reduced at the expense of extra phase shift and vice versa. A small phase shift can be tolerated (e.g. 5 degrees maximum) as part of an overall phase shift budget for the main control loop stability. AC ripple, on the other hand, is amplified by the loop compensation, especially if it includes a derivative term (e.g. PID control with high frequency boost to compensate for inertia), resulting in overload.

The temptation is to opt for a high operating frequency but this can compromise accuracy [3].

The solution is to employ the bridge as part of the control loop: The set-point ("demand") signal sets the bridge ratio and the actuator drives the transducer (mechanically) to achieve null balance. Zero output voltage means zero ripple!

For advice or more detail please contact the author.

1. Part 1, monograph 3: "Rotary capacitive displacement transducers"
2. Part 3, monograph 8: "A 16 bit binary differential capacitance bridge
3. In one version (frame scanning for a thermal imager) the carrier frequency was 2MHz so that the low-pass filter bandwidth could be as high as 10kHz.

6. The quadrature servo

6.1 In-phase and quadrature

The issue of quadrature usually arises when measuring the ratio of resistors with either parallel cable capacitance or, in the case of very low resistance values, series inductance. For more detail see the monograph "High accuracy resistors" [1]. Quadrature is much less of a problem with capacitance bridges (except with very long connecting cables) [2].

In-phase and quadrature are defined relative to master phase references: usually the logic signals used to drive the synchronous rectifiers. To generate 75Hz, for example, a phase-lock loop is used to track the (nominally) 50Hz supply with a divider in feedback followed by three T-type flip-flops (e.g. based on JK flip-flops, type CD4027). The Q and notQ output timings can be accurate to a few nano-seconds thanks to the accurate matching of gate characteristics on the same chip: -

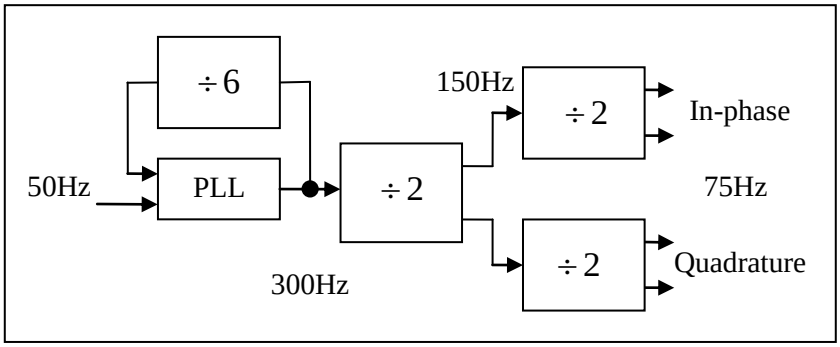


Fig. 6.1.1 The master phase reference (75Hz)

Phase locking the operating ("carrier") frequency to a precise multiple of the supply frequency affords maximum rejection of interference.

The in-phase logic signals can then be used to produce the (nominally) in-phase sinewave carrier with two second order low-pass filters, each operated at the corner frequency with a total phase shift of approximately 180deg. The higher harmonics are reduced sufficiently for the purpose and phase accuracy is not particularly important.

A typical application is a resistance bridge with a three-stage ratio transformer. A quadrature imbalance is caused mainly by cable capacitance plus a very small phase error due to the ratio transformer [3]: -

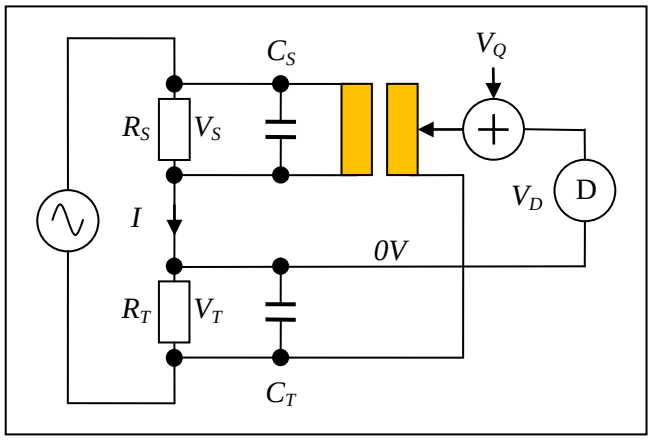


Fig. 6.1.2 A typical resistance bridge structure

1. Part 1, monograph 1: "High accuracy resistors". See section 3.
2. Part 3, monograph 2: "Single stage inductors and transformers". See section 4.5 and example calculation 4.5.1
3. Part 3, monograph 4: "Three-stage RTs". See example calculation section 2.3.

The transformer ratio is adjusted and the quadrature servo adds a signal, V_Q, to the output of the bridge so that a null is achieved for both in-phase and quadrature components: -

$$V_D = aV_S - V_T + V_Q = 0$$

$a = \frac{N_S}{N_P}$ is the transformer ratio setting. The phase error of a ratio transformer is usually much smaller than that due to the parallel capacitance and is simply added to it – equivalent to a small change in capacitance. The second order term (in-phase error) is negligible. I shall assume, therefore, that a is a real number.

A resistor in parallel with capacitance has impedance: $Z = \frac{R}{1 + j\omega RC}$

Define parameters: $\theta_S = \omega_C R_S C_S$ and $\theta_T = \omega_C R_T C_T$. Both are small (typically <0.1mrad or 100ppm)

Note that for small angles, to a good approximation (in radians): $\tan(\theta) \approx \theta$

The ideal quadrature servo needs to add a signal: $V_Q = V_T - aV_S = \frac{IR_T}{1 + j\theta_T} - \frac{aIR_S}{1 + j\theta_S}$

One can separate this into its real and imaginary parts: -

$$V_Q = \frac{IR_T}{1+\theta_T^2} - \frac{aIR_S}{1+\theta_S^2} + j\left(\frac{aIR_S}{1+\theta_S^2}\theta_S - \frac{IR_T}{1+\theta_T^2}\theta_T\right)$$

Both contributions to the imaginary component are small and the effect of the denominators is negligible.

$$\frac{IR_T}{1+\theta_T^2}\theta_T \approx IR_T\theta_T \text{ etc. } \Rightarrow \text{ Im}(V_Q) \approx j(aIR_S\theta_S - IR_T\theta_T)$$

Similarly, at or approaching null balance, the real part is also small so that: $IR_T \approx aIR_S$

One possibility is, therefore, for the quadrature servo to inject a signal: $V_Q = jaIR_S(\theta_S - \theta_T)$

This signal is, as expected, proportional to the magnitude of the current and the phase difference. It is also precisely in quadrature relative to the current but that is not easy to do in practice.

A more practical option is to derive a signal from the reference voltage, V_S: -

$$V_Q = jbV_S \quad \Rightarrow \quad V_D = (a + jb)V_S - V_T$$

At null balance (in-phase and quadrature): $V_D = 0 \Rightarrow \frac{V_T}{V_S} = a + jb$

Now: $V_T = \frac{IR_T}{1 + j\theta_T}$ and $V_S = \frac{IR_S}{1 + j\theta_S} \Rightarrow a + jb = \frac{R_T}{R_S} \times \frac{1 + j\theta_S}{1 + j\theta_T}$

With a little algebra: -

$$\Rightarrow \quad a = \frac{R_T}{R_S}\left(\frac{1+\theta_S\theta_T}{1+\theta_T^2}\right) \text{ and } \quad b = \frac{R_T}{R_S}\left(\frac{\theta_S - \theta_T}{1+\theta_T^2}\right)$$

For the very highest accuracy it is possible to add capacitance or, better still, matching cable to one side or the other so that the remaining quadrature imbalance is small. To a very good approximation, therefore: -

$$\theta_S \approx \theta_T \text{ and } \theta_S << 1 \text{ and } \theta_T << 1 \Rightarrow a \approx \frac{R_T}{R_S} \text{ and } b \approx \frac{R_T}{R_S}(\theta_S - \theta_T)$$

The phase difference is: $$\frac{b}{a} = \theta_S - \theta_T$$

Matching cable is the better choice because any loss factor ("tan(δ)") of the cable dielectric, equivalent to a small imaginary component for both θ_S and θ_T, is the same for both numerator and denominator and cancels (in the formula for the real component, *a*).

$$\theta_S = \theta_T \text{ and } \theta_S \to \theta_S(1 + j\delta) \text{ and } \theta_T \to \theta_T(1 + j\delta) \Rightarrow a = \frac{R_T}{R_S}$$

Similarly, if the signal injected has a small phase error, equivalent to a small imaginary component, θ_Q, of the parameter *b*, the result is a very small in-phase error: -

$$b \to b(1 + j\theta_Q) \Rightarrow a + jb \to a - b\theta_Q + jb \Rightarrow \frac{\Delta a}{a} = (\theta_T - \theta_S)\theta_Q$$

6.1.1 Example calculation: -

$R_S = 25\Omega$ with 1 metre of cable $(C_S = 200pF)$; $R_T = 100\Omega$ at the end of 5 metres of cable $(C_T = 1nF)$ with an operating frequency of 75Hz ($\omega = 2\pi f = 471$ radians per second).

$$\theta_S = 471 \times 25 \times 2 \times 10^{-10} \approx 2.4 \times 10^{-6} \quad \theta_T = 471 \times 100 \times 10^{-9} \approx 4.7 \times 10^{-5} \text{ and } a = \frac{R_T}{R_S} \approx 4$$

The in-phase error is: $$\frac{\delta a}{a} = \frac{1 + \theta_S\theta_T}{1 + \theta_T^2} \approx 2.2 \times 10^{-9}$$ (2.2ppb and negligible)

The quadrature imbalance is: $$\theta_S - \theta_T = -4.5 \times 10^{-5}$$ (45ppm)

The parameter *b* is: $$b = \frac{R_T}{R_S}(\theta_S - \theta_T) \approx -1.8 \times 10^{-4}$$ (180ppm)

With an operating current of 1mA (RMS) the reference voltage is 25mV (RMS) and the magnitude of the injected signal is: -

$$|V_Q| = |jbV_S| \approx 4.5\mu V \text{ (RMS)}$$

If the injected signal has a phase error of $\theta_Q = 10^{-4}$ (0.1mrad) the in-phase error is: -

$$\frac{\Delta a}{a} = (\theta_S - \theta_T)\theta_Q \approx 4.5 \times 10^{-9}$$ (4.5ppb)

The method based on $V_Q = jbV_S$ is sufficiently accurate.

6.2 A practical quadrature servo

One practical approach is based on an analogue multiplier and integrator in a negative feedback loop.

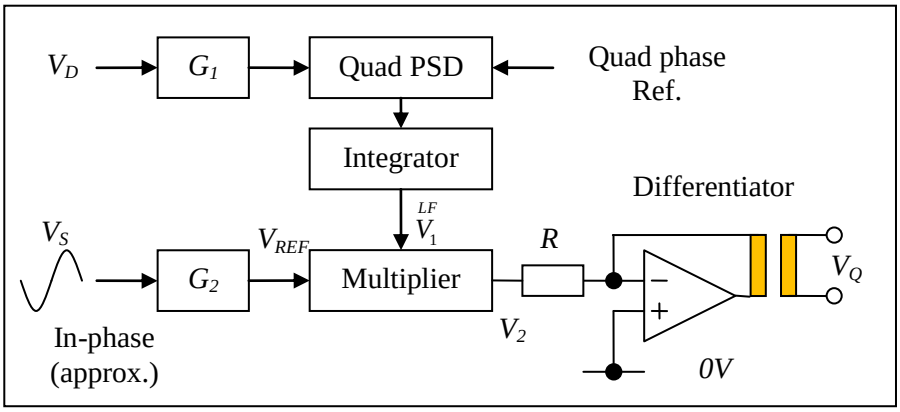


Fig. 6.2.1 A typical quadrature servo

The DC output of the integrator, together with the AC reference voltage (substantially in-phase), derived from the bridge reference resistor, are put to an analogue multiplier (four quadrant e.g. Analog Devices AD534). The reference voltage is amplified to a suitable level (maximum typically $20V_{PK\text{-}PK}$) with a high accuracy differential amplifier (gain G_2) [1: section 5]. A bridge with a pair of high accuracy voltage followers provides a convenient low impedance source for the reference voltage V_S [2].

The AC output (substantially in-phase) of the multiplier then undergoes a phase shift of precisely 90 degrees with a high accuracy differentiator based on a low phase error ferrite pot core transformer [1: section 6]. The transformer secondary is isolated so that the signal is easily added in series with the output of the ratio transformer (see fig. 6.1.2).

The synchronous rectifier/integrator functions are easily combined into a current mode circuit with a feedback capacitor, rather than a resistor: -

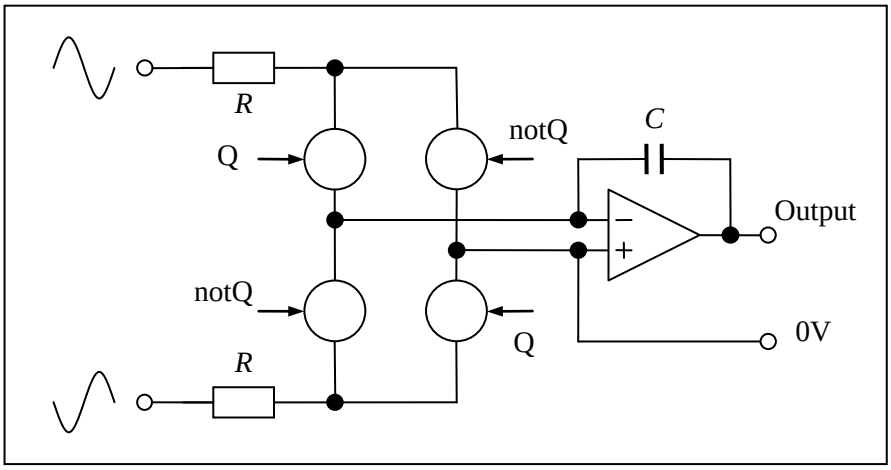


Fig. 6.2.2 A combined synchronous rectifier/integrator

Clearly it would be an advantage if the amplifier/filter stages (represented by G_1) used for the in-phase null balance could also be used by the quadrature servo. With a suitable integrator time constant, gain G_2 and transformer/mutual inductance this proves to be possible (see below).

1. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators".
2. Part 4, monograph 2: "High accuracy voltage followers". See section 1.

The overall effect of the feedback is to drive the quadrature component to zero. The output of the differentiator circuit is [1]: -

$$V_Q = j\frac{\omega_C L_M}{R}V_2$$

L_M is the mutual inductance of the transformer, ω_C is the operating frequency and V_2/R is the current through the transformer primary. V_Q is limited by the maximum output of the multiplier (typically: V_2 = ±10V$_{PK}$), the maximum current through the ferrite transformer primary (typically ±1mA$_{PK}$ with R = 10kΩ) and the mutual inductance. A suitable transformer has 16 turns for both primary and secondary (see section 6.4) on an RM10 pot core so that the mutual inductance is [1]: -

$$A_L = 400\,nH/turn^2 \quad \Rightarrow \quad L_M = N_S N_P A_L \approx 100\,\mu H$$

The maximum output at an operating frequency of 75Hz $\left(\omega_C = 471 rads^{-1}\right)$ is:-

$$V_2 = 10V_{PK} \text{ and } R = 10k\Omega \;\Rightarrow\; \left|V_Q\right|_{MAX} = 47\,\mu V_{PK}$$

The output of the multiplier is, typically: $V_2 = \dfrac{V_{REF} \times \overset{LF}{V_1}}{10V}$ with $V_{REF} = G_2 V_S$

The maximum is only achieved when the V_{REF} input to the multiplier is ±10V$_{PK}$ and the gain of the differential amplifier, G_2, is set accordingly. The symbol $\overset{LF}{V_1}$ represents the DC output of the integrator (maximum also ±10V).

The denominator factor 10V (ten Volts) is typical for a four quadrant analog multiplier. The symbol V (volts) is retained as a reminder that the equations must be dimensionally correct. The actual range of the servo depends on the gain of the differential amplifier and the differentiator as the following demonstrates: -

The bridge output (see fig. 6.1.2) is: $aV_S - V_T = \dfrac{aIR_S}{1+j\theta_S} - \dfrac{IR_T}{1+j\theta_T} \approx aIR_S(1-j\theta_S) - IR_T(1-j\theta_T)$

When the in-phase (real) component is set to null: $a = \dfrac{R_T}{R_S}$

The remaining imaginary component is, to a good approximation: $\mathrm{Im}(aV_S - V_T) \approx IR_T(\theta_T - \theta_S)$

With $\overset{LF}{V_1} = 10V$ the maximum output of the servo is: $\left|V_Q\right|_{MAX} = \dfrac{\omega_C L_M}{R}V_{REF} = \dfrac{\omega_C L_M}{R}G_2 V_S$

This is equal to but opposite the maximum quadrature output of the bridge: $\left|V_Q\right|_{MAX} = IR_T(\theta_S - \theta_T)_{MAX}$

$$V_S \approx IR_S \;\Rightarrow\; \left|V_Q\right|_{MAX} \approx \frac{\omega_C L_M}{R}G_2 IR_S = IR_T(\theta_S - \theta_T)_{MAX}$$

This corresponds to a maximum value for the parameter b: $b_{MAX} = \dfrac{R_T}{R_S}(\theta_S - \theta_T)_{MAX} = \dfrac{\omega_C L_M}{R}G_2$

1. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators". See section 5.

6.3 The servo as a control loop

The quadrature servo can be thought of as a continuous negative feedback control system. The input is the bridge output: -

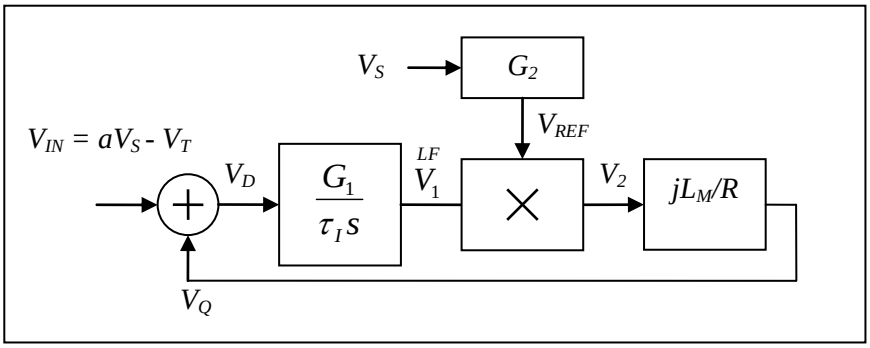


Fig. 6.3.1 The quadrature servo feedback loop

The ferrite transformer secondary is added in series with the bridge output so that the signal presented to the amplifier filter stages is: -

$$V_D = V_{IN} + V_Q$$

The quadrature servo injects a voltage proportional to the reference signal V_S (the voltage developed across the reference resistor) and phase shifted by precisely 90 degrees: -

$$V_{IN} = aV_S - V_T \quad \text{and} \quad V_Q = jbV_S \;\Rightarrow\; V_D = (a + jb)V_S - V_T$$

The null balance state (parameters *a* and *b*), for both in-phase and quadrature, is independent of the reference voltage and, therefore, the operating current. The balance remains stable as the operating current varies (randomly or due to temperature variations).

The quadrature servo main gain block consists of an amplifier, filter, synchronous rectifier and integrator, producing a slowly varying "DC" output proportional to the integral of the imaginary component. One can describe this with the usual complex representation of an integrator $(s = j\omega_{LF})$: -

$$V_1^{LF} = \frac{G_1}{\tau_I s}\,\mathrm{Im}(V_{IN} + V_Q)$$

ω_{LF} is the frequency of the slowly varying input signal $(V_{IN} = aV_S - V_T)$ and G_1 is the overall gain factor, including the synchronous rectifier. The symbol V_1^{LF} represents the low frequency complex representation. The symbols V_Q etc remain the complex representation of the AC signals at the operating frequency ω_C.

The output of the multiplier is: $$V_2 = \frac{G_2 V_S \times V_1^{LF}}{10V}$$

The differentiator provides the accurate (90 deg) phase shift: $V_Q = j\frac{\omega_C L_M}{R} V_2$

1. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators".

The injected signal is now substantially quadrature: V_2 has a small quadrature component and the transformer introduces a very small phase error, equivalent to an imaginary component of L_M. I shall represent the small error with a factor $(1+\delta)$: -

$$\mathrm{Im}(V_Q)(1+\delta) = \frac{\omega_C L_M}{R} V_2 \quad \text{with} \quad |\delta| << 1 \quad (\delta \text{ is a small and possibly complex number})$$

Combine the equations above to reveal in the (low frequency) complex representation $(s = j\omega_{LF})$: -

$$\mathrm{Im}(V_Q)(1+\delta) = \frac{\omega_C L_M}{R} \frac{G_2 V_S}{10V} \frac{G_1}{\tau_I s} \left(\mathrm{Im}(V_{IN}) + \mathrm{Im}(V_Q)\right)$$

The various gain factors and integrator characteristic represent a high gain block [1]: -

$$\mathrm{Im}(V_Q)(1+\delta) = -H(s)\left(\mathrm{Im}(V_{IN}) + \mathrm{Im}(V_Q)\right) \quad \text{with} \quad H(s) = -\frac{\omega_C L_M}{R} \frac{G_2 V_S}{10V} \frac{G_1}{\tau_I s}$$

The sign convention is chosen to be consistent with the usual definition of the open loop characteristic of a high gain block (from the non-inverting input to the output) [1]. With a little algebra: -

$$\mathrm{Im}(V_Q)\left(1 + \frac{1+\delta}{H(s)}\right) = -\mathrm{Im}(V_{IN})$$

At very low frequency the magnitude of the high gain block is very large and the term $|\delta / H(s)|$ is truly negligible. One can deduce the transfer function, to a very good approximation: -

$$\left|\frac{\delta}{H(s)}\right| << 1 \;\Rightarrow\; T(s) = \frac{\mathrm{Im}(V_Q)}{\mathrm{Im}(V_{IN})} \approx \frac{-H(s)}{1+H(s)} = \frac{-1}{1+\tau_Q s} \quad \text{with} \quad \tau_Q = \frac{R}{\omega_C L_M} \frac{10V}{G_2 V_S} \frac{1}{G_1} \tau_I$$

The interpretation is simple: The response is that of an inverter (gain = -1) with a first order low pass characteristic.

At very low frequency, to a good approximation: -

$$\tau_Q \omega_{LF} << 1 \;\Rightarrow\; |H(s)| >> 1 \;\Rightarrow\; \mathrm{Im}(V_Q) \approx -\mathrm{Im}(V_{IN}) \;\Rightarrow\; \mathrm{Im}(V_D) \approx 0$$

Given a step-change in quadrature the servo approaches balance with the usual exponential decay to zero (null balance): -

$$\mathrm{Im}(V_D) = \mathrm{Im}(V_D(t=0))\exp\left(-\frac{t}{\tau_Q}\right)$$

Five time constants $(t = 5\tau_Q)$ is usually regarded as sufficent settling time and, very approximately, a time constant of 1second is a reasonable design target.

1. Part 4, monograph 1: “High gain blocks”.

6.4 Interaction between the in-phase and quadrature balance

The in-phase balance algorithm is usually best implemented with a microcontroller or external computer. The most appropriate algorithm depends on circumstances. Rapid balance from a large change (e.g. switching to a different thermometer or resistor) usually involves initiating a sequence: set a low gain, measure the out-of-balance, calculate and correct the transformer ratio accordingly and repeat with a progressively increasing gain until the required resolution is achieved. Once null balance is achieved the algorithm can then switch to a tracking mode – similar to a linear feedback system with a suitably fixed high gain setting and time constant.

If the main gain/filter stages are common to both in-phase and quadrature servo there is a high degree of interaction. With a low gain setting, for example, the quadrature servo time constant is large and the servo is slow to achieve balance. As the gain is increased the servo speeds up. Fortunately this is not a problem since an accurate quadrature null balance is required only at the highest gain setting. With a suitable transformer, variable gain differential amplifier stage (G_2) and integrator time constant (τ_I) a reasonably good compromise between servo range and settling time can be achieved. The quadrature servo can be left to run continuously with only a small effect on the in-phase null balance process. Similarly, small changes to the transformer ratio have negligible effect on the quadrature null balance.

The time constant for the servo feedback loop is, from the previous section: -

$$\tau_Q = \frac{R}{\omega_C L_M}\frac{10V}{G_2 V_S}\frac{1}{G_1}\tau_I$$

A reasonable target is of the order 1 second (5s settling time from a large change). The optimum design considerations are as follows: -

1). The AC gain of the main amplifier/filter stages (G_1) needs to be sufficiently large so that in-phase accuracy is limited only by noise (typically $\approx 1nV_{RMS}/\sqrt{Hz}$ referred to the input of the pre-amplifier). The DC error at the input of the final stage synchronous rectifier/low-pass filter is typically <0.1mV hence the gain needs to be at least 10^5 (preferably 10^6) as long as the synchronous rectifier is not overloaded by interference. Two sets of variable gain/filter stages are usually sufficient (see section 3) though it would be wise to monitor the residual signal level and set the maximum gain accordingly. A reasonable assumption is, therefore, a maximum gain setting in tracking mode: -

Maximum gain setting: $G_1 \approx 10^6$

2). The gain of the differential amplifier, G_2, needs to be variable to compensate for large variations in operating current and reference resistor [1]. The gain is chosen so that the signal (V_{REF}) is optimum at the analogue multiplier input (typically 2 -20$V_{PK\text{-}PK}$). The choice of gain also affects the actual range of the quadrature servo though that is not critical. A reasonable assumption is $V_{REF} \approx 3V_{PK}$ (give or take a factor of three up or down). It is necessary, however, to measure the reference voltage (V_{REF}), with a further synchronous rectifier/low-pass filter and ADC channel, in order to determine the appropriate gain setting.

For practical values (approximately ×10, ×100 and ×1000) see section 6 [1]. It is debateable whether a fine gain adjustment stage is necessary. Assume: -

$$V_{REF} = G_2 V_S \approx 3V_{PK}$$

3). Typical values for the differentiator are $f_C = 25Hz$ or $f_C = 75Hz$ [1: section 5]: -

$$\omega_C = 157 \text{ or } = 471 \quad L_M \approx 100\mu H \text{ and } R = 10k\Omega \Rightarrow \frac{\omega_C L_M}{R} \approx 157\times10^{-8} \text{ or } \approx 471\times10^{-8}$$

1. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators".

4). The optimum time constant of the integrator can now be calculated. From above: -

$$\tau_I = \frac{\omega_C L_M}{R}\frac{G_2 V_S}{10V} G_1 \tau_Q \approx 157 \times 10^{-8} \times 0.3 \times 10^6 \times 1s \approx 0.47s$$

This is easily implemented with switched gain stages (typically ×1, ×10 or ×100) and a fine gain adjustment stage based on a multiplying digital to analogue converter (12-bit should suffice).

Even though the gain of the shared amplifier/filter stages can vary over a wide range the quadrature servo integrator time constant can be fixed.

The range of the quad servo can then be determined: -

$$b_{MAX} = \frac{R_T}{R_S}\left(\theta_S - \theta_T\right)_{MAX} = \frac{\omega_C L_M}{R} G_2$$

Appendix 1: Basic noise theory

A1.1. Johnson noise

In the early development of radio and thermionic devices Johnson [1] and Nyquist [2] (experimentally and theoretically, respectively) realised that there is a fundamental relationship between the random voltage observed at the terminals of a resistor, absolute temperature and bandwidth. The modern expression is: -

$$V_N = \sqrt{4kTRB} \quad \text{(Volts RMS)}$$

Where: -

V_N Is one standard deviation (the RMS value) of a very large number of instantaneous samples of voltage measurement and is referred to as the "noise voltage". The RMS value is also related to power.

$k = 1.38 \times 10^{-23} JK^{-1}$ is Boltzmann's constant (relating random kinetic (thermal) energy and temperature).

T = Absolute temperature (Kelvin)

R = Resistance (Ohm's)

B = Bandwidth (Hz) - of the (ideal) measuring instrument

Nyquist's theory is based on classical concepts in statistical thermodynamics (i.e. not quantum) and transmission line theory (see section 4 of this appendix). The "free" electrons in a conductor, with their highly random motion, behave like a perfect monatomic gas in thermal equilibrium with their surroundings. This form of noise is called "thermal" and is extremely random – conforming to the "normal" (Gaussian or "bell curve") statistical distribution.

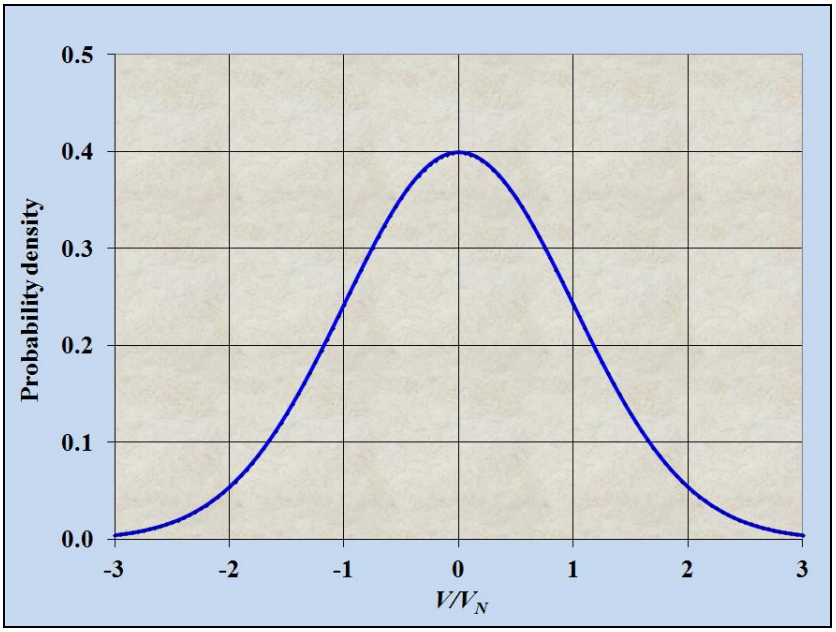


Fig. A1.1.1 The normal (Gaussian) probability distribution

Given a large number, N, of instantaneous samples, V_i , with an average of zero, the standard deviation, V_N , is defined by: -

$$V_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N} V_i^2}$$

For a normal distribution 68.3% of samples, on average, fall within the range $0 \pm V_N$. Measuring instruments with resolution limited by noise often quote two standard deviations (95.5%) or even three standard deviations (99.7%).

1. Johnson, J. B.: "Thermal Agitation of Electricity in Conductors". Phys. Review 32 (1928). P97-109.
2. Nyquist, H: "Thermal Agitation of Electric Charge in Conductors". Phys. Review 32 (1928). P110.

It is shown in many texts on probability and statistics that the RMS of a number of such random variables added together is: -

$$V_T = \sqrt{V_{N1}^2 + V_{N2}^2 + V_{N3}^2 + \ldots}$$

Whereas the instantaneous measured values of voltage add arithmetically, according to one of Kirchhoff's laws (conservation of energy), the statistical parameters, RMS, add according to a form of Pythagoras' theorem – if the random variables are statistically independent they are, in a general sense, orthogonal (mutually perpendicular).

If, for example, one connected a number of resistors in series (at the same temperature) the total voltage developed across them would have an RMS value corresponding to a single resistor with the same total resistance.

$$V_T = \sqrt{4kTB(R_1 + R_2 + R_3 + \ldots)}$$

It would seem (and it is no accident) that the quantity of interest has the units of (electrical) power. The "noise power" depends only on temperature and bandwidth: -

$$P_N = \frac{V_N^2}{R} = 4kTB$$

The concepts of "noise power" and "noise temperature" turn out to be as useful, if not more so, compared to the concept of "noise voltage". If, for example, one connected a number of resistors in parallel the "noise power" would be, as expected: -

$$P_N = V_T^2\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots\right) = \frac{V_T^2}{R_1} + \frac{V_T^2}{R_2} + \frac{V_T^2}{R_3} + \ldots$$

It is as if each resistor is able to supply a certain amount of power (e.g. to a very cold load resistor) and these add arithmetically to equal the total power, according to the law of conservation of energy. Unfortunately it is not that simple – not all of the "noise power" is delivered to the load (see below).

If one short circuits the terminals of a resistor a random "noise current" would flow according to Ohm's law, which remains valid for RMS values: -

$$I_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N} I_i^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \frac{V_i^2}{R^2}} = \frac{V_N}{R} = \sqrt{\frac{4kTB}{R}} \quad \text{(RMS)}$$

The product of "noise voltage" and "noise current" is, not surprisingly, the "noise power" as defined above: -

$$V_N I_N = P_N$$

In practice most applications fall somewhere between open circuit and short circuit. If one were to connect a warm source resistor to a cold load resistor heat would flow down the connections, in the form of a random AC current. The rate at which the heat energy flows can be interpreted as (random) electrical power: -

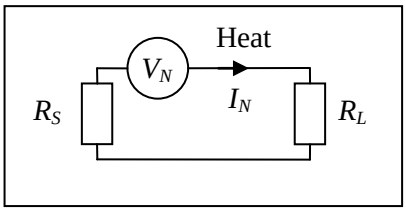


Fig. A1.1.2 A warm source and cold load

Ohm's law remains valid for the RMS values and, therefore: - $I_N = \dfrac{V_N}{R_S + R_L}$

The power delivered to the load is: $P_L = I_N^2 R_L = \dfrac{V_N^2 R_L}{(R_S + R_L)^2}$

The maximum power delivered to the load is found with simple calculus – the gradient is zero. Also, multiply top and bottom by $(R_S + R_L)^3$ and divide by V_N^2: -

$$\frac{\partial P_L}{\partial R_L} = \frac{V_N^2}{(R_S + R_L)^2} - 2\frac{V_N^2 R_L}{(R_S + R_L)^3} = 0 \quad \Rightarrow \quad R_L = R_S$$

RF engineers are familiar with this – maximum power transfer and best noise performance is when the source and load resistance are equal. Matching also prevents reflections – a particular problem with high frequency (including "digital") circuits and transmission lines. **If the load is matched to the source the actual power flow is one quarter of the "noise power"**: -

$$R_L = R_S \quad \Rightarrow \quad P_L = \frac{V_N^2}{4R_L} = \frac{P_N}{4}$$

A1.2. Low noise pre-amps

The most basic model of a real amplifier is an ideal (noiseless) amplifier with a noise voltage source in series with one of the inputs and a (statistically independent) noise current source in parallel. There may be two or more statistically independent mechanisms that give rise to these but if one assumes that all practical measures have been taken to minimise the combined effect one may as well treat each as a single source.

The noise current flows through the source impedance generating a second noise voltage in series. The source impedance may also generate noise and so the complete model is as follows: -

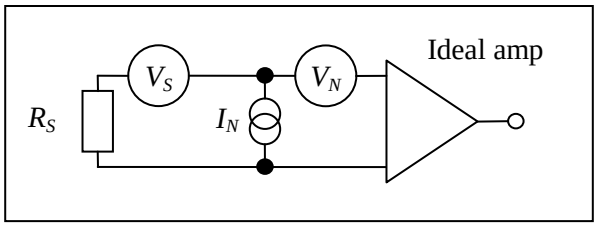


Fig. A1.2.1 A basic noise model (resistive source)

Once again one can interpret the symbols as the RMS value of the probability distributions over a specified range of frequency. Low noise amplifiers and voltage references are often specified, for example, in the range 0.1 to 10Hz. The RMS value of the net signal at the input of the amplifier is calculated according to: -

$$V_T^2 = V_S^2 + V_N^2 + I_N^2 R_S^2 \quad \text{(RMS values)}$$

More generally noise is specified and measured/displayed in terms of spectral density – with units of either V^2/Hz or $V/\sqrt{Hz}$, defined in the usual way: employing the symbol Δ to aid interpretation: -

$$\frac{d(V_T^2)}{df} = \lim_{\Delta f \to 0} \frac{\Delta(V_T^2)}{\Delta f} \quad (\text{units: } V^2/Hz) \quad \text{and} \quad \bar{e}_n = \sqrt{\frac{d(V_T^2)}{df}} \quad (\text{units: } V/\sqrt{Hz})$$

Where $\Delta(V_T^2)$ is the small but finite noise voltage (squared) in the small but finite bandwidth Δf.

Noise spectral density often varies with frequency. The total noise, referred to the input of the amplifier in the frequency range f_1 to f_2 is, therefore: -

$$V_T^2 = \int_{f_1}^{f_2} \left(\frac{d\left(V_S^2\right)}{df}(f) + \frac{d\left(V_N^2\right)}{df}(f) + \frac{d\left(I_N^2\right)}{df}(f) R_S^2 \right) df$$

What if the source is partly inductive or capacitive? One can think of the current noise, in each infinitesimal bandwidth, as a pure sinewave, albeit with random magnitude and phase. The phase distribution is uniform (the same probability density from 0 to 360 degrees) and remains so regardless of the imaginary part of the source impedance. The voltage developed does, however, depend on the magnitude of the source impedance and, therefore, the frequency. The contribution from the noise current is, therefore: -

$$R_S \to Z_S \quad \Rightarrow \quad \int_{f_1}^{f_2} \frac{d\left(I_N^2\right)}{df}(f) \left| Z_S(f) \right|^2 df$$

In practice most datasheets specify noise in units of $V/\sqrt{Hz}$ though often omitting the words "spectral density". Publications frequently include equations that appear to be dimensionally inconsistent. The symbol $\bar{e}_n$, for example, would normally be associated with a voltage: -

$$\bar{e}_n = \sqrt{\frac{d\left(V_T^2\right)}{df}}$$

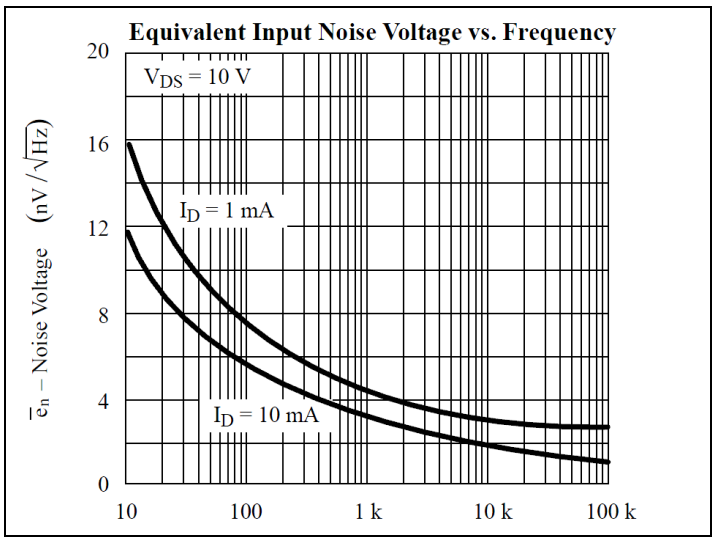


Fig. A1.2.1 Noise voltage spectral density for a low noise dual JFET (U430)

Fortunately most applications are concerned with a narrow range of frequency over which the noise spectral density can be regarded as constant so that calculations are quite simple: -

$$\Delta V_N^2 = \frac{d\left(V_N^2\right)}{df} \Delta f = \left(\bar{e}_n\right)^2 \Delta f \quad \Rightarrow \quad \Delta V_N = \bar{e}_n \sqrt{\Delta f} \quad \text{etc.}$$

Typical values for a dual matched JFET (U430) are: -

Current (mA)	$V_N \left(nV/\sqrt{Hz}\right)$	$I_N \left(fA/\sqrt{Hz}\right)$	P_N (W)	R_N (MΩ)	T_N (K)
1	4	≈1 (@25ºC)	4×10^{-24}	4.0	0.07

A1.3 Optimum noise matching

The minimum value for V_T coincides with the minimum value of V_T^2 so that, with a little algebra, one can find the required condition, assuming the source noise and source resistance are constant: -

$$d\left(V_T^2\right)=0 \;\Rightarrow\; 2V_N dV_N + 2R_S^2 I_N dI_N = 0$$

It is also reasonable to assume that the noise power of the pre-amp is constant – one can trade off a reduced noise voltage at the expense of increased noise current, pro-rata, so that: -

$$P_N = V_N I_N \text{ (constant)} \;\Rightarrow\; dV_N = -P_N \frac{dI_N}{I_N^2} = -V_N \frac{dI_N}{I_N} \;\Rightarrow\; -\frac{V_N^2}{I_N} dI_N + R_S^2 I_N dI_N = 0 \;\Rightarrow\; \frac{V_N}{I_N} = R_S$$

The ratio of noise voltage to noise current is the pre-amp "noise resistance": $\frac{V_N}{I_N} = R_N$

If the source is partly reactive: $R_N = |Z_S|$

Best noise performance is achieved, therefore, when the noise resistance of the pre-amp matches the magnitude of the source impedance at the operating frequency.

A1.4. Some notes on Nyquist's theory

With half a blank page left I could not resist some comments on Nyquist's astonishing theory. During the 1920s much progress was made in quantum theory (see, for example, [1]). Nyquist employed a similar concept in his theory of noise: the quantisation of the electromagnetic field due to boundary conditions. His argument goes something like the following: -

Consider an ideal transmission line (characteristic impedance *R*) with warm resistors (resistance *R*) at both ends. Thermal equilibrium is reached and then the two ends are simultaneously short circuited, trapping the electromagnetic energy (all power is reflected back from the short circuits). One can regard the length of line as "vibrating at its natural frequencies". Each mode of vibration is one degree of freedom with average energy *kT*. Nyquist deduced the density of modes per Hz of bandwidth and bingo!

In the last few of lines of his paper Nyquist discussed the latest (then) thinking on statistical thermodynamics in the light of the new quantum theory. He pointed out that the energy per degree of freedom (the principle of equal energy partition or "peep") consistent with Planck's theory is: -

$$E = \frac{h\nu}{\exp\left(\frac{h\nu}{kT}\right)-1}$$

$h = 6.626\times10^{-34}\,Js$ is Planck's constant and $\nu =$ frequency (Hz). The small value of Planck's constant means that for frequency range below optical (c.f. the "ultra-violet catastrophe"): -

$$\frac{h\nu}{kT} << 1 \;\Rightarrow\; \exp\left(\frac{h\nu}{kT}\right) \approx 1 + \frac{h\nu}{kT} \;\Rightarrow\; E \approx kT$$

1. Dirac P.A.M.: "The Quantum Theory of the Emission and Absorption of radiation". Proc. Royal Soc. (London) A114 pp. 243-265 (1927).

Appendix 2: Band-pass filter analysis

A2.1 The transfer function

A practical choice for the band-pass filter (inverting) is based on a single op-amp. Calculating practical component values can be a bit tricky but, fortunately, this has been done for the frequencies of interest. The choice of subscripts (*1A* and *1B*) in the following circuit will become clear later (the parallel combination).

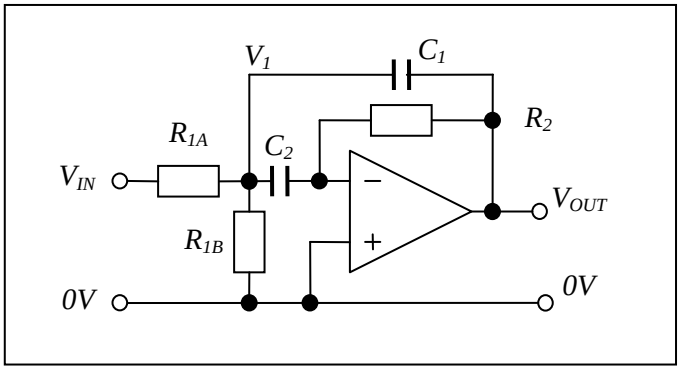


Fig. A2.1.1 A multiple feedback band-pass filter

If one assumes an ideal op-amp the inverting input is a virtual earth (0V) and, according to Ohm's law and Kirchhoff's laws in the complex representation $(s = j\omega)$: -

$$V_{OUT} = -V_1 sR_2C_2 \quad \text{and} \quad \frac{V_{IN} - V_1}{R_{1A}} + (V_{OUT} - V_1)sC_1 - V_1 sC_2 - \frac{V_1}{R_{1B}} = 0$$

$$\Rightarrow \quad \frac{V_{IN}}{R_{1A}} - \frac{V_1}{R_{1A}} + V_{OUT}sC_1 - V_1 sC_1 - V_1 sC_2 - \frac{V_1}{R_{1B}} = 0$$

Eliminate V_1: -

$$V_1 = -\frac{V_{OUT}}{sR_2C_2} \quad \Rightarrow \quad \frac{V_{IN}}{R_{1A}} + \frac{1}{R_{1A}}\frac{V_{OUT}}{sR_2C_2} + V_{OUT}sC_1 + \frac{V_{OUT}}{sR_2C_2}sC_1 + \frac{V_{OUT}}{sR_2C_2}sC_2 + \frac{1}{R_{1B}}\frac{V_{OUT}}{sR_2C_2} = 0$$

Multiply both sides by R_{1A} and sR_2C_2: -

$$sR_2C_2V_{IN} + V_{OUT}\left(1 + sC_1R_{1A}sR_2C_2 + R_{1A}sC_1 + R_{1A}sC_2 + \frac{R_{1A}}{R_{1B}}\right) = 0$$

The transfer function is, therefore: -

$$\frac{V_{OUT}}{V_{IN}} = \frac{-sR_2C_2}{s^2R_{1A}R_2C_1C_2 + sR_{1A}(C_1 + C_2) + (R_{1A} + R_{1B})/R_{1B}}$$

Multiply top and bottom by $R_{1B}/(R_{1A} + R_{1B})$ to reveal the standard form: -

$$\frac{V_{OUT}}{V_{IN}} = \frac{-2\xi G(s/\omega_N)}{s^2/\omega_N^2 + 2\xi(s/\omega_N) + 1}$$

ω_N is the natural frequency, ξ the damping ratio $(Q = 1/2\xi)$ and G is the peak gain (at the natural frequency).

Equating coefficients: $\frac{1}{\omega_N^2} = \frac{R_{1A}R_{1B}}{R_{1A}+R_{1B}} R_2C_1C_2 = R_1R_2C_1C_2$ with $R_1 = \frac{R_{1A}R_{1B}}{R_{1A}+R_{1B}}$

R_1 is R_{1A} in parallel with R_{1B} (hence the choice of subscripts).

Also: $\frac{2\xi}{\omega_N} = \frac{R_{1A}R_{1B}}{R_{1A}+R_{1B}}(C_1+C_2) = R_1(C_1+C_2)$ and $\frac{2\xi G}{\omega_N} = \frac{R_2R_{1B}}{R_{1A}+R_{1B}}C_2 = \frac{R_2}{R_{1A}}R_1C_2$

The natural frequency, damping ratio and peak gain are, therefore: -

$$\omega_N = \frac{1}{\sqrt{R_1R_2C_1C_2}} \qquad \xi = \frac{1}{2}\frac{R_1(C_1+C_2)}{\sqrt{R_1R_2C_1C_2}} \text{ and } G = \frac{R_2}{R_{1A}}\frac{C_2}{(C_1+C_2)}$$

In practice it is possible for the capacitors to have the same value: -

$$C_1 = C_2 \Rightarrow \omega_N = \frac{1}{\sqrt{R_1R_2}C} \qquad \xi = \sqrt{\frac{R_1}{R_2}} \text{ and } G = \frac{1}{2}\frac{R_2}{R_{1A}}$$

Some designers prefer the "Quality factor": $Q = \frac{1}{2\xi} = \frac{1}{2}\sqrt{\frac{R_2}{R_1}}$

Freq (Hz)	Q	R_2 (kΩ)	R_{1A} (kΩ)	R_1 (Ω)	R_{1B} (Ω)	C_1=C_2 (nF)
25	2.083	39	19.5	2247	2540	680
50	5	200	100	2000	2041	159
75	2.022	39	19.5	2385	2717	220
100	10	200	100	500	503	159
150	10	200	100	500	503	106

Fig. A2.1.2 Practical component values

A2.2 Bandwidth

The band-pass filter transfer function (with gain = 1) is, in normalised form $(s = j\omega/\omega_N = jx)$: -

$$T(s) = \frac{2\xi s}{s^2+2\xi s+1} = \frac{2\xi jx}{1-x^2+2\xi jx}$$

Divide top and bottom by $2\xi jx$: -

$$T(s) = \frac{1}{1+j\tan(\theta)} \text{ where } \tan(\theta) = \frac{1}{2\xi}\left(\frac{x^2-1}{x}\right)$$

This form of the transfer function can be quite useful for calculating magnitude and phase. First is the phase from an easily calculable form: -

$$\tan(\theta) = Q\left(x-\frac{1}{x}\right) \text{ where } Q = \frac{1}{2\xi} \text{ is the quality factor}$$

Then the magnitude: $$|T(s)| = \frac{1}{\sqrt{1+\tan^2(\theta)}} = \cos(\theta)$$

Note also: $$T(s) = \cos(\theta)\exp(-j\theta)$$

It also provides a quick route to the bandwidth, best defined by the -3dB points, where the gain is $\frac{1}{\sqrt{2}}$: -

$$|T(s)| = \frac{1}{\sqrt{2}} \quad \Rightarrow \quad \sqrt{1+\tan^2(\theta)} = \sqrt{2} \quad \Rightarrow \quad \tan(\theta) = \pm 1$$

This produces two quadratic equations each with two possible solutions – one needs to select the correct ones: -

$$\tan(\theta) = \frac{1}{2\xi}\left(\frac{x^2-1}{x}\right) = \pm 1 \quad \Rightarrow \quad x^2 - 1 = \pm 2\xi x$$

The solution for the quadratic of the form $ax^2 + bx + c = 0$ is by using the usual formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Case 1: -

$$x^2 + 2\xi x - 1 = 0$$

Case 2: -

$$x^2 - 2\xi x - 1 = 0$$

Providing solutions: -

$$x_1 = -\xi \pm \sqrt{\xi^2 + 1}$$

$$x_2 = \xi \pm \sqrt{\xi^2 + 1}$$

The positive square roots must be the valid solutions in both cases. Assume, for example, the damping ratio is very small (highly resonant with a narrow bandwidth). Taking the positive square root provides sensible values – just above and below the natural frequency: -

$$\xi << 1 \quad \Rightarrow \quad x_1 \approx 1 - \xi \quad \text{and} \quad x_2 = 1 + \xi$$

Taking the negative square root does not: $x_1 \approx x_2 \approx -1$

The difference gets us to the bandwidth: $$x_2 - x_1 = \xi + \sqrt{\xi^2 + 1} - \left(-\xi + \sqrt{\xi^2 + 1}\right) = 2\xi$$

The -3dB bandwidth is, therefore: $$\Delta\omega = 2\xi\omega_N$$

Appendix 3: Low-pass/notch filter analysis

A3.1 The basic low-pass filter

The basic idea is to employ a band-pass filter (inverting) and a conventional low-pass filter with two inputs: -

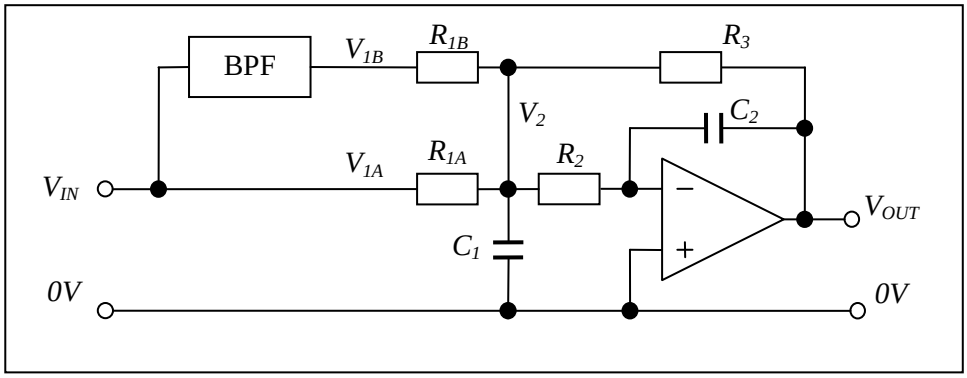


Fig. A3.1.1 A combined low-pass and notch filter.

If one assumes an ideal op-amp, Ohm's law and Kirchhoff's laws, in the complex representation $(s = j\omega)$: -

$$V_{OUT} = -\frac{1}{sR_2C_2}V_2 \quad \text{and} \quad I = \frac{V_{1A}-V_2}{R_{1A}} + \frac{V_{1B}-V_2}{R_{1B}} + \frac{V_{OUT}-V_2}{R_3} - \frac{V_2}{R_2} - V_2 sC_1 = 0$$

$$\Rightarrow \quad \frac{V_{1A}}{R_{1A}} + \frac{V_{OUT}sR_2C_2}{R_{1A}} + \frac{V_{1B}}{R_{1B}} + \frac{V_{OUT}sR_2C_2}{R_{1B}} + \frac{V_{OUT}}{R_3} + \frac{V_{OUT}sR_2C_2}{R_3} + \frac{V_{OUT}sR_2C_2}{R_2} + V_{OUT}s^2R_2C_2C_1 = 0$$

$$\Rightarrow \quad V_{OUT}\left(\frac{sR_2C_2}{R_{1A}} + + \frac{sR_2C_2}{R_{1B}} + \frac{1}{R_3} + \frac{sR_2C_2}{R_3} + \frac{sR_2C_2}{R_2} + s^2R_2C_2C_1\right) = -\frac{V_{1A}}{R_{1A}} - \frac{V_{1B}}{R_{1B}}$$

$$\Rightarrow \quad V_{OUT}\left(1 + sR_2C_2\left(\frac{R_3}{R_{1A}} + \frac{R_3}{R_{1B}} + \frac{R_3}{R_2}\right) + s^2R_3R_2C_2C_1\right) = -V_{1A}\frac{R_3}{R_{1A}} - V_{1B}\frac{R_3}{R_{1B}}$$

The result is a second order low pass filter (inverting) with two inputs. In normalised form $(s = j\omega/\omega_N)$: -

$$V_{OUT} = \left(V_{1A}\frac{R_3}{R_{1A}} + V_{1B}\frac{R_3}{R_{1B}}\right)\left(\frac{-1}{1 + 2\xi s + s^2}\right)$$

The damping ratio and natural frequency are: -

$$2\xi = R_2C_2\left(\frac{R_3}{R_{1A}} + \frac{R_3}{R_{1B}} + \frac{R_3}{R_2}\right)\omega_N \quad \text{with} \quad \omega_N = \frac{1}{\sqrt{R_3R_2C_2C_1}}$$

$$\Rightarrow \quad \xi = \frac{1}{2}\sqrt{\frac{R_2C_2}{R_3C_1}}\left(\frac{R_3}{R_{1A}} + \frac{R_3}{R_{1B}} + \frac{R_3}{R_2}\right)$$

A3.2 A basic band-pass filter

The 25Hz notch filter is less critical than in the amplifier/filter units and a much lower quality factor is tolerable. The most basic band-pass filter has a damping ratio of 1 (Q = 0.5): -

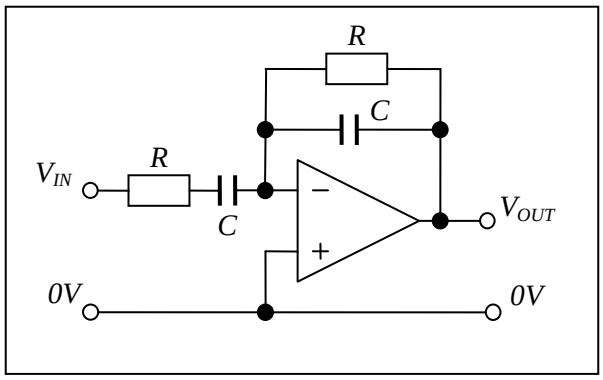


Fig. A3.2.1 A basic band-pass filter

If one assumes an ideal op-amp, Ohm's law and Kirchhoff's laws, the transfer function is, in the complex representation $(s = j\omega)$: -

$$\frac{V_{OUT}}{V_{IN}} = -\left(R \parallel \frac{1}{sC}\right) \div \left(R + \frac{1}{sC}\right)$$

With a little algebra: $$\frac{V_{OUT}}{V_{IN}} = \frac{-sRC}{1 + 2sRC + s^2R^2C^2}$$

The damping ratio and natural frequency are: -

$$2\xi = 2RC\omega_N \quad \text{with} \quad \omega_N = \frac{1}{RC} \quad \Rightarrow \quad \xi = 1$$

The maximum gain is 0.5

Practical component values (for 25Hz) are $C = 470nF$ and $R = 13.5k\Omega$ (An 8k2 fixed resistor plus a 10k variable).

Low noise BJT pre-amplifiers

1. Introduction

The following circuits employ a differential "long tail pair" of matched low noise bipolar junction transistors (BJTs). The differential structure and constant current source ensure a high level of rejection of common mode voltage at the inputs and power supply variations. DC offset is low over a wide range of temperature.

Noise performance is excellent in applications where the source impedance is between 1kΩ and 100kΩ. If the source impedance is lower than 1kΩ it is best to use an impedance matching transformer. If the source impedance is significantly higher than 100kΩ (e.g. with a capacitance bridge) the contribution due to noise current becomes dominant and it is better to use a pre-amplifier based on a matched pair of JFETs. For more details see the monograph "A low noise JFET pre-amplifier" by the same author.

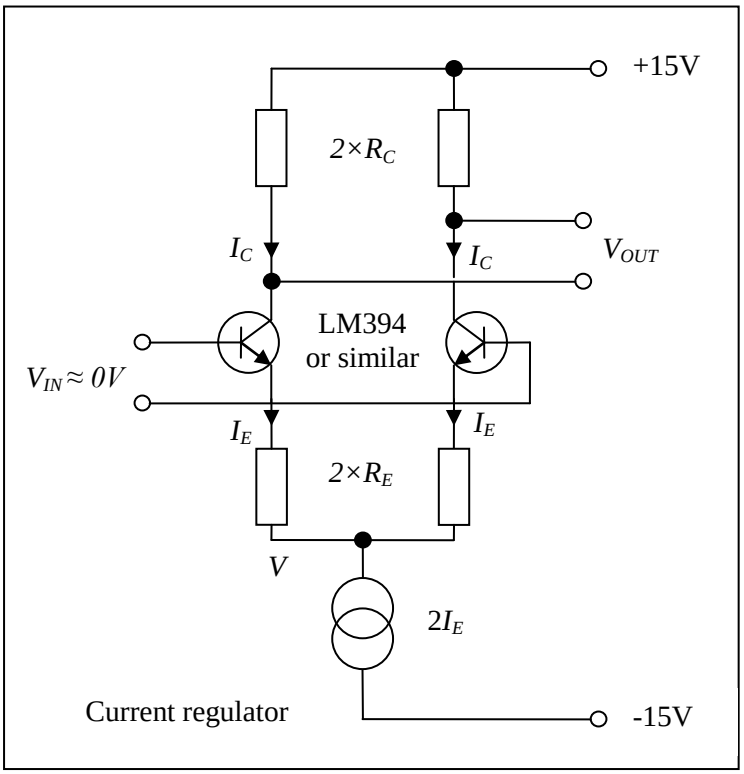


Fig. 1 Outline schematic of a long-tail pair.

As a rule of thumb the gain is, to a reasonably good approximation (see section 3 for details): -

$$G = \frac{dV_{OUT}}{dV_{IN}} \approx \frac{R_C}{R_E + R_{EO}}$$

R_C is the collector resistance and $R_E + R_{EO}$ is the emitter resistance (component) in series with the emitter output resistance. The parameter R_{EO} is easily calculated from the operating current.

For a more accurate calculation replace the collector resistance with the effective collector resistance (the collector resistance in parallel with the output resistance of the transistor).

The long tail pair can be used on its own, if accuracy and stability of the gain is not critical (e.g. as the pre-amp of a null detector) or as the first stage, inside the loop, of a composite amplifier/integrator, employing an operational amplifier with negative feedback for greater accuracy (e.g. as the first stage of a high accuracy voltage follower).

The choice of operating current depends mainly on the source resistance and the need for optimum impedance matching (see section 3 for details).

2. A review of basic BJT theory (low frequency)

2.1 Current gain and output conductance

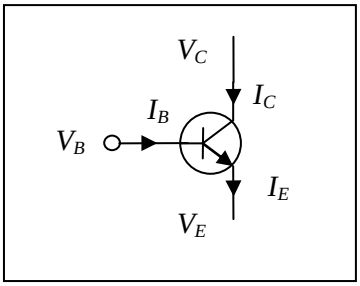


Fig. 2.1.1 A basic model for the BJT $\left(I_E = I_C + I_B\right)$

The collector current depends mainly on the base-emitter current but it is also a function of the collector-base voltage due to "base width modulation". The latter effect is small but not negligible. For a given V_{CB} the collector current is approximately proportional to the base current: -

$$I_C = I_C\left(I_B, V_{CB}\right) \approx h_{DC} I_B \qquad \text{with } h_{DC} = \frac{I_C}{I_B} \text{ approximately constant}$$

Where h_{DC} is usually referred to as the "DC current gain". In order to analyse circuits, however, one needs to model small signal (AC) characteristics, including the dependence on V_{CB}: -

$$dI_C = \frac{\partial I_C}{\partial I_B} dI_B + \frac{\partial I_C}{\partial V_{CB}} dV_{CB}$$

The first partial derivative defines the small signal current gain and the second defines the small signal output conductance (in data sheets often referred to as h_{FE} and h_{OE} respectively): -

Define the parameters: $\qquad h_{FE} = \frac{\partial I_C}{\partial I_B} \qquad h_{OE} = \frac{\partial I_C}{\partial V_{CB}} = \frac{1}{R_{CO}}$

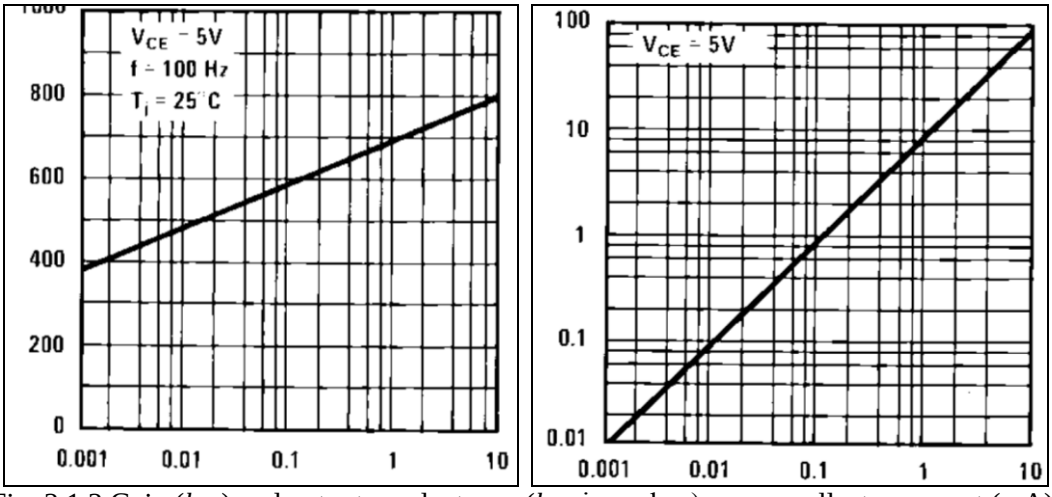


Fig. 2.1.2 Gain (h_{FE}) and output conductance (h_{OE} in μmhos) versus collector current (mA).
(LM394 datasheet, Dec. 1994. Courtesy National Semiconductor corp.)

The small signal current gain and output conductance increase with collector current and, for the LM394 type: -

$$h_{FE} \approx 105 \log_{10}\left|I_C\right| + 1010 \qquad \text{and} \qquad h_{OE} = \frac{1}{R_{CO}} \approx \frac{I_C}{100}$$

For most types of matched BJT pairs the output conductance is accurately proportional to collector current (over many decades), though the constant of proportionality may vary. The small signal current gain, on the other hand, can vary substantially.

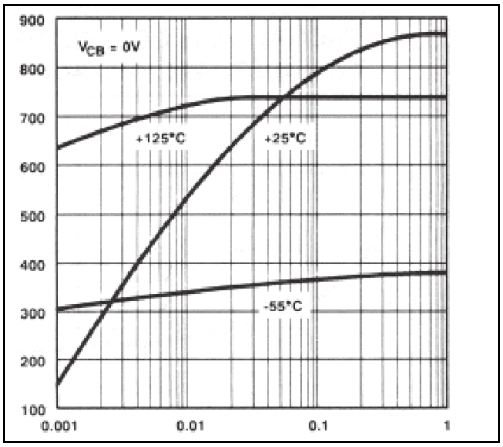


Fig. 2.1.3 Gain (h_{FE}) versus collector current (mA) for an SSM2210 (Courtesy Analogue Devices).

The DC current gain is, however, sufficiently large so that one can assume, for most practical purposes: -

$$h_{DC} >> 1 \quad \Rightarrow \quad I_C >> I_B \quad \text{and} \quad I_E = I_C + I_B \quad \Rightarrow \quad I_E \approx I_C \quad \text{and} \quad dI_E \approx dI_C$$

Less obvious is the relationship between the small signal current gain and the DC current gain. The latter is always lower but not usually given in data sheets. It is easily measured but, fortunately, the difference is sufficiently small for a design rule of thumb, as the following nice bit of analysis demonstrates: -

The collector current is approximately proportional to base current (the gradient increases slightly) and a reasonably good fit is, therefore, the linear equation: -

$$I_C \approx kI_B \qquad \text{with } k \text{ a constant}$$

The gradient (h_{FE}) is roughly proportional to the log of collector current and a reasonably good model is an equation of the form: -

$$h_{FE}(I_C) = \frac{\partial I_C}{\partial I_B} \approx \alpha \ln(I_C) + \gamma \quad \Rightarrow \qquad \frac{\partial I_C}{\partial I_B}(I_B) \approx \alpha \ln(kI_B) + \gamma$$

Integrating results in: $I_C \approx \alpha(\ln(kI_B) - 1)I_B + \gamma I_B$

The constant of integration must be zero so that: $I_B = 0 \quad \Rightarrow \quad I_C = 0$

It is then clear that the DC current gain is: $h_{DC} = \frac{I_C}{I_B} = \alpha(\ln(kI_B) - 1) + \gamma \quad = \alpha \ln(I_C/e) + \gamma$

The difference, at the same operating current, expressed as a ratio is: $\frac{h_{FE} - h_{DC}}{h_{FE}} = \frac{\alpha}{h_{FE}}$

For the LM394 at $I_C = 1mA;\ h_{FE} \approx 700$ and $\alpha = \frac{105}{\ln|10|} \approx 46$. The error is about 7% and one has, at least for the first iteration of the design process, the rule of thumb: $h_{DC}(I_C) \approx h_{FE}(I_C)$

N.B. a better estimate is: $h_{DC}(I_C) \approx h_{FE}(I_C/e)$ (i.e. h_{FE} at about one third the operating current.)

2.2 Base current

Commercially available matched pairs have properties closely matching the behaviour of an ideal BJT. Specifically, to a very good approximation, the base-emitter junction behaves like an ideal PN junction: -

$$V_{BE} = V_T \ln\left|\frac{I_B}{I_S}\right| \qquad \text{or} \qquad I_B = I_S \exp\left(\frac{V_{BE}}{V_T}\right)$$

I_S = a small current depending on the doping level of the silicon.

$k = 1.38\times10^{-23}\,J/K$ is Boltzman's constant.

T = absolute temperature (nominally 293K at 20 ºC)

$e = 1.6\times10^{-19}C$ is the quantum of charge (e.g. on a proton)

$V_T = \frac{kT}{e} \approx 25mV$ (at 20ºC) is a voltage corresponding to thermal energy of the electrons.

Note that V_T is temperature sensitive and PN junctions are often employed as linear temperature sensors: -

$$\frac{dV_T}{dT} = \frac{k}{e} \approx 86\,\mu V/\deg .$$

Note also that a measure of conformity is the linearity of V_{BE} versus log(I_C). Linearity is typically better than 0.1%, over a very wide range of collector current (100pA to 1mA as V_{BE} varies from about 200 to 600mV), and this property is widely used for log/antilog and multiplier/divider circuit functions.

2.3 Emitter output resistance

The output resistance at the emitter is, from above, assuming a constant temperature: -

$$dI_B = \frac{I_S}{V_T}\exp\left(\frac{V_{BE}}{V_T}\right)dV_{BE} \qquad \Rightarrow \qquad dI_B = \frac{I_B}{V_T}dV_{BE}$$

From section 2.1: $$dI_C = h_{FE}\frac{I_B}{V_T}dV_{BE} + \frac{dV_{CB}}{R_{CO}}$$

The small signal current gain (h_{FE}) is approximately the same as the DC current gain (see section 2.1) so that, in many cases, it is sufficiently accurate to conclude: -

$$h_{FE}I_B \approx I_C \qquad \Rightarrow \qquad dI_C \approx \frac{I_C}{V_T}dV_{BE} + \frac{dV_{CB}}{R_{CO}}$$

Hold the collector and base voltages constant and take a small bit of extra current from the emitter (by an amount dI_E). The emitter voltage must decrease so that V_{BE} increases by dV_{BE}. One can interpret this in terms of a voltage source in series with a resistor - the emitter output resistance: -

$$dV_{CB} = 0 \quad \text{and} \quad dI_C \approx dI_E \qquad \Rightarrow \qquad R_{EO} = \frac{dV_{BE}}{dI_E} \approx \frac{V_T}{I_C} \approx \frac{25mV}{I_C}$$

Rule of thumb: 25Ω at 1mA of current.

N.B. This low output resistance is often useful hence the frequently employed emitter follower circuit as an output stage, especially when impedance matching to a 50Ω transmission line $(I_C = 0.5mA)$.

A useful corollary is the reasonably accurate model: $$dI_C \approx \frac{dV_{BE}}{R_{EO}} + \frac{dV_{CB}}{R_{CO}}$$

2.4 BJT noise performance

The basic noise model is an ideal (noise free) amplifier with a noise voltage source in series with the input and a noise current source in parallel with the source impedance (usually a source resistance). The source resistance also generates noise, V_S, depending on resistance and temperature (Johnson noise): -

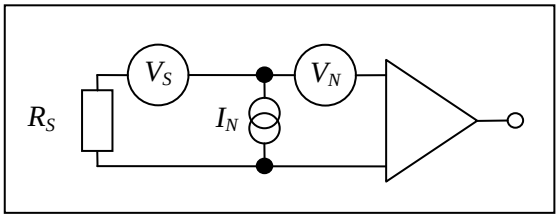


Fig. 2.4.1 Basic noise model for the amplifier

The noise current flows through the source impedance creating a third noise voltage source in series with V_N. The three sources are statistically independent and conform to a normal (Gaussian) probability distribution and so combine as follows: -

The total noise voltage at the input is: $V_T = \sqrt{V_S^2 + (I_N R_S)^2 + V_N^2}$ with $V_N = \sqrt{V_{N1}^2 + V_{N2}^2}$

Where V_S, V_N, etc. and I_N represent the standard deviation (root mean square - RMS) of the probability distribution. V_N. and I_N are normally specified, in data sheets, in terms of spectral density (volts and amps per square root of bandwidth) and as a (graph) function of frequency (see fig. 2.4.2). Johnson noise, on the other hand, can be easily calculated and the spectral density does not vary with frequency.

The random voltage, as measured across the terminals of a resistor (open circuit), and the random current that would flow through a short circuit can be characterised as random variables with normal distributions with RMS values: -

$$V_S = \sqrt{4kTR_S B} \quad \text{and} \quad I_S = \sqrt{\frac{4kTB}{R_S}} \quad \text{so that} \quad P_J = V_S I_S = 4kTB$$

Where: -

$k = 1.38 \times 10^{-23}\, J/K$ = Boltzmann's constant.

$T =$ is the absolute temperature ($T \approx 293K$ at 20ºC).

$B =$ Bandwidth of the measuring instrument, usually a spectrum analyser and often specified as 1Hz.

$P_J = 4kTB$ is the "Johnson noise power". It does not refer to any actual flow of power but it is a useful concept.

For an ideal BJT as the collector current increases the transistor input noise voltage reduces but the noise current increases. The product (noise power) remains roughly constant. It can be shown that, for a useful range of (high) frequency and (low) collector current for an ideal BJT: -

$$V_{N1} = kT\sqrt{\frac{2B}{eI_C}}$$

Where: -

$V_{N1} =$ Component 1 of the total noise voltage.

$e = 1.60 \times 10^{-19} C$ is the quantum of electrical charge (i.e. on a proton).

$I_C =$ Operating (collector) current.

See later for component 2.

The other main source of noise is primarily due to the base current. It is remarkable that the quantisation of charge is directly responsible for this observable phenomenon. With a base current, I_B, the average number of electrons flowing through the base in time, T, is: -

$$N = \frac{I_B T}{e}$$

Where $e = 1.60 \times 10^{-19} C$ is the quantum of electrical charge.

These are thermally generated electrons and the rate at which they pass through the base is highly random. According to a fundamental theory of probability the number can vary (RMS or one standard deviation) by $\sqrt{N}$.

The RMS current is, therefore: $$I_N = \frac{e\sqrt{N}}{T} = \sqrt{\frac{eI_B}{T}}$$

This is an average over a time T and a well known theory (Fourier transforms) equates this process of averaging to an equivalent ideal filter with bandwidth $2B$ Hertz.

$$\frac{1}{T} \equiv 2B \quad \Rightarrow \quad I_N = \sqrt{2eI_B B}$$

In practice one can select the collector current which is related to the base current by the DC current gain: -

$$I_C = h_{FE} I_B \;\Rightarrow\; I_N = \sqrt{\frac{2eI_C B}{h_{FE}}}$$

h_{FE} = DC current gain of the transistor (approximately equal to the small signal current gain)

The product of noise voltage and noise current is the noise power: $$P_N = V_{N1} I_N = 2kTB\sqrt{\frac{1}{h_{FE}}}$$

In practice data sheets usually provide noise performance in 1Hz of bandwidth (B = 1).

The noise resistance and noise temperature are: $$R_N \approx \frac{V_{N1}}{I_N} = \frac{kT}{eI_C}\sqrt{h_{FE}} \quad \text{and} \quad T = \frac{P_N}{4k}$$

The formulae work reasonably well over a very wide range (decades) of frequency but only for low operating current. In practice there is a significant ("base spreading") resistance, R_{BB}, in series with the base which, being at ambient temperature, also generates noise. For the LM394 this is about 40Ω and for the SSM2210 it is typically 28Ω. Also, in datasheets, the small signal current gain is not specified accurately - noise calculations are very much "ball park" and one ends up relying on measured performance as part of the design optimisation process.

The total voltage noise is, therefore, a combination of the ideal BJT and the Johnson noise of the extra resistance: -

$$V_N = \sqrt{V_{N1}^2 + V_{N2}^2} \quad \text{with} \quad V_{N2} = \sqrt{4kTR_{BB}B}$$

In addition, at very low frequency (typically <20Hz) the voltage noise spectral density increases, at a rate inversely proportional to frequency (hence the name "1/f noise"), due to a further mechanism thought to be related to surface contamination of the transistors. The noise current spectral density also increases at low frequency but, for the LM394, the corner frequency increases with operating current, making the best choice of operating current a little more complicated. In practice it is found that an operating frequency of 25Hz or above and a low operating current (<1mA) avoids the problem of 1/f noise.

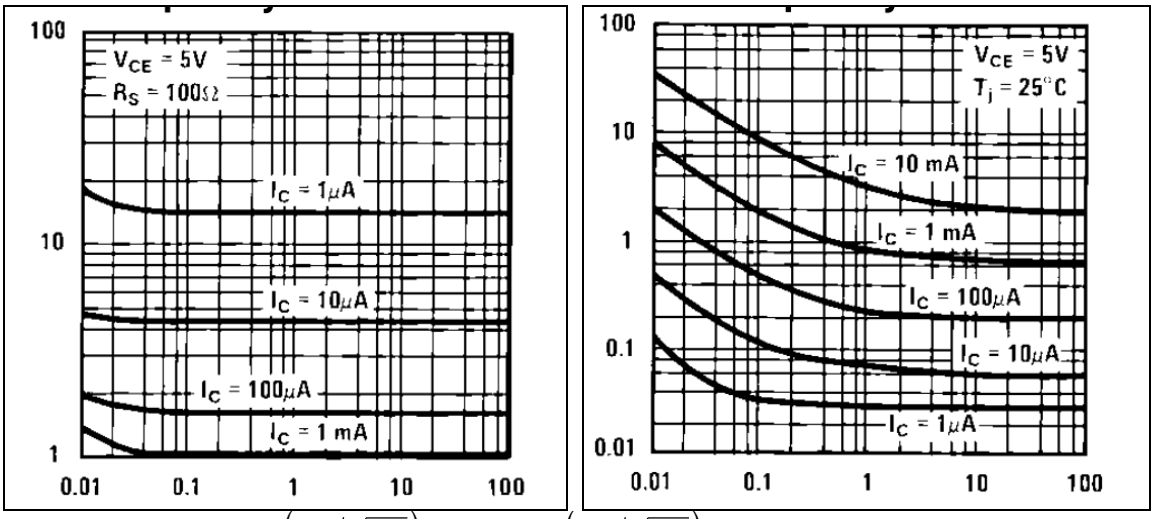


Fig. 2.4.2 Noise voltage $\left(nV/\sqrt{Hz}\right)$ and current $\left(pA/\sqrt{Hz}\right)$ versus frequency (*kHz*) for the LM394

Optimum noise performance is when the noise resistance of the transistor is the same as the source impedance (noise matching). The equivalent noise resistance of the amplifier is, for low operating current and high frequency, approximately: -

$$R_N \approx \frac{V_N}{I_N} = \frac{kT}{eI_C}\sqrt{h_{FE}}$$

Given the source impedance one could use this to select the optimum operating current but, in many cases, this is not sufficiently accurate. In practice it is necessary to use the datasheet information or, better still, the actual measured performance, rather than the theoretical model. Also, the source impedance is often too low for a direct match and it is better, therefore, to use an impedance matching transformer and a relatively high noise resistance pre-amplifier. The transformer ratio can be made selectable to cater for a wide range of source impedance [1].

2.4.1 Example calculation

It is interesting to compare the equivalent noise resistance of an LM394 BJT at, say, 1mA operating current and 20ºC. According to the datasheet the small signal current gain is typically 700 but can be as low as 300. If one employs the theoretical model and the typical value: -

$$R_N \approx \frac{kT}{eI_C}\sqrt{h_{FE}} \quad \text{and} \quad 300 < h_{FE} \approx 700 \quad \Rightarrow \; 438 < R_N \approx 690\Omega$$

If, on the other hand one relies on typical figures (fig. 2.4.2) from the data sheet.

From the datasheet at 25Hz and 1mA: $V_N \approx 1.1nV/\sqrt{Hz}$ and $I_N \approx 5\,pA/\sqrt{Hz} \quad \Rightarrow \quad R_N \approx 220\Omega$

The difference is considerable and measured performance is closer to the latter. The problem at low frequency is 1/f noise which is not predicted by the theory (a result of the manufacturing process). Fortunately this is still higher than the source impedance typically encountered and so the (higher than expected) noise current is not a problem. Also, noise performance only suffers when there is a major mismatch (typically >3:1). For lower noise current the SSM2210, with lower 1/f corner frequency, is preferable.

From the SSM2210 datasheet at 25Hz and 1mA: -

$$V_N \approx 1nV/\sqrt{Hz} \quad \text{and} \quad I_N \approx 1pA/\sqrt{Hz} \quad \Rightarrow \quad R_N \approx 1k\Omega$$

1. Part 3, monograph 5: “Noise matching transformers”

3. Analysis of the long tail pair

3.1 Voltage gain

For this analysis I shall assume, in addition to Kirchhoff's laws and Ohm's law: -
a) Perfectly matched transistors and resistors.
b) Ideal current source (the current is independent of the voltage, *V*).
c) Symmetrical inputs ($0V \pm dV_{IN}$) and outputs ($V_C \pm dV_{OUT}$).

With a load resistance in the collector circuit the collector-base voltage is no longer constant – the collector voltage decreases as the collector current increases.

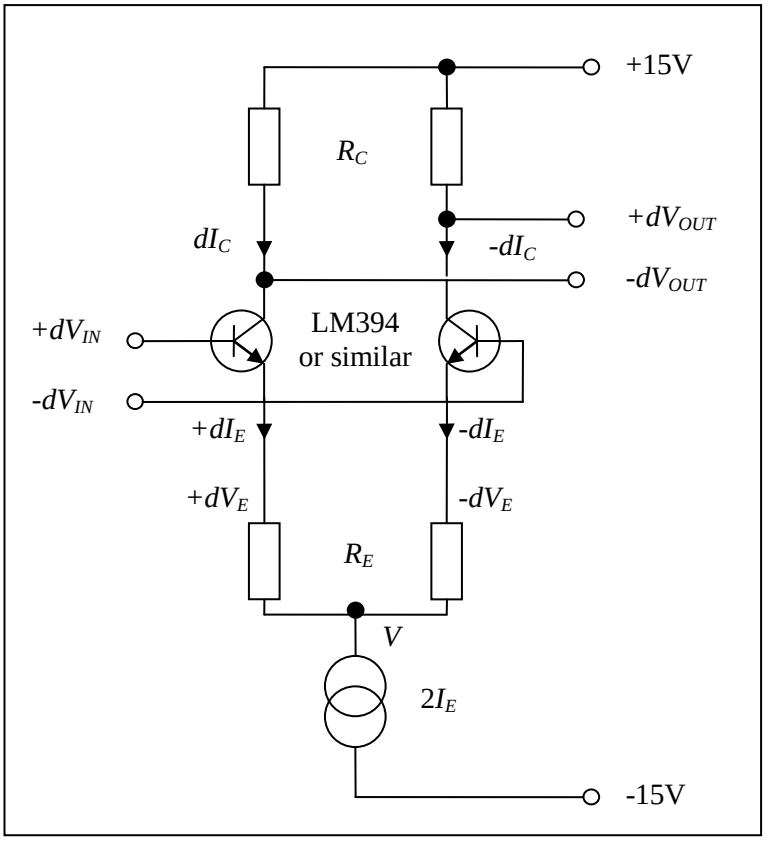


Fig. 3.1.1 Small signal analysis of the long-tail pair

From section 2.3: $$dI_C = \frac{dV_{BE}}{R_{EO}} + \frac{dV_{CB}}{R_{CO}}$$

By definition, referring to fig. 3.1.1: -

$$dV_{BE} = dV_{IN} - dV_E \qquad \text{and} \qquad dV_{CB} = dV_{OUT} - dV_{IN}$$

Substitute these into the above: -

$$dI_C = \frac{dV_{IN}}{R_{EO}} - \frac{dV_E}{R_{EO}} + \frac{dV_{OUT}}{R_{CO}} - \frac{dV_{IN}}{R_{CO}}$$

also $$dI_C = -\frac{dV_{OUT}}{R_C} \qquad dV_E = R_E dI_E \quad \text{and} \quad dI_E \approx dI_C \quad \Rightarrow \quad dV_E \approx -\frac{R_E}{R_C} dV_{OUT}$$

Re-arrange so that, to a good approximation: $$-dV_{OUT}\left(\frac{1}{R_C}+\frac{1}{R_{CO}}+\frac{R_E}{R_{EO}R_C}\right)\approx dV_{IN}\left(\frac{1}{R_{EO}}-\frac{1}{R_{CO}}\right)$$

Now $\left(\frac{1}{R_C}+\frac{1}{R_{CO}}=\frac{1}{R_{CE}}\right)$ is the reciprocal of the effective collector resistance (conductance), hence the usual (small signal) model of a current source in parallel with the output resistance and collector resistance.

Also: $$R_{CO} >> R_{EO} \quad \Rightarrow \quad G=\frac{dV_{OUT}}{dV_{IN}}\approx -\left(\frac{R_{EO}}{R_{CE}}+\frac{R_E}{R_C}\right)^{-1}$$

Strictly speaking the gain of each side is negative but one normally refers only to the magnitude.

Maximum gain is achieved with zero emitter resistances: $$G_{MAX}\approx\frac{R_{CE}}{R_{EO}}$$

A good choice of collector voltage is about 10V. This permits maximum dynamic range whilst keeping V_{CB} > 5V. One can conclude, in many cases (power supply of ±15V): -

$$I_C R_C \approx 5V$$

From the second graph, fig. 2.1.2, the output conductance is proportional to the collector current according to: -

$$h_{OE}=\frac{1}{R_{CO}}\approx\frac{I_C}{100} \quad \Rightarrow \quad R_{CO}\approx 20R_C$$

Hence the rule of thumb (for calculating gain within 5%): $$R_{CE}\approx R_C \;\Rightarrow\; G\approx\frac{R_C}{R_{EO}+R_E}$$

3.1.1 Example calculation

A frequently chosen operating current is 0.1mA so that the typical current gain is 600 and calculations based on $I_E \approx I_C$ should be in error by less than 0.2%. The collector output resistance is 1MΩ. Suitable collector resistors are 47kΩ and the emitter output resistance is 250Ω.

The effective collector resistance is: $$R_{CE}=\frac{R_C R_{CO}}{R_C+R_{CO}}=44.9k\Omega$$

The maximum gain (zero emitter resistances): $$G_{MAX}=\frac{R_{CE}}{R_{EO}}=\frac{44.9k}{250}=179$$

For a gain of 100, employing the rule of thumb: $R_{EO}+R_E=470\Omega \;\Rightarrow\; R_E=220\Omega$ (an E24 value)

$$G\approx\frac{R_C}{R_{EO}+R_E}=\frac{47k}{470}=100$$

The more accurate formula is not much different: $$G=\left(\frac{250}{44.9k}+\frac{220}{47k}\right)^{-1}=97.6$$

N.B. As R_E increases the gain reduces. The emitter resistors provide a modicum of negative feedback and, as usual, this helps to reduce sensitivity to variations in transistor characteristics and improves linearity.

3.2 Differential input resistance

If one holds one input at a constant voltage and increases the other by a small amount the base current also increases. The ratio is the differential input resistance.

By definition: $$dV_{BE} = dV_{IN} - dV_E \quad \Rightarrow \quad dV_{IN} = dV_{BE} + dV_E$$

The input resistance is, therefore: $$R_{IN} = \frac{dV_{IN}}{dI_B} = \frac{dV_{BE}}{dI_B} + \frac{dV_E}{dI_B}$$

The first part, from section 2.2: $$V_{BE} = V_T \ln\left|\frac{I_B}{I_S}\right| \quad \Rightarrow \quad \frac{dV_{BE}}{dI_B} = \frac{V_T}{I_B} \approx h_{DC} R_{EO}$$

The second part, by Ohm's law: $$\frac{dV_E}{dI_B} = h_{FE} R_E$$

The DC current gain is approximately the same as the small signal current gain and the input resistance is, with sufficient accuracy in most cases: -

$$R_{IN} \approx h_{FE}\left(R_{EO} + R_E\right)$$

3.3 Common mode rejection (mismatched resistors)

Whereas data sheets often lack the necessary detail it is claimed that the matching of the transistors makes it possible to achieve a common mode rejection ratio (CMRR) of 120dB (10^6) or more, by employing very low tolerance/matched collector resistors (usually specified at 0.1%). This can be overkill in many applications and maximum CMRR can be achieved with 1% or even 5% resistors if one employs a current source with sufficiently high output resistance. Also, it is frequently necessary to deliberately unbalance the collector resistors in order to null the input offset voltage (see section 3.5)

If one increases the voltage at both inputs by *dV* the voltage, *V*, at the current source (see figs. 1 and 3.1.1) increases by approximately the same amount. Ideally the current does not change and the collector voltages remain the same. In practice the current source has a finite output conductance, equivalent to a large resistor, R_{OS}, in parallel with an ideal current source. The current will change by a small amount: -

$$dI = \frac{dV}{R_{OS}}$$

If one assumes that the BJTs are perfectly matched the change in current is shared equally. If the collector resistances are exactly the same the result is a change in the output common mode but not the difference. The resistors have a tolerance $\left(R_C \pm \Delta R_C\right)$, however, and, in the worst case (one resistor higher and the other lower than nominal), the difference between the outputs would change by: -

$$R_C \pm \Delta R_C \quad \Rightarrow \quad dV_{OUT} = 2\Delta R_C \frac{dI}{2} = \Delta R_C \frac{dV}{R_{OS}}$$

The differential gain is, approximately (section 3.1): $$G \approx \frac{R_C}{R_{EO} + R_E}$$

The CMRR is usually defined as the differential gain divided by the common mode gain: -

$$G \div \frac{dV_{OUT}}{dV} \approx \frac{R_C}{\Delta R_C} \times \frac{R_{OS}}{R_{EO} + R_E}$$

For maximum CMRR one can choose the maximum gain $(R_E = 0)$. Also, for a current source with suitably chosen BJT (high R_{CO}), the output resistance is typically (see sections 4.1 and 2.3 respectively): -

$$R_{OS} \approx \frac{3\times10^4}{I_C} \quad \text{and} \quad R_{EO} \approx \frac{25mV}{I_C} \quad \Rightarrow \quad CMRR \approx 1.2\times10^6 \times \frac{R_C}{\Delta R_C}$$

A CMRR of better than 10^6 (120dB) is thus easily achieved with 1% tolerance resistors.

3.4 Common mode rejection (mismatched transistors)

In practice the CMRR is limited by the matching of the BJTs. The effect of an infinitesimal change in collector voltage is negligible (second order) so that, to a very good approximation: -

$$dI_C = \frac{dV_{BE}}{R_{EO}} + \frac{dV_{CB}}{R_{CO}} \approx \frac{dV_{BE}}{R_{EO}}$$

The change in current is the same but the balance is altered slightly by the mismatch.

$$dI \ = \frac{dV_E}{R_{OS}} = dI_1 + dI_2 \quad \text{with} \quad dI_1 \approx \frac{dV_{BE}}{R_{EO1}} \quad \text{and} \quad dI_2 \approx \frac{dV_{BE}}{R_{EO2}}$$

The change in current is very nearly the same in each side and, therefore: -

$$dI \ = \frac{dV_E}{R_{OS}} \approx \frac{dV_{BE}}{2R_{EO}} \quad \Rightarrow \quad dV_{BE} \approx \frac{2R_{EO}}{R_{OS}} dV_E$$

The change in base-emitter voltage is very much less than the change in base voltage (the emitters "follow" the gate common mode). The change in the output is: -

$$dV_{OUT} = R_C(dI_1 - dI_2) = R_C\left(\frac{1}{R_{EO1}} - \frac{1}{R_{EO2}}\right)dV_{BE}$$

Which simplifies, with a little algebra: -

$$dV_{OUT} = 2\frac{\Delta R_{EO}}{R_{EO}}\frac{R_C}{R_{OS}}dV_E$$

The common mode gain is, therefore: $dV_E \approx dV_B \quad \Rightarrow \quad \frac{dV_{OUT}}{dV_G} \approx 2\frac{\Delta R_{EO}}{R_{EO}}\frac{R_C}{R_{OS}}$

The CMMR is usually defined as the differential gain divided by the common mode gain: -

$$G \div \frac{dV_{OUT}}{dV_G} \approx \frac{R_{CE}}{R_{EO}}\frac{R_{EO}}{2\Delta R_{EO}}\frac{R_{OS}}{R_C}$$

The effective collector resistance is usually not much lower than the collector resistor (component): -

$$R_{CE} \approx R_C \quad \Rightarrow \quad CMRR \approx \frac{R_{EO}}{2\Delta R_{EO}}\frac{R_{OS}}{R_{EO}}$$

Manufacturers do not specify the mismatch in the emitter output resistance characteristics but, fortunately, one can relate this to a parameter that is specified – the current gain, based on the fundamental properties (see section 2.2).

$$I_B = I_S \exp\left(\frac{V_{BE}}{V_T}\right) \quad \text{and} \quad I_C = h_{FE} I_B$$

From the analysis in section 2.3: $R_{EO} = \frac{V_T}{I_C}$

The thermal voltage $V_T = \frac{kT}{e} \approx 25mV$ (at 20ºC) is a fundamental property and one can be fairly confident that it matches accurately within a pair. The current gain, on the other hand, is known to vary considerably, depending on the manufacturing process (level of doping). The emitter output resistance is related to the DC current gain by: -

$$R_{EO} = \frac{V_T}{h_{FE} I_B}$$

Once again one can be confident that the base current matches reasonably well if the base-emitter voltage is the same and the tolerance is, therefore the simple relationship.

$$\frac{\Delta R_{EO}}{R_{EO}} = -\frac{\Delta h_{FE}}{h_{FE}}$$

We are usually interested in the magnitude only and so the CMRR is: -

$$CMRR \approx \frac{h_{FE}}{2\Delta h_{FE}} \frac{R_{OS}}{R_{EO}}$$

The data sheet for the low cost version of the LM394 has a typical mismatch of 1% and the SSM2210 data sheet quotes 0.5%. Both quote a maximum of 5%.

Parameter	Conditions	LM194			LM394			LM394B/394C			Units
		Min	Typ	Max	Min	Typ	Max	Min	Typ	Max	
Current Gain Match, (h_{FE} Match) $= \frac{100\ [\Delta I_B]\ [h_{FE(MIN)}]}{I_C}$	V_{CB} = 0V to V_{MAX} I_C = 10 μA to 1 mA I_C = 1 μA		0.5 1.0	2		0.5 1.0	4		1.0 2.0	5	% %

Once again the high output resistance of the BJT current source reduces the effect of transistor mismatch and a CMRR of 10^6 (120dB) is readily achieved.

3.5 Input offset voltage and temperature coefficient

The base-emitter voltage for an ideal PN junction is, according to theory: -

$$V_{BE} = V_T \ln\left|\frac{I_B}{I_S}\right| \quad \text{with} \quad V_T = \frac{kT}{e} \approx 25mV \text{ (at 20°C)}$$

For each transistor in a slightly mismatched pair, therefore: $V_{BE1} = V_T \ln\left|\frac{I_{C1}}{h_{FE1}I_{S1}}\right|$ and $V_{BE2} = V_T \ln\left|\frac{I_{C2}}{h_{FE2}I_{S2}}\right|$

The difference is: $\Delta V_{BE} = V_T \ln\left|\frac{I_{C1}h_{FE2}I_{S1}}{I_{C2}h_{FE1}I_{S2}}\right|$

The temperature coefficient is: $\frac{d(\Delta V_{BE})}{dT} = \frac{k}{e}\ln\left|\frac{I_{C1}h_{FE2}I_{S1}}{I_{C2}h_{FE1}I_{S2}}\right| = \frac{\Delta V_{BE}}{T}$

With both inputs connected to 0V and the emitters connected together the base-emitter voltages are identical. Any mismatch in the transistors results in an imbalance in collector currents and, if the collector resistors are identical, an offset at the output.

$$\Delta V_{BE} = 0 \quad \Rightarrow \quad \frac{I_{C1}h_{FE2}I_{S1}}{I_{C2}h_{FE1}I_{S2}} = 1 \quad \text{and} \quad \frac{d(\Delta V_{BE})}{dT} = 0 \quad \text{and} \quad \Delta V_{OUT} = (I_{C1} - I_{C2})R_C$$

The output offset can then be eliminated by either: -
a). Applying an offset to one of the inputs or
b). Adjusting the collector resistors: $\Delta V_{OUT} = I_{C1}R_{C1} - I_{C2}R_{C2} = 0V$
The latter preserves the condition for minimum temperature coefficient and is recommended. The mismatch in collector resistors reduces the common mode rejection ratio slightly but this is not usually a problem with a high output resistance BJT current source. This theory is not perfect but sufficiently accurate for most purposes.

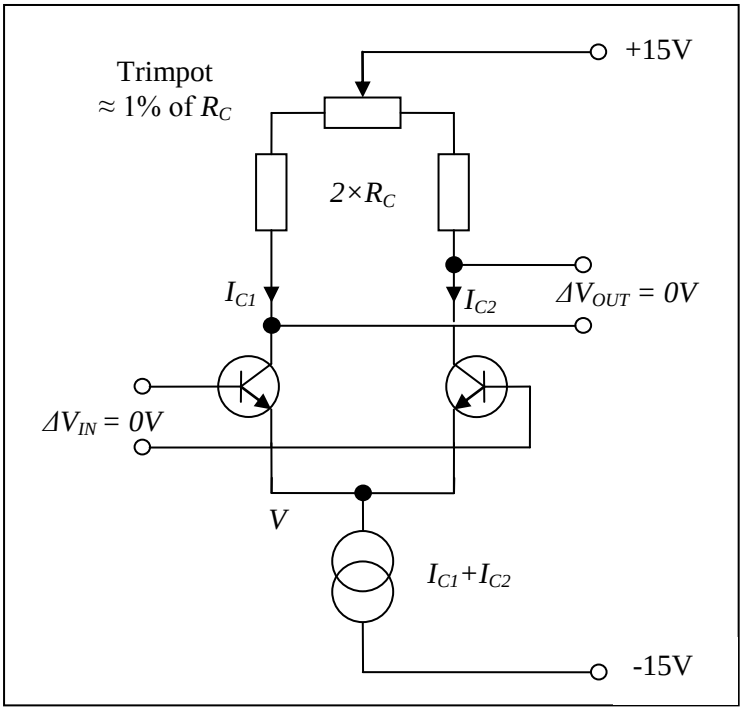


Fig. 3.5.1 Offset trim for retaining good temperature matching

Any residual temperature sensitivity, due either to the transistor mismatch or from the source, can be significantly reduced by a combination of offset trim methods. See the LM394 application note AN-222 for details.

4. Practical circuits

4.1 A basic current regulator

A very simple current regulator employs a low cost, high gain BJT with guaranteed minimum collector output resistance (e.g. BC109). The base is held at a constant voltage including a pn junction for temperature compensation for the transistor V_{BE} though some claim that two diodes or even a (red?) LED provides a better match. The 47nF capacitor provides low impedance at high frequency preventing any AC signal on V_C from getting through to the base via the collector-base capacitance (≈5pF).

An increase in V_C tends to increase I_C depending on the inherent collector output resistance of the transistor. The result is an increase in V_E and a reduction in the base current, counteracting the change. The (negative) feedback thus increases the effective output resistance of the transistor.

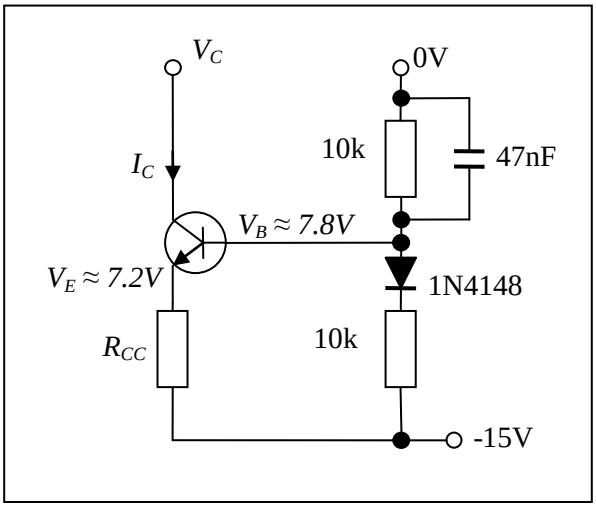


Fig. 4.1.1 Circuit model for the current regulator

From section 3.1: $$dI_C = \frac{dV_B}{R_{EO}} - \frac{dV_E}{R_{EO}} + \frac{dV_C}{R_{CO}} - \frac{dV_B}{R_{CO}}$$

$$dV_B = 0 \quad \text{and} \quad dV_E = R_{CC} dI_E \quad \Rightarrow \quad dI_C = -\frac{R_{CC}}{R_{EO}} dI_E + \frac{dV_C}{R_{CO}}$$

To a good approximation: $$dI_E \approx dI_C \quad \Rightarrow \quad R_{OS} = \frac{dV_C}{dI_C} \approx R_{CO}\left(1 + \frac{R_{CC}}{R_{EO}}\right)$$

The effective output resistance is, therefore, boosted by a large factor.

The base is normally held at about half the negative power supply so that the voltage across R_{CC} is about 7V. Also to a good approximation (rule of thumb) the emitter output resistance is inversely proportional to current: -

$$R_{CC} \approx \frac{7.2V}{I_E} \quad \text{and} \quad R_{EO} \approx \frac{25mV}{I_E} \quad \Rightarrow \quad R_{OS} \approx R_{CO}\left(1 + \frac{7.2V}{25mV}\right) \approx 290 R_{CO}$$

The output resistance of a suitably chosen BJT (e.g. BC109A) is: $R_{CO} \approx \frac{100k\Omega}{I_C(mA)}$ so that, very approximately: -

$$R_{OS} \approx \frac{3 \times 10^4}{I_C}$$

4.2 A fully differential pre-amp

A long tail pair, followed by a differential in/out amplifier, makes an excellent pre-amplifier for a resistance bridge null detector. Gain accuracy and stability, linearity and DC offsets are relatively unimportant – the most important thing is very low noise and a fully differential structure to avoid earth loop problems. The following was constructed as a stripboard prototype with an LM394 matched pair and some 0.1% 10k and 100Ω resistors scavenged from an old resistance box. A TL072 low cost dual op-amp is perfectly adequate for the differential amplifier (P channel JFET input stage so that in input range is up to the positive supply voltage). The back to back diodes protect the matched pair from excessive input signals. Local power supply decoupling of 100nF (ceramic) is usually sufficient if a very good quality regulated PSU is employed (recommended).

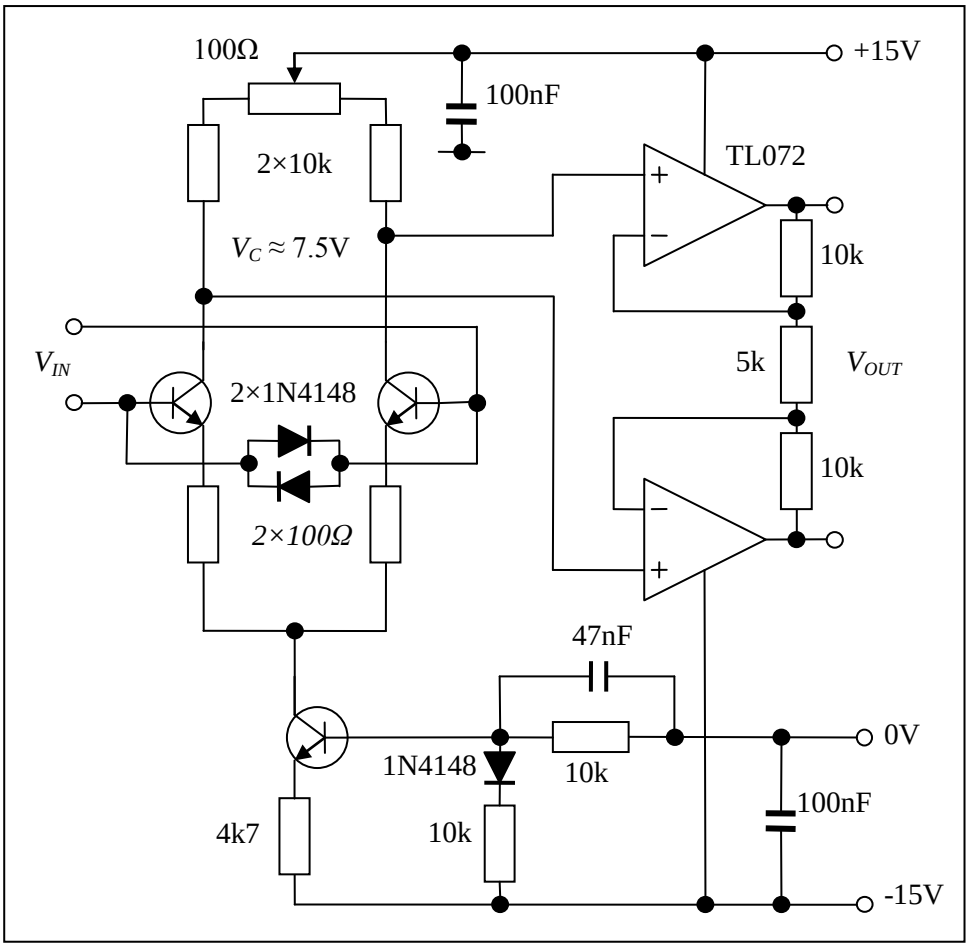


Fig 4.2.1 A basic long tail pair pre-amp and differential amplifier

The long tail pair has a gain of approximately 200 followed by the differential amplifier gain of 5 (overall ×1000). In practice the next stage is usually a band-pass filter with differential input stage and single-ended output stage, after which earth loops cease to be an issue. For test purposes a differential amplifier, based on a single op-amp with a gain of ten, would facilitate connection to a spectrum analyser.

The inputs should be connected by a screened twisted pair and the layout of the front end should be to minimise the area through which stray magnetic flux can flow (especially the loop consisting of V_{IN} and across the transistor bases and the loop consisting of the 100Ω resistors and emitters). It is probably not necessary to screen the whole amplifier but, if one does, then the screen (also the twisted pair) should be connected to local power supply 0V.

An operating current of 0.75mA was chosen so that the collectors are at approximately 7.5V and well within the input range of the op-amps, including the maximum AC signal. The outputs of the op-amps are also biased at 7.5V which limits the maximum signal but this is not a major problem. The upper limit for the TL072 is +13V and so the maximum output is 11V peak-peak (22V differential).

4.3 A pre-amp with feedback

If the gain needs to be accurate and stable it is recommended that feedback be applied to one side of the long tail pair. One of the inputs is labelled 0V (signal) and must be connected to the 0V (usually a virtual earth) side of the bridge output. To avoid an earth loop the following stage can employ a differential input (the difference between V_{OUT} and signal 0V). Better still is to avoid an earth loop completely by employing an impedance matching transformer. The signal 0V can then be connected to the local power supply 0V. The recommended collector resistors are chosen so that, at the operating current, the collectors are at approximately 10V and well within the input range of the LF356 (also a P channel JFET input stage). The action of feedback maintains the AC signal on the collectors zero (or very small) so that maximum signal handling is limited only by the range of the op-amp output (typically ±13V and only a little more than the previous example because it is single-ended).

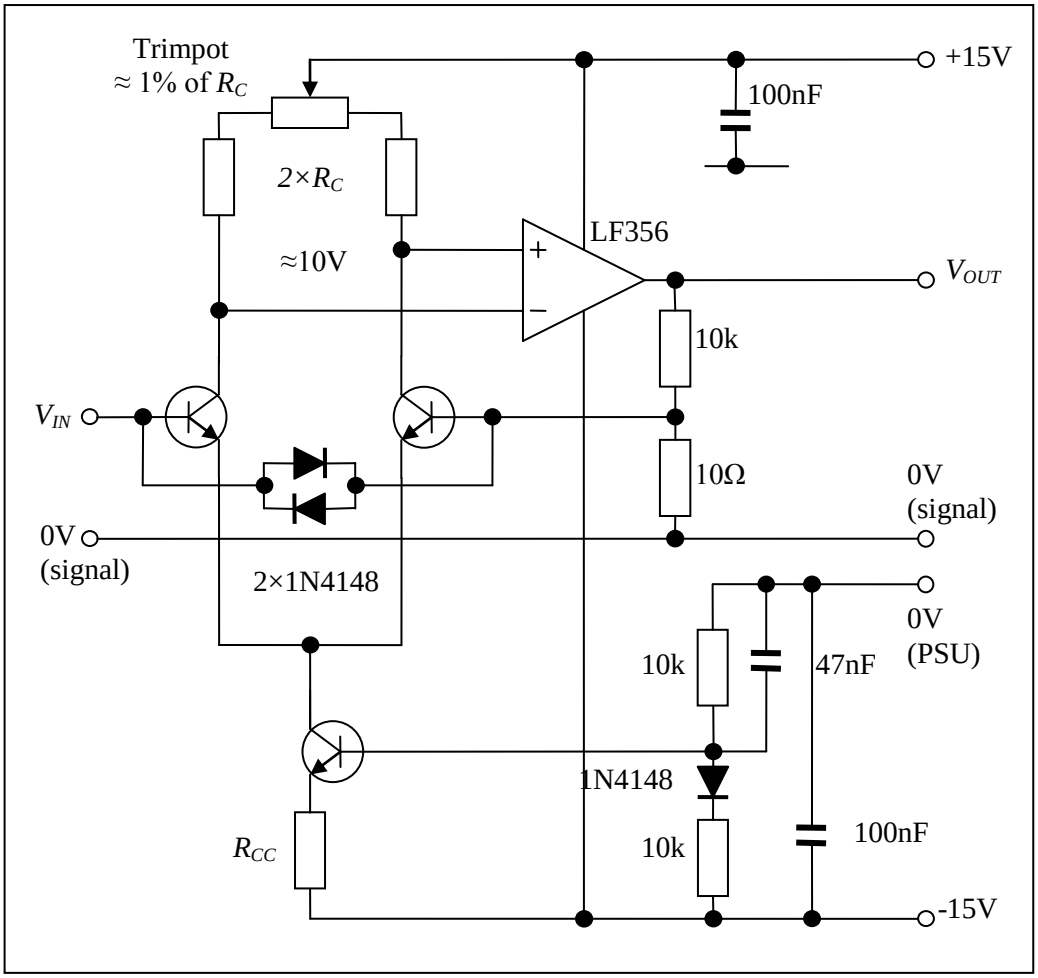


Fig. 4.1 A low noise pre-amplifier (gain = 1000)

With a closed loop gain of 1000 there is no problem with stability, though the bandwidth is reduced, compared to the unity gain frequency of the op-amp. The gain of the front end is, very approximately: -

$$R_C \approx \frac{5V}{I_C} \quad \text{and} \quad R_{EO} \approx \frac{25mV}{I_C} \quad \Rightarrow \quad G \approx \frac{R_C}{R_{EO}} \approx 200$$

This would tend to increase the bandwidth but the much lower feedback factor (10^{-3}) results in an overall reduction of the bandwidth to one fifth (0.2) of the op-amp (approx 1MHz with a LF356: GBWP = 5MHz). This is not a problem for most applications. The 10Ω in the feedback divider contributes negligible noise.

A large DC offset at the output can be a nuisance and so a DC offset trim is included. With a 470k resistor up to 320μV of mismatch can be corrected. Lowest temperature coefficient of DC offset is achieved when the operating currents of the two transistors (and collector voltages) are precisely the same. For the very best DC performance, therefore, it may be worth including an offset trim on the LF356 as well – for adjusting the output offset.

4.4 A low noise stage for an HAVF

With a little modification the previous example can be converted to a one-plus-integrator suitable for the first stage of a high accuracy voltage follower (e.g. for the active drive of a three-stage ratio transformer) [1 and 2].
The feedback resistor is replaced by a capacitor. The time constant required usually means quite a large capacitor (≈1-2μF with a 10Ω resistor). The load impedance is then too low for the op-amp. A simple solution is to increase the 10Ω resistor by a factor of 10 and reduce the capacitor accordingly. The Johnson noise of the 100Ω resistor is then comparable to the transistors but not critical. The resulting circuit works well at low frequency (sine wave - no high frequency signal present). The back to back diodes (2×1N4148) across the capacitor help restore stable operation after transient overload conditions (e.g. power-up) [2].

With a capacitor in the feedback circuit the feedback factor, at high frequency, increases to 100%. The closed loop gain falls to unity and so, for stability, the loop gain needs to be reduced with a snubber (470Ω in series with 330pF). See below for details.

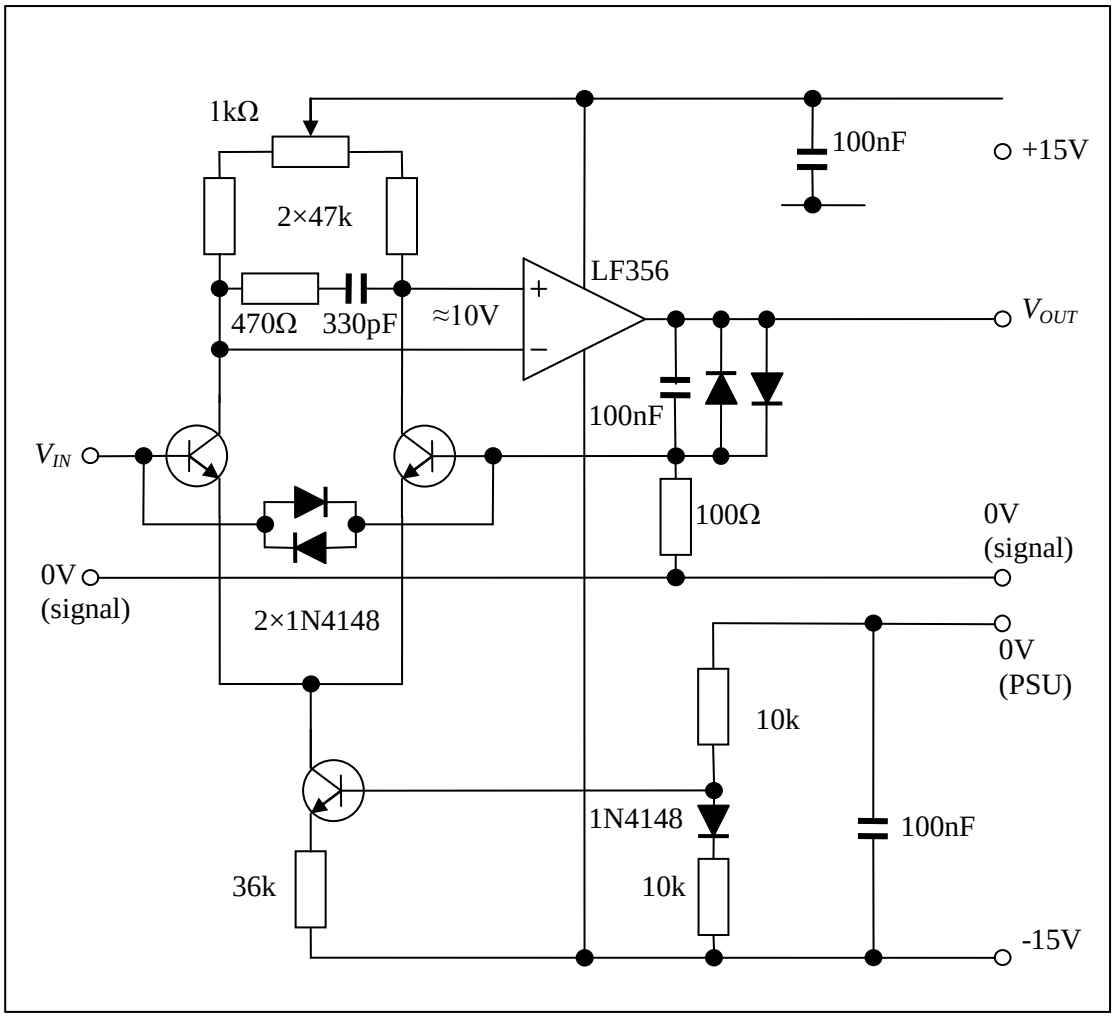


Fig. 4.4.1 The first (low noise) one-plus-integrator stage of a high accuracy voltage follower (F18 type).

With the F18 type ratio transformer the follower voltage noise contributes very little to the overall performance but the noise current needs to be low. Also, the very low resistance of the second energising winding requires excellent DC performance. Noise current and DC performance is improved with low operating current (lower self-heating) and an operating current of 0.1mA proves to be a good compromise. With the F17 type transformer, however, the follower noise is more important and DC less critical (higher winding resistances) and an operating current of 1mA is recommended. In both cases the DC offset adjustment is essential.

1. Part 4, monograph 1: “High gain blocks”
2. Part 4, monograph 2: “High accuracy voltage followers”

4.4.1 Stability and the snubber

The best way to reduce the open loop gain is a snubber across the collectors. At very low frequency the impedance of the snubber is very high and it has no effect. The gain of the long tail pair is approximately 180 (see section 3.1.). At a frequency corresponding to twice the collector resistance and the snubber capacitance the gain starts to fall. At the frequency corresponding to the snubber resistance and capacitance the impedance reaches its minimum and the gain levels out. To be on the safe side the snubber resistance is chosen to be 200 times smaller than the combined collector resistances and the gain levels out to just below unity. As a result the overall open loop gain, including the op-amp, falls below unity just below the frequency at which the op-amp gain is unity and the stability margin is about the same as the op-amp.

Pole frequency, at which the LTP gain starts to fall: $$f_1 = \frac{1}{2\pi 2R_C C_S} = 5.13kHz$$

Zero frequency, at which the LTP gain levels out: $$f_2 = \frac{1}{2\pi R_S C_S} = 1.03MHz$$

The overall open loop transfer function is, therefore, LTP gain with a pole and a zero followed by an integrator, representing an ideal op-amp: -

$$T(s) = G \times \frac{1+\tau_2 s}{1+\tau_1 s} \times \frac{1}{\tau_3 s}$$

$$\tau_1 = 2R_C C_S = 3.1\times10^{-5} s \qquad \tau_2 = R_S C_S = 1.55\times10^{-7} s$$

The time constant for the op-amp corresponds to the gain-bandwidth product (GBWP) also known as the unity gain frequency (5MHz for an LF356): -

$$\tau_3 = \frac{1}{2\pi f_{GBWP}} \approx 3.2\times10^{-8} s$$

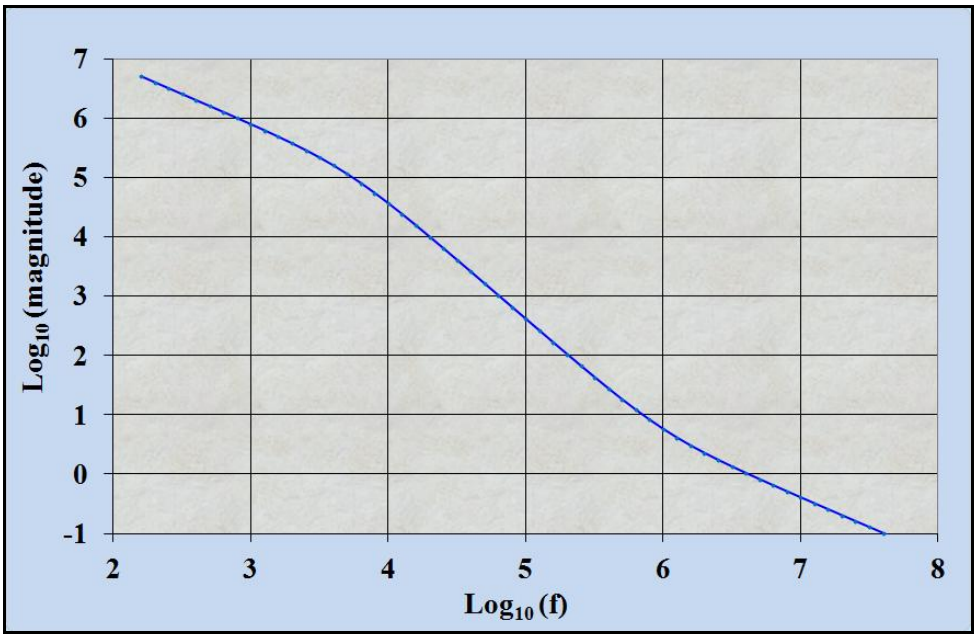


Fig. 4.4.1.1 Open loop gain (frequency in Hz)

The overall open loop gain initially falls at first order (one decade per decade) with a phase shift of 90 degrees. The pole at $5kHz$ increases the slope to second order until the zero frequency of $1MHz$ after which it reverts to first order. The phase shift when the gain passes through unity (0dB), with an ideal op-amp, is approximately 100 degrees, providing 80 degrees phase margin. In practice most op-amps have internal frequency compensation so that the phase margin is at least 45 degrees and so the overall phase margin is reduced by 10 degrees to at least 35 degrees. There is some peaking in the closed loop frequency response at the upper cut-off frequency but this should not be a problem if no high frequency signals are present.

4.5 A useful variant

A useful compromise, for a null detector pre-amp, is to apply feedback to one of the emitters. The 10kΩ feedback resistor and emitter output resistance determine the feedback factor and, therefore, the closed loop gain. The closed loop gain is not as accurate or stable as example 4.3 (the emitter output resistance is temperature sensitive) but the fully differential input stage avoids the problem of earth loops. The 10k resistor to 0V on the other emitter restores symmetry and balance.

The operating current is set precisely with a zener voltage reference and an op-amp applying negative feedback. The action of the feedback is to maintain the collector voltage of the first transistor at precisely 5.1V below the +15V supply. The current through the 5k1 collector resistor is, therefore, maintained at a stable 1mA. The diode in series with the emitter resistors avoids latch-up under transient conditions. The other op-amp also applies feedback to maintain the other collector at precisely the same voltage. Both feedback loops require a snubber to reduce loop gain at high frequency in order to ensure stability.

A low cost dual op-amp (TL072) is perfectly adequate for most applications.

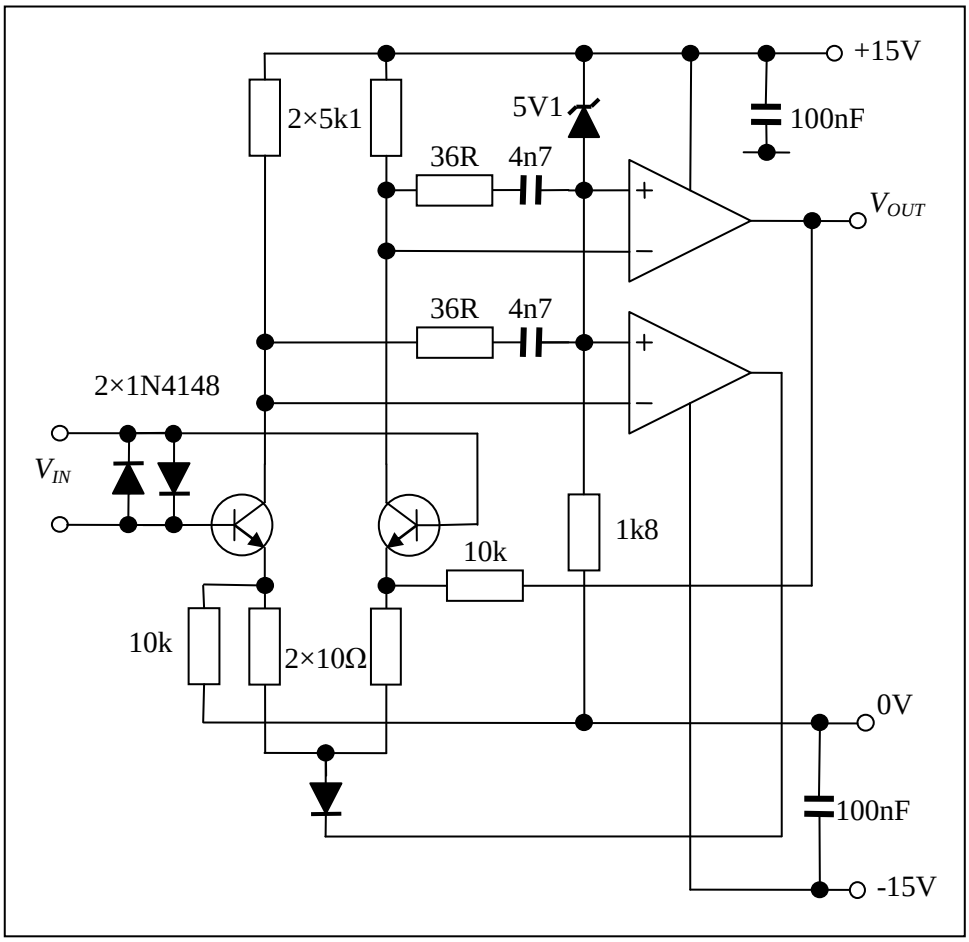


Fig. 4.5.1 A useful variant with fully differential input.

The gain is easily calculated, at least approximately. For the case above (1mA operating current): -

$$G = 1 + \frac{R_F}{R_{EO}} \quad \text{with} \quad R_{EO} \approx \frac{25mV}{I_C} \quad \Rightarrow \quad G \approx \frac{10k\Omega}{25\Omega} = 400$$

If a DC offset trim is required the same technique as above may be employed. A 100Ω collector trimpot will usually suffice.

Low noise JFET pre-amplifiers

1. Introduction

Junction field effect transistors (JFETs) have a higher noise voltage, compared to bipolar junction transistors (BJTs) [1] but much lower noise current. Their noise resistance is much higher and suited to applications where the source impedance is greater than about 100kΩ.

Whereas excellent integrated circuits are available (e.g. the Analogue Devices AD 645 op-amp and the Burr-Brown INA126 instrumentation amplifier) the absolute best noise performance is achieved with discrete transistors or proprietary matched pairs.

The most frequently used pre-amplifier circuit is the differential in/out long tail pair. The long tail pair can be used on its own, if accuracy and stability of the gain is not critical (e.g. as the pre-amp of a null detector) or as the first stage, inside the loop, of a composite amplifier/integrator, employing an operational amplifier with negative feedback for greater accuracy. In high accuracy applications JFETs are mostly employed at the front end of a charge amplifier, as part of a capacitance bridge, typically operating at 1.6kHz.

The choice of operating current depends mainly on the type of JFET. Generally speaking a higher current results in a lower noise voltage and higher noise current but not in simple inverse proportion as with BJTs. The best way to match source resistance is by selecting the right type of JFET (larger area devices have lower noise resistance). The best choice of operating current is at the upper end of the device capability as the gain is also higher and the noise voltage (usually the main issue) is lowest. One must be aware, however, of self heating and heat dissipation (good ventilation and a heat sink help) as noise current increases significantly with temperature.

DC performance is not particularly good, in terms of both initial matching and temperature stability. The best (and most expensive) proprietary matched pairs have a mismatch of 10mV or more and often require an offset trim.

Other parameters, on the other hand, are quite well matched, especially when the DC is trimmed (precisely matched operating current) and a common mode rejection ratio of over 100dB (10^5) is readily achieved.

An alternative to the long tail pair is a differential source follower. The main advantage is speed (no Miller capacitance) and a bandwidth of many MHz The main disadvantage is low gain (approximately ×1) and the need for a low noise amplifier stage to follow. There are, however, a number of applications where a source follower pre-amplifier could be a useful extra. In a multi-purpose instrument, for example, a wide range of source impedance can be catered for with a BJT front end and a source follower. The source follower is switched in for the higher impedance ranges.

An N-type JFET is basically a thin conduction channel of N-type semiconductor material embedded in a P-type substrate forming a high quality PN junction. At each end of the conduction channel connections are brought to the surface with N-type material (drain and source). An insulating depletion region is formed at the PN junction [2].

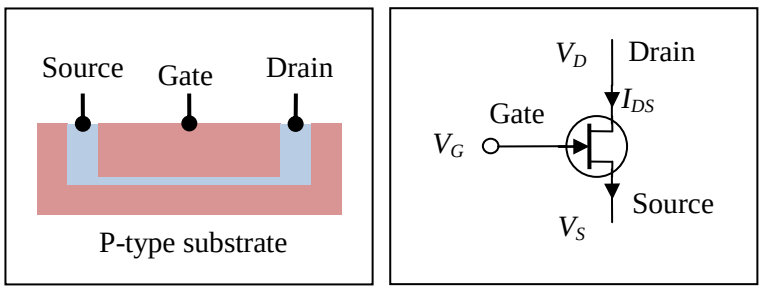


Fig. 1.1 A basic model for the JFET (N-type)

1. Part 5, monograph 2: “Low noise BJT pre-amplifiers.”
2. Part 5, monograph 4: “JFET Theory.”

A single connection is made to the substrate (the gate). The extent of charge carrier depletion and, therefore, the shape and thickness of the conduction channel depend on the relative voltages of all three connections. Whereas some devices are precisely symmetrical, with regard to drain and source, others are designed to minimise on-chip capacitance between the gate and drain. At low frequency the drain and source are usually interchangeable.

In most applications the PN junction is reverse biased and the current flowing in or out of the gate is negligible (a few pA at ambient temperature). The current flowing "down the drain" is almost exactly the same as that flowing "out of the source" (hence the silly names). The most frequently used mode of operation is the "pinched-off" mode where the drain-gate voltage is strongly reverse biased and the conduction channel at the drain end is extremely thin ("pinched off"). In this mode the drain-source current, I_{DS}, depends mainly on the gate-source voltage but it is also a function of the drain-gate voltage. The latter effect is small but not negligible. According to a fairly basic model, to a good approximation [1]: -

$$V_{GS} \geq V_P \quad \Rightarrow \quad I_{DS} = I_{DS}(V_{GS}, V_{DG}) \approx k(V_{GS} - V_P)^2 \qquad \text{with } k \text{ and } V_P \text{ approximately constant}$$

and

$$V_{GS} \leq V_P \quad \Rightarrow \quad I_{DS} = 0$$

Where: $V_{GS} = V_G - V_S$ and $V_{DG} = V_D - V_G$

Both V_{GS} and V_P are numerically negative. The small signal analysis is usually defined by: -

$$dI_{DS} = \frac{\partial I_{DS}}{\partial V_{GS}} dV_{GS} + \frac{\partial I_{DS}}{\partial V_{DG}} dV_{DG}$$

The first partial derivative defines the small signal forward transconductance and the second defines the small signal output conductance (in data sheets often referred to as g_{FS} and g_{OS} respectively): -

Define parameters: $$g_{FS} = \frac{\partial I_{DS}}{\partial V_{GS}} \qquad g_{OS} = \frac{\partial I_{DS}}{\partial V_{DG}} = \frac{1}{R_{DO}}$$

g_{FS} and g_{OS} are themselves functions of the drain-source current. R_{DO} defines the drain output resistance and is analogous to the collector output resistance as in the case of BJTs. The parameters k, V_P, g_{FS} and g_{OS} depend on the type of device and also vary from device to device of the same type. Fortunately there are some fundamental relationships between them which allow some useful design rules of thumb. The main one is that characteristics match very closely for devices of the same type when operated at precisely the same current [1]. One need not pay top dollar to achieve excellent performance with either an integrated pair or with two single transistors glued together.

An equally valid form of the equation, preferred by this author, employs the (numerically positive) variable source-gate voltage (V_{SG}).

$$V_{SG} \leq -V_P \quad \Rightarrow \quad I_{DS} \approx k(V_{SG} + V_P)^2 \quad \text{and} \quad V_{SG} \geq -V_P \quad \Rightarrow \quad I_{DS} = 0$$

Take the square root for the linear equation: $$\sqrt{I_{DS}} \approx \sqrt{k}(V_{SG} + V_P)$$

The other mode of operation is as a voltage controlled resistor (VCR). This has a potential benefit which is explored in section 4.

1. Part 5, monograph 4: "JFET Theory."

2. The long tail pair

The most frequently used JFET pre-amplifier is the differential "long tail pair" of matched low noise transistors operated in the pinched-off mode. The differential structure and constant current source ensure a high level of rejection of common mode voltage at the inputs and power supply variations. DC offset is not particularly accurate or stable, with respect to temperature, compared to a BJT matched pair or even a low cost op-amp. An input offset of 10mV is considered a good one.

The JFET high input impedance is easily damaged by electrostatic discharge and low leakage back-back diodes are recommended for protection. These are basically JFETs with drain and source connected together to form the cathode.

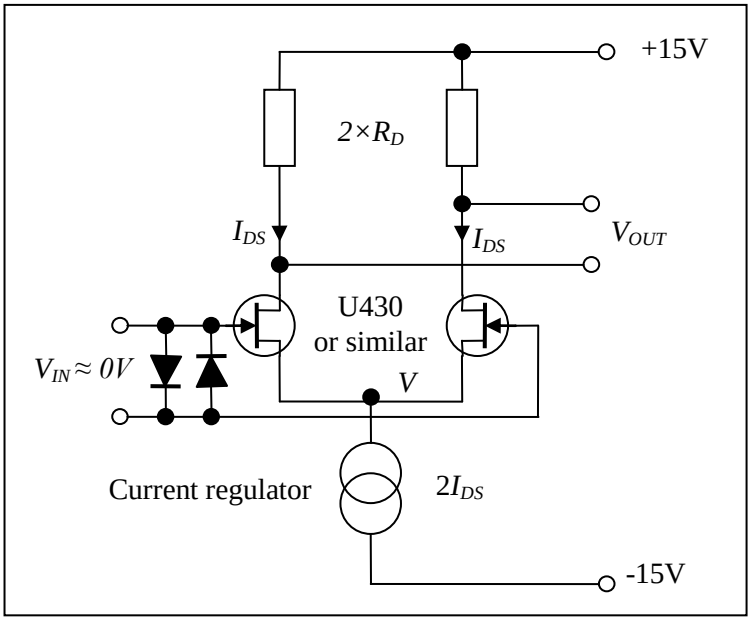


Fig. 2.1 Outline schematic of a long-tail pair.

As a rule of thumb the gain is, to a good approximation (see appendix section A1): -

$$G = \frac{dV_{OUT}}{dV_{IN}} \approx g_{FS} R_{DE} \quad \text{with} \quad R_{DE} = \frac{R_D R_{DO}}{R_D + R_{DO}} \quad \text{and} \quad R_{DO} = \frac{1}{g_{OS}}$$

g_{FS} is the forward transconductance of the JFET and R_{DE} is the effective drain resistance (the component resistance in parallel with the drain output resistance, R_{DO}).

The choice of operating current depends mainly on the type of JFET. Generally speaking a higher current results in lower noise voltage and higher noise current but not in simple inverse proportion as with BJTs. The best way to match source resistance is by selecting the right type of JFET (larger area devices have lower noise resistance). The best choice of operating current is usually a trade-off between voltage noise and stage gain. Whereas higher current results in higher forward transconductance the gain is lower – the drain resistors can be higher. See the monograph "JFET theory" for more detail [1].

Other parameters, on the other hand, are quite well matched, especially when the DC is trimmed (precisely matched operating current) and a common mode rejection ratio of over 100dB (10^5) is readily achieved.

Typical values for a dual matched JFET (U430) are: -

Current (mA)	$V_N\ (nV/\sqrt{Hz})$	$I_N\ (fA/\sqrt{Hz})$	P_N (W)	R_N (MΩ)	T_N (K)
1	4	≈1 (@25ºC)	4×10^{-24}	4.0	0.07

1. Part 5, monograph 4: "JFET Theory". See section 2.6.

3. A differential source follower

An alternative to the long tail pair is a differential source follower. The main advantage is speed (no Miller capacitance) and a bandwidth of many MHz The main disadvantage is low gain (approximately ×1) and the need for a low noise amplifier stage to follow. There are, however, a number of applications where a source follower pre-amplifier could be a suitable add-on extra. In a multi-purpose instrument, for example, with a low noise BJT front end, the source follower could be switched in for a high source impedance range.

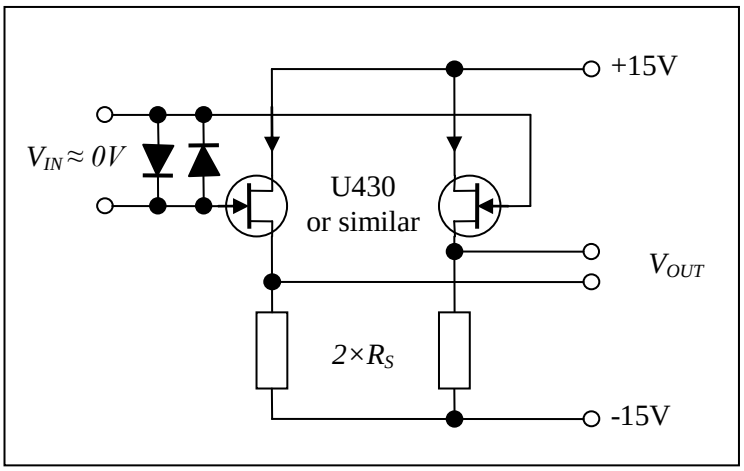


Fig. 3.1 A differential source follower

The input impedance is extremely high and the output resistance is moderately low – suitable for input to a BJT matched pair. According to Ohm's law and JFET theory [1], for each transistor: -

$$dV_S = R_S dI_{DS} \quad \text{and} \quad dI_{DS} = g_{FS} dV_{GS} + g_{OS} dV_{DG}$$

The change in drain-gate voltage has a very small affect on the gain: $g_{OS} \approx 0 \;\Rightarrow\; dI_{DS} \approx g_{FS} dV_{GS}$

With a little algebra, for each JFET: $dV_{GS} = dV_G - dV_S \;\Rightarrow\; G = \dfrac{dV_S}{dV_G} = \dfrac{g_{FS} R_S}{1 + g_{FS} R_S}$

The output resistance is also easy to calculate. If one holds the input constant and then takes a small bit of extra current from the source, the source voltage must fall according to: -

$$dI_{DS} \approx g_{FS} dV_{GS} = g_{FS}\left(dV_G - dV_S\right) = -g_{FS} dV_S \quad \Rightarrow \quad R_{SO} = -\frac{dV_S}{dI_{DS}} = \frac{1}{g_{FS}}$$

The result is typically between 100Ω and 1kΩ. With a following input capacitance of a few pF a bandwidth in excess of 100MHz is easily achieved.

A differential follower can also be used on its own, if accuracy and stability of the gain is not critical (e.g. as the pre-amp of a null detector). It can also be used as the first stage, inside the loop, of a composite op-amp, employing a low noise (BJT input) operational amplifier with negative feedback. A source follower would be particularly useful in the front end of a high speed charge amplifier, operating at a frequency of 160kHz or even higher, as part of a high speed capacitive transducer signal conditioner. Whereas such measurement systems are only moderately accurate (to 0.01% at best) they are often key subsystems in much higher accuracy measurement and control systems (e.g. nano-positioners and high accuracy optical instruments).

With a gain of only ×1 the stability of a feedback system is readily achieved.

1. Part 5, monograph 4: "JFET Theory."

4. An ultra-low input current pre-amplifier

The input (leakage) current of a JFET, in pinch-off mode, is largely dependent on the drain-gate voltage as well as temperature. Unfortunately there is very little information available on gate leakage in voltage controlled resistance (VCR) mode. A typical datasheet contains the following: -

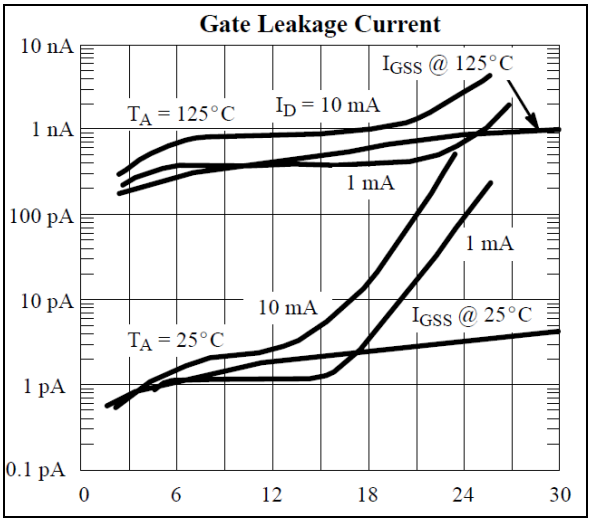


Fig. 4.1 Gate leakage current versus drain-gate voltage (J112 courtesy Siliconix)

It should be possible to operate a JFET pair in VCR mode with much reduced drain-gate and source-gate potential differences, resulting in very much reduced leakage current and, therefore, much lower current noise. The J112, for example, is a symmetrical device and, if the drain and source are maintained at equally positive and negative potentials, relative to the gate, the net leakage current should be extremely small. A reasonably good model is an ideal (zero leakage) JFET with two very large resistors (typically $>10^{12}\Omega$) with leakage currents of the order <100fA (10^{-13}A) through each. In practice this would require the transistor package to be extremely clean (degrease with fresh cotton wool and methylated spirits) with complete screening/guarding and a very good insulator (PTFE or sapphire) for the gate terminal post [1].

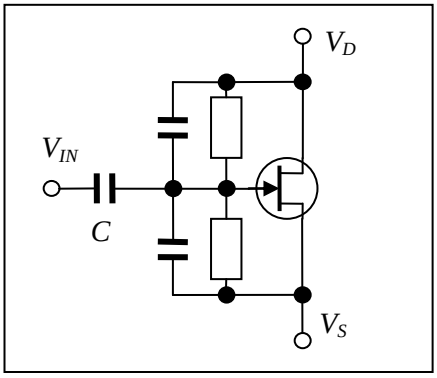


Fig. 4.2 A symmetrical JFET model

The resistors are temperature sensitive as well as non-linear, depending on the potential differences. It should be possible, however, to employ negative feedback to adjust the drain and gate DC potentials on a relatively long timescale, relative to the operating frequency, with guaranteed loop stability over a useful range of conditions.

It is also possible to arrange for the drain and source voltages to "follow", almost exactly, the gate AC potential thus "bootstrapping" the drain-gate and source-gate resistances and capacitances [2].

1. Yeager, John. & Hrusch-Tupta, Mary Anne et al: "Low Level Measurement", 5th edition.
Keithley Instruments Inc. For more practical tips on very high resistance and very low current measurements see sections 2 and 4.
2. Part 2, monograph 3: "An ultra-high input impedance high-pass filter"

A practical circuit could be a differential cascode stage with a matched (low noise) BJT pair. This is preferable to a long-tail pair (operated in the VCR mode) as the gain is much greater (see appendix A3). The resistors (R_1 and R_2) determine the drain-source operating voltage (typically 0.1V). The diode compensates for the base-emitter voltage and its variation with temperature. If the BJTs are perfectly matched the voltage developed across R_2 is then imposed across both JFETs. The value of resistor, R_3, is chosen so that the drain and source voltages are approximately equal and opposite relative to 0V (V_C = 0V). Fine adjustment is then provided by the control voltage.

The 100kΩ potentiometer allows for a slight mismatch between the JFETs (up to 15mV difference in pinch-off voltage). The input JFET must have the lower "on" resistance when tested with a meter (V_{GS} = 0V). The potentiometer is then set for zero output when the input gate is connected to 0V.

The input capacitors represent the ground capacitance, including the input capacitance of the JFET and a transducer capacitance/reference and/or the feedback capacitance if configured as a charge amplifier.

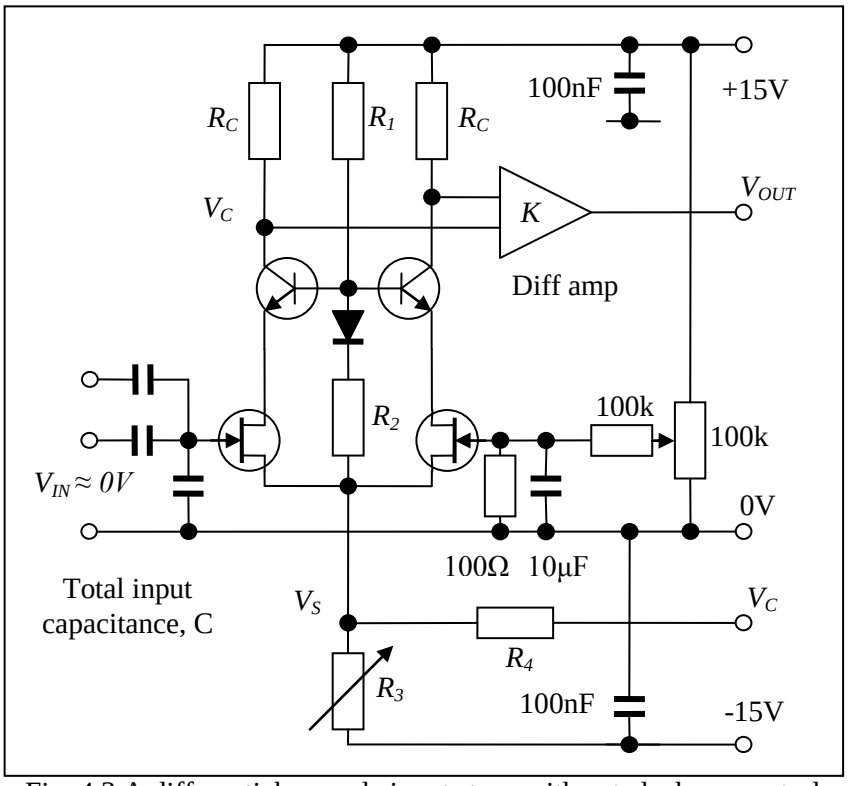


Fig. 4.3 A differential cascode input stage with gate leakage control

If the JFETs, BJTs and collector resistors are perfectly matched the output should not change, at least not instantaneously, when the control voltage is adjusted. If, however, there is a net leakage current the potential at the gate of the input JFET will gradually ramp up or down resulting in a slowly varying output. This "DC" component can then be fed back, with a suitable dynamic response, to automatically adjust the leakage current and maintain the gate voltage at 0V. The adjustment range of V_S is small and depends on the ratio of R_3 and R_4 (typically a control signal of ±10V is reduced to ±0.1V). A reasonably good model would be, in the (very low frequency) complex representation $(s = j\omega)$: -

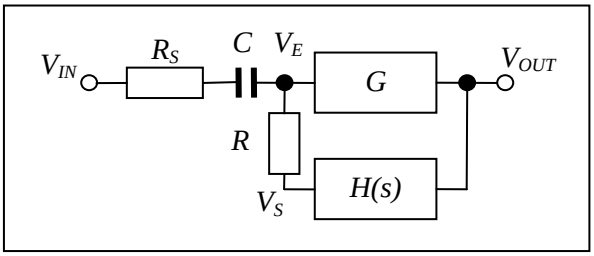


Fig. 4.4 The pre-amp feedback model

At the operating frequency the signal passes through the capacitor, is amplified, and appears at the output.

The most basic feedback network is an integrator, based on an op-amp $\left(H(s) = -\frac{1}{\tau_I s} \right)$. The differential amplifier must be connected the right way round to ensure that the feedback is negative.

The JFETs are, in effect, constant current sources (inputs constant) and the biasing resistors (R_1 and R_2) can be relatively large – the JFET source node is a high resistance point. The attenuation of the control signal is determined, therefore, almost entirely by resistors R_3 and R_4. The effective time constant of the integrator is increased so that with integrator components R_I and C_I : -

$$R_4 = 100R_3 \quad \Rightarrow \quad \tau_I \approx R_I C_I \frac{R_3 + R_4}{R_3} \approx 101 R_I C_I$$

For the model (see fig. 4.4) with negligible source resistance $(R_S = 0)$ and $\tau = RC$: -

$$H(s) = -\frac{1}{\tau_I s} \;\Rightarrow\; V_S = -\frac{G}{\tau_I s} V_E = -\frac{G}{\tau_I s}\left(V_{IN} + (V_S - V_{IN})\frac{1}{1+\tau s} \right)$$

$$\Rightarrow\; V_S\left(1 + \frac{G}{\tau_I s(1+\tau s)} \right) = -\frac{G}{\tau_I s}\left(\frac{\tau s}{1+\tau s} \right) V_{IN}$$

The transfer function, from V_{IN} to V_S is, therefore: $\dfrac{V_S}{V_{IN}} = \dfrac{-G\tau s}{G + \tau_I s + \tau_I \tau s^2}$

Divide top and bottom by G and the result is a second order band-pass characteristic: -

$$\frac{V_S}{V_{IN}} = \frac{-\tau s}{1 + \tau_I' s + \tau_I' \tau s^2} \quad \text{with} \quad \tau_I' = \frac{\tau_I}{G}$$

This time the effective time constant of the integrator is reduced - by the preceding gain. In standardised and normalised form $(s = j\omega/\omega_N)$: -

$$\frac{V_S}{V_{IN}} = \frac{-ks}{1 + 2\xi s + s^2}$$

The natural frequency is: $\omega = \dfrac{1}{\sqrt{\tau_I' \tau}}$ and damping ratio: $\xi = \dfrac{1}{2}\sqrt{\dfrac{\tau_I'}{\tau}}$

The constant k is related to the peak gain: $\dfrac{V_S}{V_{IN}}(s = j\omega_N) = \dfrac{-\tau}{\tau_I'} = \dfrac{-k}{2\xi}$

Loop stability depends on the damping ratio. A reasonable compromise is slight underdamping $(\xi \approx 0.5)$ so that the time constants need to be approximately the same.

For stable operation, therefore: $\tau_I' \approx \tau$

A large source resistance will also affect stability, limiting the range of applications. There is little point in detailed analysis as the leakage resistance is highly unpredictable. It will be necessary to experiment with a prototype.

The transfer function from V_{IN} to the main signal output is: -

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{H(s)}\frac{V_S}{V_{IN}} = \frac{\tau_I' \tau s^2}{1 + \tau_I' s + \tau_I' \tau s^2}$$

As expected this is a second order high-pass filter with the same natural frequency and damping ratio as above.

From the monograph "JFET theory" the current in VCR mode is [1. See section 2.5]: -

$$V_D + V_S = 0 \;\Rightarrow\; I_{DS} \approx \frac{V_{DS}}{R_0}\left\{1 - \frac{V_G}{V_P}\right\} \;\Rightarrow\; dV_C = R_C dI_{DS} \approx -V_{DS}\frac{R_C}{R_0}\frac{dV_G}{V_P}$$

By definition, when $V_G = 0V$: $$\frac{V_{DS}}{R_0} = I_{DS} \;\Rightarrow\; \frac{dV_C}{dV_G} \approx -\frac{R_C I_{DS}}{V_P}$$

The overall gain, including differential amplifier (gain = *K*) is, therefore: -

$$G = \frac{dV_{OUT}}{dV_G} \approx K\frac{R_C I_{DS}}{V_P}$$

4.1 Example calculation

The J112 has an "on" resistance of about 50Ω and a reasonable choice of operating current is 2mA each side. When stabilised this would result in drain and source voltages of ±50mV relative to the gate (0V). With a drain-gate and source gate resistances of the order $10^{12}\Omega$ the resulting leakage current is 50*f*A in each (at 20ºC). This is comparable with the best MOSFETs with the advantage of lower noise voltage (MOSFETs tend to be noisy). Fine adjustment, by the action of feedback, then reduces the net average current to near zero, limited only by the leakage current of the input capacitors.

The input capacitor could be constructed with PTFE or, even better, sapphire insulators with leakage currents at the limit of measurement (≈10^{-17}A or 10 atto-amps). According to theory the random fluctuation in each of the leakage currents (current noise) is predicted to be [1. See section 3]: -

$$\Rightarrow\; I_N = \sqrt{2eI_L B} \quad \text{and} \quad I_L \approx 50\,fA \;\Rightarrow\; I_N \approx 0.13\,fA/\sqrt{Hz}$$

Where $e = 1.60 \times 10^{-19} C$ is the quantum of electrical charge (i.e. on a proton) and *B* is the bandwidth in Hz.

A typical application would have an input capacitance of 10pF. This could be a very low leakage component capacitor or capacitance transducer. The input capacitance represents the total capacitance at the input, including ground capacitance and any feedback capacitance, as in the case of a charge amplifier. The resulting time constant is around 10s $\left(R \approx 10^{12}\Omega\right)$ and a natural frequency of 0.16Hz (high-pass cut-off). This should facilitate operation at half the power supply frequency (25 or 30Hz) or even lower.

The BJTs should be able to operate correctly with a collector-base voltage as low as 5V, allowing a DC voltage drop across the collector resistors of 10V and collector resistances of 5kΩ. With a pinch-off voltage of -2V the JFET gain is approximately ×5: -

$$\frac{R_C I_{DS}}{V_P} \approx 5$$

1. Part 5, monograph 4: "JFET theory".

The gain of the differential amplifier needs to be sufficient for the noise contribution of the following stage to be negligible. A convenient value is K = 20 so that the overall gain, including JFET stage, is ×100 and compensates for the effect of the attenuator on the feedback time constant. The actual integrator component values then determine the time constant.

The voltage noise contribution from the collector resistors is negligible compared to that generated by the noise current of the JFETs and the noise voltage added by the BJTs. According to theory the "Johnson" current noise and voltage noise generated by a resistor are [1. See section 2.4]: -

$$I_S = \sqrt{\frac{4kTB}{R_S}} \quad \text{and} \quad V_S = \sqrt{4kTR_SB}$$

Where $k = 1.38 \times 10^{-23}\, J/K$ = Boltzmann's constant, $T =$ is the absolute temperature ($T \approx 293K$ at 20ºC) and $B =$ Bandwidth in Hz. With a drain-source resistance of 50Ω the noise current is: -

$$I_S \approx 18\, pA/\sqrt{Hz} \quad \text{This flows through 5k}\Omega \quad \Rightarrow \quad V_{N1} \approx 91 nV/\sqrt{Hz} \text{ at the collectors.}$$

The noise generated by the collector resistor, on the other hand, is negligible: $V_{N2} \approx 9.1 nV/\sqrt{Hz}$

The noise of the BJT, referred to the input, at 2mA operating current, is of the order $1nV/\sqrt{Hz}$ (SSM2010). This is then amplified, by approximately the ratio of the collector to the emitter resistance (×100), also resulting in a noise component of around $V_{N3} \approx 100 nV/\sqrt{Hz}$. The three noise sources are statistically independent and combine as the root mean square (RMS) [2. See appendix A1]. Also, the same noise from the other side combines to increase the overall noise by a factor of $\sqrt{2}$

$$V_T = \sqrt{V_{N1}^2 + V_{N2}^2 + V_{N3}^2 + \ldots} \times \sqrt{2} \quad \Rightarrow \quad V_T \approx 189 nV/\sqrt{Hz}$$

The equivalent voltage noise, referred to the input (divide by the voltage gain), is very approximately $V_N \approx 40 nV/\sqrt{Hz}$ and significantly more than one would expect in the pinched-off mode (typically $5 nV/\sqrt{Hz}$).

As a ball-park estimate the noise resistance of the pre-amp is the ratio of the noise voltage and noise current: -

$$R_N = \frac{V_N}{I_N} \approx \frac{40 nV/\sqrt{Hz}}{0.13\, fA/\sqrt{Hz}} \approx 308 M\Omega$$

Suitable values for R_2 and R_1 are 1kΩ and 150kΩ respectively, consuming a modest extra 0.1mA of supply current. The noise voltage developed by R_2 and R_3 is common to both sides and is easily rejected by the differential amplifier. Suitable values for R_3 and R_4 are: -

$$R_3 = \frac{14.95V}{4.1mA} = 3.65k\Omega \quad \Rightarrow \quad R_4 = 365k\Omega$$

A fixed and variable resistor, connected in series, is probably the best solution for R_3.

Finally, given the unpredictability of the integrator time constant required, it could be argued that a more adaptive control loop, consisting of a microcontroller (with ADC and DAC), would be preferable. The algorithm could continuously analyse the response of the loop and self optimise accordingly. See, also, section 5.4 and [3].

1. Part 5, monograph 4: "Low noise BJT pre-amplifiers".
2. Part 5, monograph 1: "Null detectors – the basics".
3. Part 2, monograph 3: "An ultra-high input impedance high-pass filter"

5. Charge amplifiers

5.1 Op-amp based charge amplifiers

The main low noise JFET circuit of interest in high accuracy electronics is a charge amplifier, as part of a null detector in a capacitance bridge or signal conditioner [1], [2] and [3]. Excellent performance is possible even with a low cost JFET input op-amp. Numerous types are available with typical RMS input noise of $12nV/\sqrt{Hz}$.

The circuit is very simple – basically an inverting amplifier [4] with input capacitors (typically a variable and reference capacitor) and a feedback capacitor. The action of feedback is to maintain the inverting input at 0V ("virtual earth"). The ground capacitance usually consists of coax cable (connecting the transducer to the charge amplifier) plus stray and input capacitance of the op-amp and capacitance in the transducer. With zero signal voltage across no current flows through and its existence does not affect the measurement, apart from contributing to the noise gain. For best noise performance the feedback capacitor should be smaller than the total ground capacitance, subject to the requirement for closed loop bandwidth and negligible contribution due to noise current. A smaller feedback capacitor means higher gain but lower bandwidth and more phase error at the operating frequency. The feedback resistor provides a route for the DC leakage current and needs to be very large to ensure low phase error and negligible noise current (typically 1GΩ with 10pF or 100pF at 1.6kHz). The precise values of the feedback resistor and capacitor are not critical and DC at the output is not usually a problem. **Much more important are the provision of an overall screen and the avoidance of earth loops [5].**

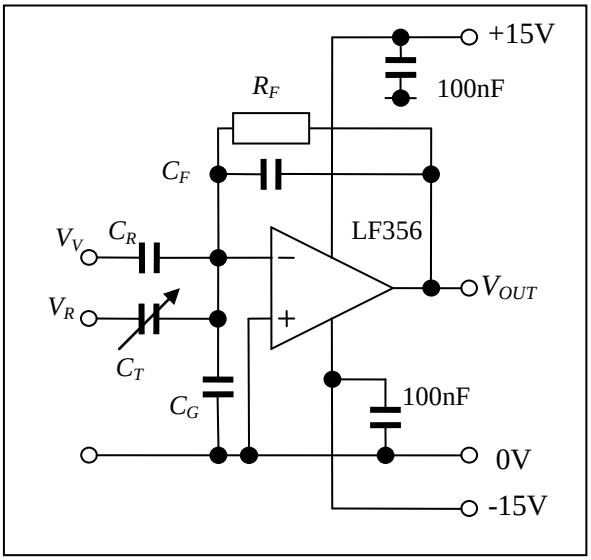


Fig. 5.1.1 An op-amp based charge amplifier

1. Part 1, monograph 3: "Rotary capacitive displacement transducer."
2. Part 1, monograph 4: "Linear capacitive displacement transducers."
3. Part 1, monograph 5: "Variable gap displacement transducers".
4. Part 4, monograph 1: "High gain blocks". See section 3.
5. Part1, monograph 2: "High accuracy single capacitors". See section 2.

Stability and noise performance depend on the size of the ground capacitance, C_G. The feedback factor, at the operating frequency, is determined by the feedback capacitor and the ground capacitance in parallel with the total input capacitance, $C_T + C_R$. The latter is usually negligible and the feedback factor is, to a good approximation: -

Feedback factor: $$F = \frac{C_F}{C_T + C_R + C_G} \approx \frac{C_F}{C_G}$$

The closed loop upper frequency response is reduced, compared to the open loop bandwidth of the op-amp by the feedback factor: -

Closed loop bandwidth: $$\omega_{CL} \approx F\omega_B = \frac{1}{\tau_{CL}}$$

Where ω_B is usually referred to as the unity gain frequency or gain-bandwidth product.

Typically the op-amp bandwidth is 5MHz (LF356). With a ground capacitance of 100pF (1m of coax cable) the feedback capacitor can be as low as 10pF ($F \approx 0.1$) in which case the closed loop bandwidth is 500kHz and the phase shift is negligible at an operating frequency of 1.6kHz.

Similarly, the feedback resistor also limits the lower frequency bandwidth. The overall result is a band-pass characteristic with widely separated poles. It is shown elsewhere [1], and in many elementary texts, that the band-pass characteristic, with respect to the V_R input, can be described with a transfer function of the form $(s = j\omega)$: -

$$T_R(s) = \frac{V_{OUT}}{V_R} = -\frac{Z_2}{Z_1} \times \left(\frac{H(s)}{H(s) + \beta(s)} \right)$$

Where Z_1 and Z_2 are the input and feedback impedances respectively. $H(s) = 1/\tau_B s$ is the open loop characteristic of the op-amp (an integrator) and $\beta(s)$ is the reciprocal of the feedback factor.

$$\Rightarrow \quad T_R(s) = -\frac{C_T}{C_F} \times \left(\frac{sR_F C_F}{1 + sR_F C_F} \right) \times \left(\frac{1}{1 + \tau_{CL} s} \right)$$

The net phase shift is approximately: $$\theta \approx \frac{1}{\omega R_F C_F} - \frac{\omega}{\omega_{CL}}$$

One can consider the op-amp noise voltage to be in series with either input. With zero input voltages the equivalent circuit is a non-inverting amplifier: -

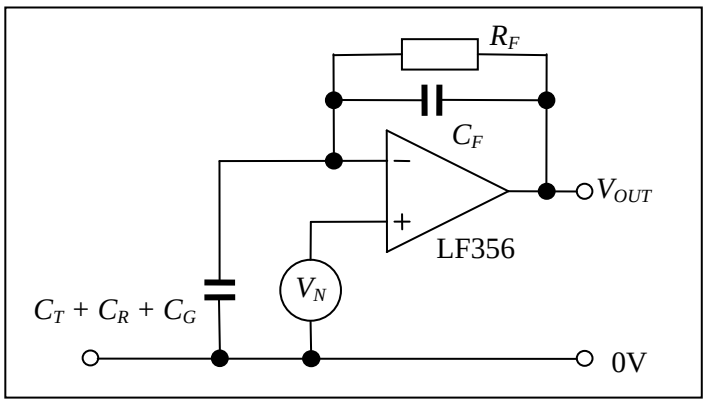


Fig. 5.1.2 Voltage noise model

1. Part 4, monograph 1: “High gain blocks”. See section 2.

The noise gain is, therefore: $$G_N = 1 + \frac{C_T + C_R + C_G}{C_F} \approx \frac{1}{F}$$

Note that the noise gain is independent of frequency.

For an LF356 the noise voltage is typically $12nV/\sqrt{Hz}$ (RMS) and, with a feedback factor of 0.1, the noise level at the output would be $132nV/\sqrt{Hz}$.

The current noise model consists of current sources due to the op-amp and the feedback resistor. The action of feedback (virtual earth) ensures that the net noise current flows only through the feedback capacitor.

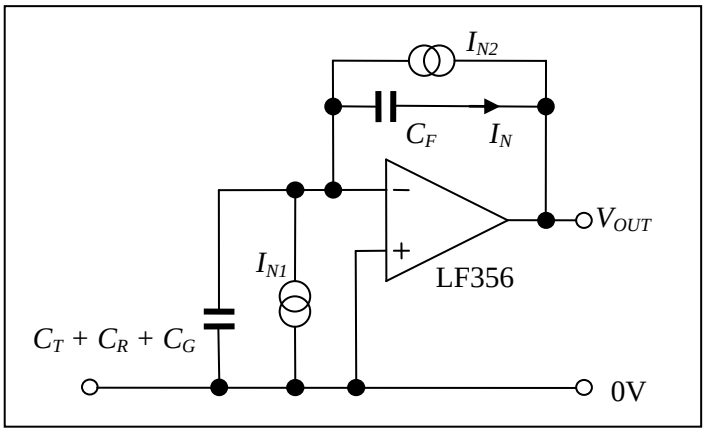


Fig. 5.1.3 Current noise model

The current sources are statistically independent so that the RMS net current flowing through C_F is: -

$$I_N = \sqrt{I_{N1}^2 + I_{N2}^2}$$

The noise at the output is, therefore: $$V_{OUT} = \frac{I_N}{\omega C_F} \quad \text{(RMS due to noise current)}$$

Note that the noise at the output now depends on the magnitude of the capacitive impedance and, therefore, frequency. Ideal noise matching occurs when the contributions from the noise voltage and noise current are the same: -

$$V_N\left(1 + \frac{C_T + C_R + C_G}{C_F}\right) = \frac{I_N}{\omega C_F} \quad \Rightarrow \quad \frac{V_N}{I_N} = \frac{1}{\omega\left(C_T + C_R + C_G + C_F\right)}$$

One could interpret this as noise resistance match to the magnitude of the total capacitive impedance connected to the virtual earth node. In practice the noise current of the feedback resistor dominates and it is not practical to trade off voltage noise against current noise.

For example: The noise current of a 1GΩ resistor at room temperature (300K) is $4\,fA/\sqrt{Hz}$ and higher than a suitable op-amp or matched pair (typically $< 1\,fA/\sqrt{Hz}$). The impedance of 100pF capacitance at 1.6kHz is 1MΩ and the contribution to noise at the output is only $40nV/\sqrt{Hz}$. This is less than the contribution due to noise voltage.

Rule of thumb: One may as well choose a low noise JFET op-amp with a noise resistance a little above the ideal.

5.2 A charge amplifier with long tail pair front end

The best noise performance is achieved by adding a matched pair at the front end of a reasonably low noise op-amp to make a composite op-amp. Typical values for a dual matched JFET (U430) are: -

Current (mA)	$V_N\left(nV/\sqrt{Hz}\right)$	$I_N\left(fA/\sqrt{Hz}\right)$	P_N (W)	R_N (MΩ)	T_N (K)
1	4	≈1 (@25ºC)	4×10^{-24}	4.0	0.07

The extra gain and phase shift of the long tail pair (LTP) can result in instability, depending on the feedback factor. A snubber (series resistor and capacitor) across the drains may be necessary.

A DC offset trim is included which can correct for up to 15mV of offset. The 100Ω resistor may contribute a small amount to the noise and a large capacitor (e.g. Tantalum) is added to reduce the AC impedance to 10Ω at 1.6kHz. Unlike BJT offset trim it is better to keep the drain resistors accurately matched so that, with zero differential output, the pair are operating at precisely the same current and their characteristics are, therefore, better matched.

Both inputs are maintained at close to 0V so common mode rejection is not an issue and a BJT constant current source may be a bit of overkill. The balanced structure and current source do, however, also provide a high level of power supply rejection. With a low noise power supply a JFET based current regulator or even a resistor would suffice. Otherwise see the monograph "Low noise BJT pre-amplifiers" for more detail.

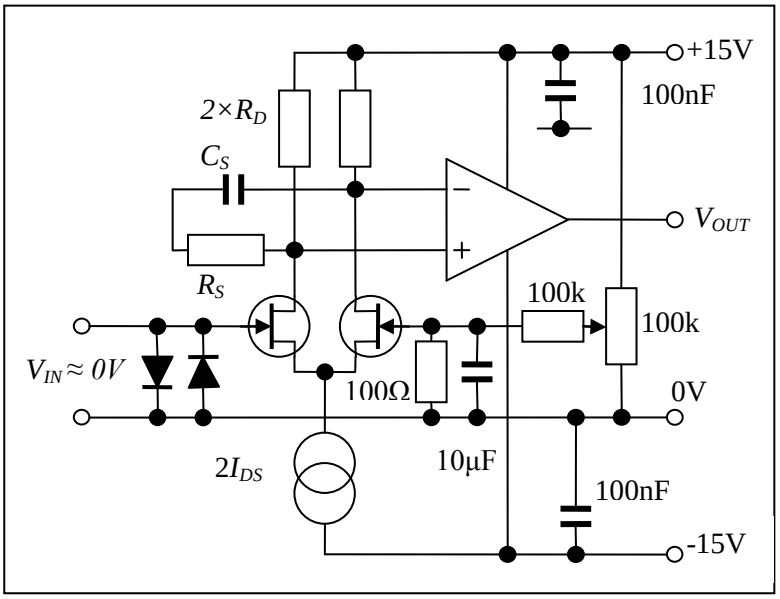


Fig. 5.2.1 A composite op-amp with a long tail pair

5.2.1 Example calculation

The author found, in a batch of ten, seven single JFETs (low cost, low resistance types J112) with excellent noise characteristics (typically $5nV/\sqrt{Hz}$). Two of those had a closely matching pinch-off voltage (10mV difference) and very closely matching gain characteristic. The models used in text books and application notes are: -

$$I_{DS} = I_{DSS}\left(1-\frac{V_{GS}}{V_P}\right)^2 = k\left(V_{GS}-V_P\right)^2 \quad \Rightarrow \quad k = \frac{I_{DSS}}{V_P^2}$$

The parameter *k* is the most useful but, unfortunately, it is not provided in data sheets. The values obtained by linear regression were 7.89×10^{-3} and 7.83×10^{-3} respectively (dimensions: amps/volt^2) – a match of about 1%. Similarly the values of pinch-off voltage were calculated by extrapolation to zero drain-source current.

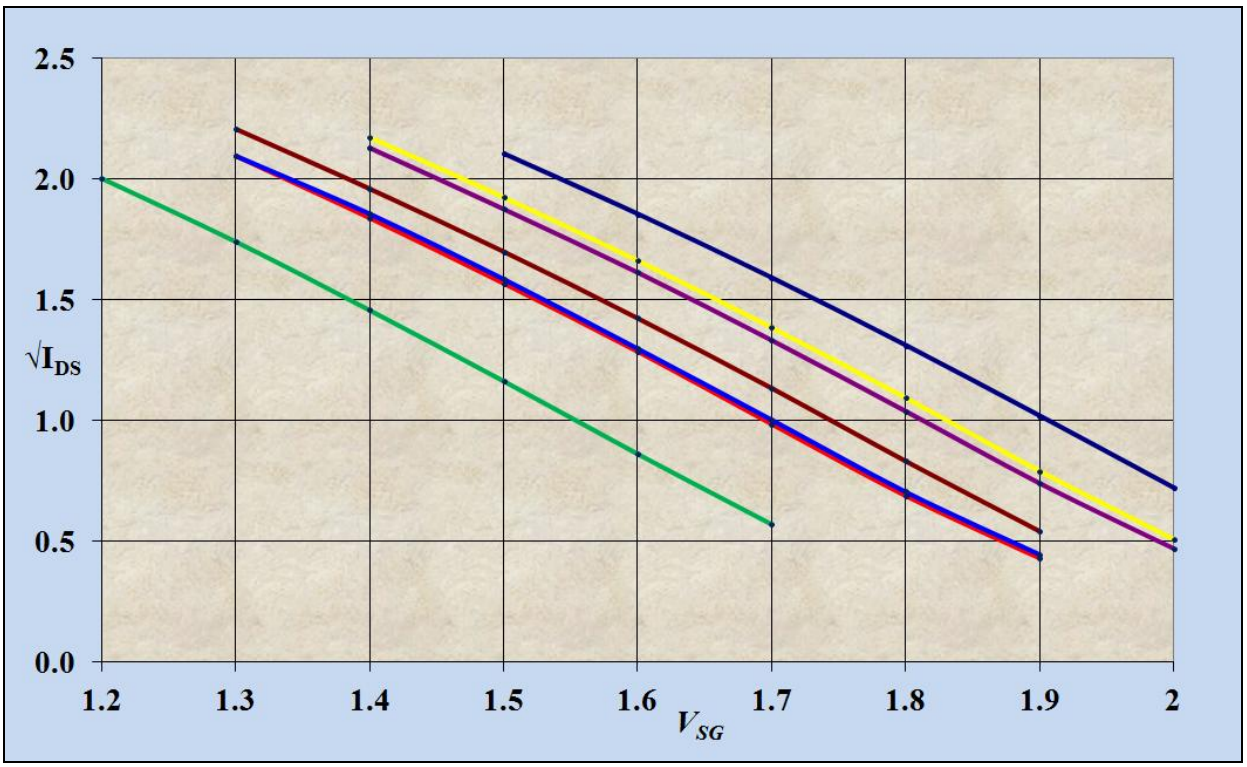


Fig. 5.2.1.1 A batch of J112 JFETs

The gate-source voltage is: $$V_{GS} = \sqrt{\frac{I_{DS}}{k}} + V_P$$

The forward transconductance is: $$\frac{\partial I_{DS}}{\partial V_{GS}} = 2k\left(V_{GS} - V_P\right) = 2\sqrt{kI_{DS}}$$

For low noise the recommended operating current is about 1mA (not critical). The average gate-source voltage required is, therefore: -

$$V_{GS} = \sqrt{\frac{1}{7.86}} - 2.056 = -1.70V$$

The gates are operating at about 0V so that the source voltage is +1.7V.

The average forward transconductance is: $$g_{FS} = 2\sqrt{7.86 \times 10^{-6}} = 5.61mS$$

(This is consistent with the data sheet that states a typical value of 6mS)

To operate well within the pinched-off mode the drain-gate voltage needs to be at least twice the pinch-off voltage. Five volts should suffice so one can use a pair of 10kΩ drain resistors (0.1% tolerance, scavenged from an old decade box). The data sheet is not very helpful with drain output resistance (specified as 25µS at 1mA but a drain-gate voltage of 20V). The effective drain resistance (component resistance in parallel with the drain output resistance) is, therefore, 20% lower than the actual resistance: -

$$R_{DE} = \frac{R_{DO}R_D}{R_{DO} + R_D} = \frac{40k \times 10k}{40k + 10k} = 8k\Omega$$

The gain is, therefore, approximately: $$G \approx g_{FS}R_{DE} = 5.61mS \times 8k \approx 45$$

A typical application requires, for example, a ground capacitance in the range 100 – 500pF (extra long cable) and, to retain a reasonably high frequency response, the feedback capacitor is chosen to be 100pF (1% polystyrene, scavenged from an old capacitive decade box). The feedback factor can be as high as 0.5 (100pF feedback into 100pF ground capacitance) and one needs to reduce the gain, at high frequency, inside the feedback loop by an extra factor of 22.5 in order to retain stability (feedback factor times the extra gain). To be on the safe side I reduced the gain by 25 and the snubber resistor needs to be a factor of 12.5 lower than the drain resistors (think of two snubber resistors in series with the centre connected to +15V). The result is 640Ω and, to be on the safe side again (it's not critical), I used a 620Ω as I had plenty in stock.

The main aim of the snubber is to reduce the gain of the long tail pair (LTP) to unity well before the op-amp gain passes through 0dB. The LF356 has lots of phase margin to spare and experiment (and analysis) shows that the LTP gain must start to level out to unity at a frequency of one fourth the op-amp unity gain frequency (5MHz).

Zero frequency, at which the long tail pair (LTP) gain levels out: $$f_2 = \frac{1}{2\pi R_S C_S} \approx 1MHz$$

The snubber capacitance needs to be, therefore: $$C_S = \frac{1}{2\pi R_S f_2} \approx 260\,pF$$

Once again this is not critical and a component value of 270pF should be satisfactory.

Pole frequency, at which the LTP gain starts to fall: $$f_1 = \frac{1}{2\pi 2R_D C_S} \approx 30kHz$$

This is well above the most frequently used operating frequency (1.6kHz) and the extra open loop gain at low frequency helps with accuracy and low phase error. The resulting charge amplifier is, in fact, usable up to 16kHz.

Clearly there is scope for experimentation if one uses a different op-amp (e.g. a much higher gain-bandwidth product) and requires more or less stability margin. For more detail and another example on snubber design and stability see the monograph "Low noise BJT pre-amplifiers" [1].

For higher frequency of operation (e.g. 1MHz) the op-amp can be replaced by an uncompensated RF amplifier (e.g. the ubiquitous μA733). To ensure stability the snubber resistor is made much smaller or eliminated altogether so that the LTP provides the dominant pole. Lower drain resistors and higher operating current ensure bandwidth up to. Say, 50MHz. The size of the capacitor depends on the feedback factor and some experimentation with stability margin and layout may be necessary. With modern superfast (internally compensated) op-amps, however, there is another way…

1. Part 5, monograph 2: "Low noise BJT pre-amplifiers". See section 4.4.1.

5.3 A charge amplifier with source followers.

For the ultimate high speed a charge amplifier can be implemented with a differential source follower and low noise high speed op-amp. The source followers do not add extra gain inside the loop and loop stability is simply an issue of feedback factor. The CLC426 (230MHz unity gain and $1.6nV/\sqrt{Hz}$), for example, has internal compensation for stable operation with a feedback factor up to 1.

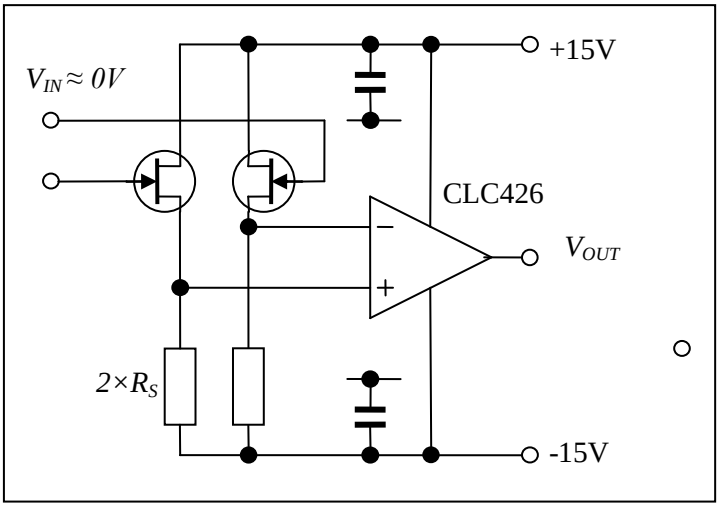


Fig. 5.3.1 A composite op-amp with source followers

Such a design is not for the faint heart and considerable RF knowledge and understanding is required (plus some really good test kit, especially the oscilloscope). Also, this is off at a tangent and the reader, if really interested, should contact the author for further advice.

5.4 An ultra-low leakage charge amplifier

With an adaptive (microcontroller controlled) ultra-low input current pre-amp (see section 4) it should be possible to construct a very sensitive charge amplifier. The basic idea is to allow the leakage current control loop to stabilise, with the input disconnected, before switching to measurement mode. In measurement mode the control loop is disabled, holding the zero leakage condition, and the input is connected via ultra-low leakage (glass encapsulated) reed relays. The reeds need to be well screened from the activating coils to prevent charge injection.

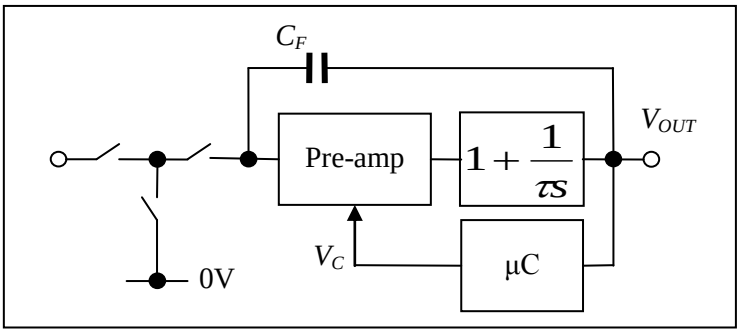


Fig. 5.4.1 An ultra-low leakage charge amplifier

The differential amplifier in the pre-amp can be given a first order low-pass characteristic. Extra gain in the loop is then provided by a one-plus integrator. This has the same time constant as the diff-amp so that the combined open loop gain is that of an integrator, equivalent to an op-amp. A second one-plus-integrator may be added, equivalent to a two-stage high gain block, for higher accuracy [1].

With such a charge amplifier it should be possible to measure average currents as low as a few atto-amps (< 100 electrons/s). If you are interested in building and testing a prototype please contact the author.

1. Part 4, monograph 1: “High gain blocks”.

6. Transducers incorporating a charge amp or single JFET

For the ultimate noise performance the charge amp (or single JFET), feedback capacitor and resistor can be incorporated within the transducer (or physically close) to minimise ground capacitance [1]. Noise current becomes an issue and the feedback resistor needs to be very large (typically > 1GΩ). Care must be taken to keep the op-amp or JFET cool as leakage current doubles for every 11 degrees. Current noise increases in proportion to $\sqrt{I_L}$.

A surface mount style op-amp takes up little space and presents no interfacing problems as the output impedance is very low.

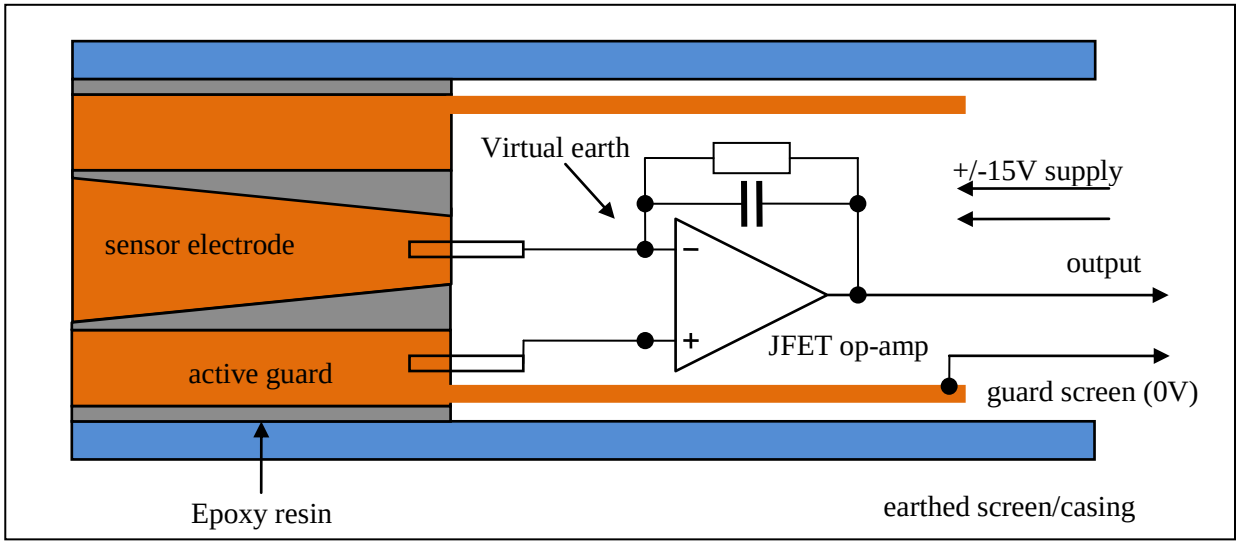


Fig. 6.1 A variable gap transducer with built-in charge amplifier [1]

The drain output of a single JFET in pinched-off mode, on the other hand, presents much higher impedance and an active guard (driven coax screen) is recommended.

The best solution is to operate the JFETs in voltage controlled resistor (VCR) mode with leakage current control. This mode eliminates the need for a high value resistor. The on resistance is typically 50Ω and interference ceases to be a problem. Such a design, however, would require a very low noise stage to follow and the cascode circuit is recommended. See section 4.

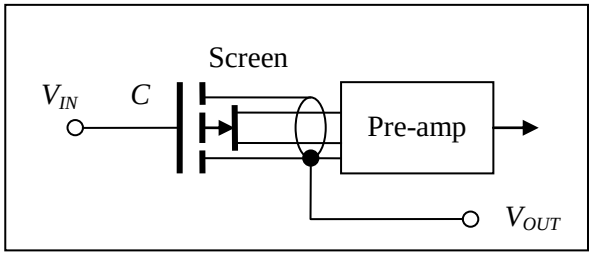


Fig.6.2 A single JFET in VCR mode in a special high-pass filter [2]

In a null-balance bridge system it is also possible to eliminate the feedback capacitor (i.e. not a charge amplifier). The bridge can still be balanced to a null though the gain of the pre-amp stage is less well defined due to the stray capacitance unless the screen is actively driven. Once again, the special high-pass filter circuit is recommended [2].

For further advice please contact the author.

1. Part 1, monograph 5: “Variable gap capacitive displacement transducers”. See, for example, section 2.1.
2. Part 2, monograph 3: “An ultra-high input impedance high-pass filter”.

Appendix: Long tail pair theory

A1. Voltage gain

The analysis is simplified if one takes advantage of the symmetry of the circuit. Consider equal and opposite infinitesimal changes to the inputs. The increase in current down one side is matched by a decrease in the other and the source voltages remain constant. The gain must be the same if one of the inputs is held constant (e.g. in a charge amplifier one input is connected to 0V).

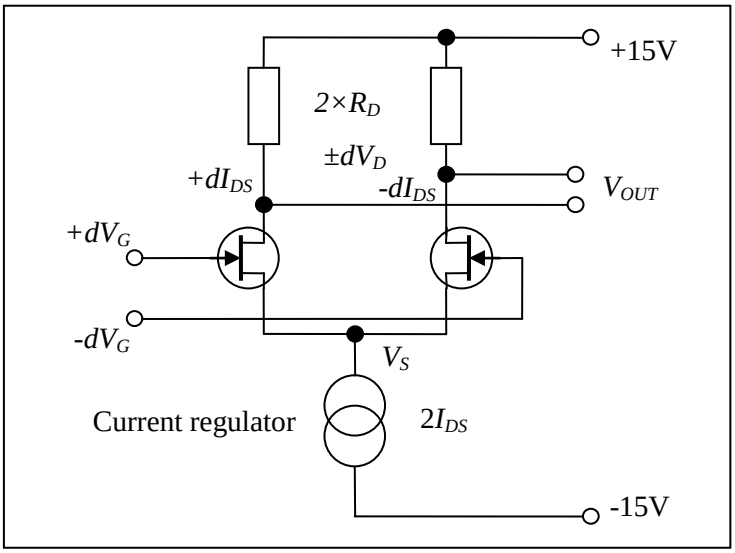


Fig. A1.1 Model for gain calculation

By definition, in the pinched-off mode: -

$$dI_{DS} = g_{FS} dV_{GS} + \frac{dV_{DG}}{R_{DO}} \quad \text{with} \quad dV_S = 0 \;\Rightarrow\; dV_{GS} = dV_G - dV_S = dV_G \quad \text{and} \quad dV_{DG} = dV_D - dV_G$$

g_{FS} = Forward transconductance and R_{DO} = drain output resistance (reciprocal of output conductance)

$$\Rightarrow \quad dI_{DS} = g_{FS} dV_G + \frac{dV_D - dV_G}{R_{DO}}$$

By Ohm's law: $\quad dV_D = -R_D dI_{DS} \quad \Rightarrow \quad dI_{DS} = g_{FS} dV_G - \dfrac{R_D dI_{DS} + dV_G}{R_{DO}}$

With a little algebra: -

$$\Rightarrow \quad \left(1 + \frac{R_D}{R_{DO}}\right) dI_{DS} = g_{FS} dV_G - \frac{dV_G}{R_{DO}} \quad \Rightarrow \quad dI_{DS} = \frac{g_{FS} R_{DO} dV_G - dV_G}{R_{DO} + R_D}$$

The voltage gain is: $\quad \dfrac{dV_{OUT}}{dV_{IN}} = \dfrac{dV_D}{dV_G} = -\dfrac{R_D dI_{DS}}{dV_G} = -\dfrac{g_{FS} R_{DO} R_D - R_D}{R_{DO} + R_D} = -g_{FS} R_{DE} + \dfrac{R_D}{R_{DO} + R_D}$

$$R_{DE} = \frac{R_{DO} R_D}{R_{DO} + R_D} = \text{Effective drain resistance } (R_D \text{ in parallel with } R_{DO})$$

Usually: $\quad g_{FS} R_{DE} >> 1 \quad \text{and} \quad \dfrac{R_D}{R_{DO} + R_D} << 1 \;\Rightarrow\; G \approx g_{FS} R_{DE}$

A2. Common mode rejection ratio

A2.1 Mismatched drain resistors

If one increases the voltage at both inputs $\left(dV_G\right)$ and the total current remains constant then the voltage, V_S, at the current source increases by the same amount. In practice the current source has a finite output conductance, equivalent to a large resistor, R_{OS}, in parallel with an ideal current source. The current will change by a small amount and there is a very small difference between dV_G and dV_S: -

$$dI = \frac{dV_S}{R_{OS}} \quad \text{and} \quad dV_S \approx dV_G \quad \Rightarrow \quad dI \approx \frac{dV_G}{R_{OS}}$$

If one assumes that the JFET pair are perfectly matched then the change in current is shared equally. If the resistances are exactly the same the result is a change in the output common mode but not the difference. The resistors have a tolerance $\left(R_D \pm \Delta R_D\right)$, however, and, in the worst case (one resistor higher and the other lower than nominal), the difference between the outputs would change by: -

$$R_D \pm \Delta R_D \quad \Rightarrow \quad dV_{OUT} = 2\Delta R_D \frac{dI}{2} = \Delta R_D \frac{dV_G}{R_{OS}}$$

The differential gain is, approximately (see previous section): $G \approx g_{FS} R_{DE}$

The CMRR is usually defined as the differential gain divided by the common mode gain. To a good approximation, therefore: -

$$G \div \frac{dV_{OUT}}{dV_G} \approx g_{FS} R_{DE} \times \frac{R_{OS}}{\Delta R_D}$$

The effective drain resistance is usually not much lower than the drain resistor (component) and, therefore: -

$$R_{DE} \approx R_D \quad \Rightarrow \quad CMRR \approx g_{FS} R_{OS} \times \frac{R_D}{\Delta R_D}$$

At a relatively high operating current (1 – 10mA) the forward transconductance is at least 1 - 10 mmho (milli-Siemens) and a typical output resistance of a BJT current source is at least $10^7\Omega$. The common mode rejection ratio is, therefore, a minimum 10^6 (120dB) with 1% tolerance resistors.

A2.2 Mismatched JFETs

In practice the CMRR is limited by the matching of the JFETs. The effect of an infinitesimal change in drain voltage is negligible (second order) and to a very good approximation, therefore: -

$$dI_{DS} = g_{FS} dV_{GS} + \frac{dV_{DG}}{R_{DO}} \approx g_{FS} dV_{GS}$$

The change in current is the same but the balance is altered slightly by the mismatch in JFETs.

$$dI = \frac{dV_S}{R_{OS}} = dI_1 + dI_2 \quad \text{with} \quad dI_1 \approx g_{FS1} dV_{GS} \quad \text{and} \quad dI_2 \approx g_{FS2} dV_{GS}$$

The changes in current are very nearly the same: -

$$dI_1 \approx dI_2 \quad \text{and} \quad dI = \frac{dV_S}{R_{OS}} \quad \Rightarrow dV_{GS} \approx \frac{dV_S}{2 g_{FS} R_{OS}}$$

The change in gate-source voltage is very much less than the change in gate voltage (the source voltage "follows" the gate common mode). The change in the output is: -

$$dV_{OUT} = R_D(dI_1 - dI_2) = R_D \Delta g_{FS} dV_{GS} \approx \frac{R_D \Delta g_{FS}}{2 g_{FS} R_{OS}} dV_S$$

The common mode gain is, therefore: $dV_S \approx dV_G \quad \Rightarrow \quad \frac{dV_{OUT}}{dV_G} \approx \frac{R_D \Delta g_{FS}}{2 g_{FS} R_{OS}}$

The CMMR is usually defined as the differential gain divided by the common mode gain: -

$$G \div \frac{dV_{OUT}}{dV_G} \approx g_{FS} R_{DE} \times \frac{2 g_{FS} R_{OS}}{\Delta g_{FS} R_D}$$

The effective drain resistance is usually not much lower than the drain resistor (component): -

$$R_{DE} \approx R_D \quad \Rightarrow \quad CMRR \approx 2 g_{FS} R_{OS} \times \frac{g_{FS}}{\Delta g_{FS}}$$

As before the product $g_{FS} R_{OS} > 10^4$ and a mismatch of no more than 1 or 2% is possible so that a CMRR of 10^6 (120dB) is readily obtained.

With charge amplifiers both inputs are firmly anchored at local PSU 0V and a high CMRR is usually overkill – a JFET current regulator or resistor to supply the current is usually sufficient.

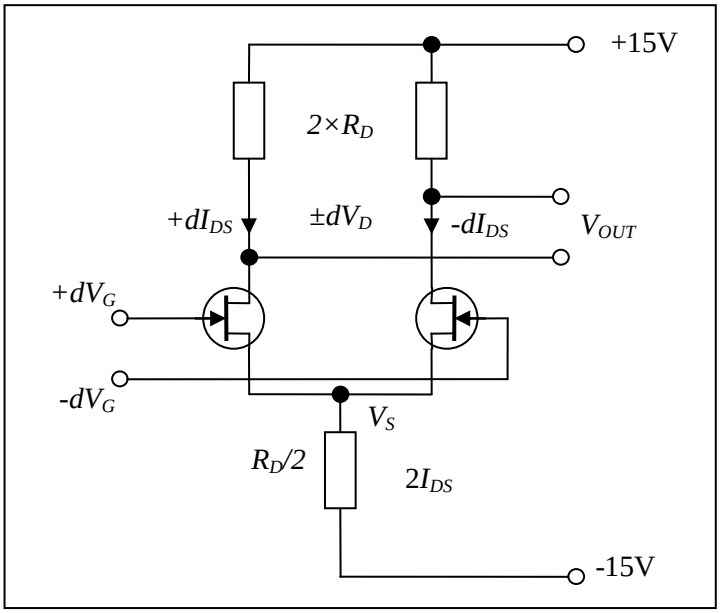
The matched pair (U430) datasheet quotes the following, re: matching: -

Matching								
Differential Gate-Source Voltage	$\lvert V_{GS1}-V_{GS2}\rvert$	V_{DG} = 10 V, I_D = 10 mA	25					mV
Saturation Drain Current Ratio[d]	$\frac{I_{DSS1}}{I_{DSS2}}$	V_{DS} = 10 V, V_{GS} = 0 V	0.95	0.9	1	0.9	1	
Transconductance Ratio[d]	$\frac{g_{fs1}}{g_{fs2}}$	V_{DS} = 10 V, I_D = 10 mA, f = 1 kHz	0.95	0.9	1	0.9	1	
Gate-Source Cutoff Voltage Ratio[d]	$\frac{V_{GS(off)1}}{V_{GS(off)2}}$	V_{DS} = 10 V, I_D = 1 nA	0.95	0.9	1	0.9	1	
Differential Gate Current	$\lvert I_{G1}-I_{G2}\rvert$	V_{DG} = 10 V, I_D = 5 mA	–2					pA
Common Mode Rejection Ratio	CMRR	V_{DG} = 5 to 10 V, I_D = 10 mA	75					dB

A3. The long-tail pair in VCR mode

With suitable drain resistors and operating current it is possible to operate in the voltage controlled resistor mode. One possible application is an ultra-low leakage current amplifier (see section 4), with drain and source voltages equal and opposite, relative to the inputs (0V). The main difference, compared to the differential cascode recommended there, is a much lower voltage gain, requiring the following stage to have a very low voltage noise and low noise resistance (i.e. a noise matching transformer [1]).

Here again one can take advantage of the circuit symmetry. Consider equal and opposite changes at the inputs.



According to theory the drain source resistance is [2]: -

$$V_D + V_S = 0V \;\Rightarrow\; R_{DS} = R_0\left(\frac{V_P}{V_P - V_G}\right) \;\Rightarrow\; dR_{DS} = R_0 V_P \frac{dV_G}{(V_P - V_G)^2} = \frac{R_{DS}}{V_P - V_G} dV_G$$

At $V_G = 0$, therefore: $\Rightarrow\; dR_{DS} = \frac{R_0}{V_P} dV_G$ (N.B. V_P is numerically negative)

The current remains approximately constant, being determined mainly by the power supply voltage and the drain resistance.

$$R_D >> R_{DS} \text{ and } V_D \approx V_S \approx 0V \;\Rightarrow\; I_{DS} \approx \frac{V_+}{R_D}$$

The change in drain voltage is determined, therefore, almost entirely due to the change in drain-source resistance: -

$$G = \frac{dV_D}{dV_G} \approx \frac{I_{DS} dR_{DS}}{dV_G} = \frac{I_{DS} R_0}{V_P}$$

With a typical R_{DS} resistance of 50Ω and pinch-off voltage of -2V and an operating current of 2mA (c.f. example calculation, section 4.1) the gain is: -

$$G \approx \times 0.05$$

1. Part 3, monograph 5: “Noise matching transformers”
2. Part 5, monograph 4: “JFET theory”. See section 2.5

A circuit for measuring tanδ

1. Introduction

The ideal capacitor has purely negative imaginary impedance. In practice, however, even high quality polymer dielectric capacitors have impedance with a real part, which can be modelled as a small series resistance, which varies with frequency [1]: -

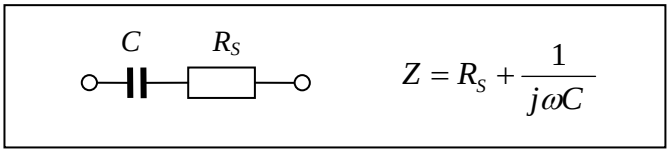


Fig. 1.1 Equivalent circuit for a practical capacitor

The ratio of the real part to the imaginary part is known as the "tanδ": the tangent of the acute angle to the negative imaginary axis of the complex number which represents the impedance. The angle is very small so that, to a very good approximation (in radians): -

$$\tan\delta = \omega R_S C \approx \delta$$

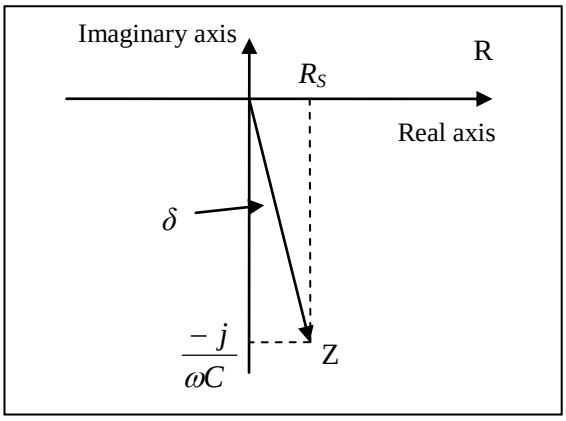


Fig. 1.2 The complex representation of a capacitor (real part exaggerated)

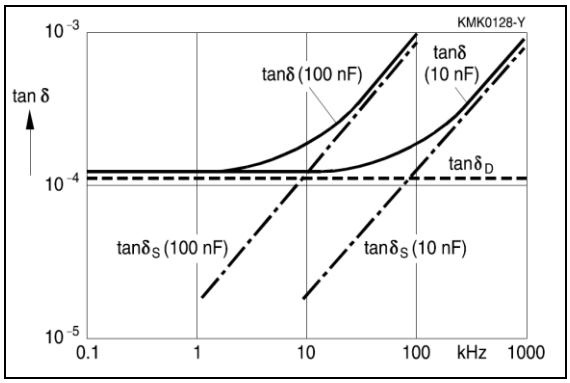


Fig. 1.3 Tanδ for polypropylene dielectric capacitors (courtesy EPCOS AG 2015 [2])

At low frequency the angle is constant and one can employ the concept of a complex capacitance [1]: -

$$\text{constant } \delta = \omega R_S C \;\Rightarrow\; R_S = \frac{\delta}{\omega C} \;\Rightarrow\; Z = \frac{1+j\delta}{j\omega C} \;\Rightarrow\; C \to \frac{C}{(1+j\delta)}$$

1. Part 1, monograph 6: "Low phase error capacitors and inductors".
2. Film Capacitors: General technical information, Sept 2015. See section 2.3.2.

1.1 The general complex representation

For this analysis I shall use the general complex representation of a sinusoid ($s = \sigma + j\omega$) where the real part represents either an exponentially growing (σ positive) or exponentially decaying (σ negative) sinusoid: -

$$V_{OUT} = V_P e^{st} \quad \text{with} \quad s = \sigma + j\omega$$

$$\Rightarrow \quad V_{OUT} = V_P e^{\sigma t} e^{j\omega t}$$

The time constant τ_D of the decaying sine wave is related to σ . This is easily seen if one expresses the decaying sinusoid in the usual form, with a numerically positive time constant: -

$$V_{OUT} = V_P e^{\sigma t} e^{j\omega t} = V_P \exp\left(-\frac{t}{\tau_D}\right)\exp(j\omega t)$$

$$\Rightarrow \quad \sigma = -\frac{1}{\tau_D}$$

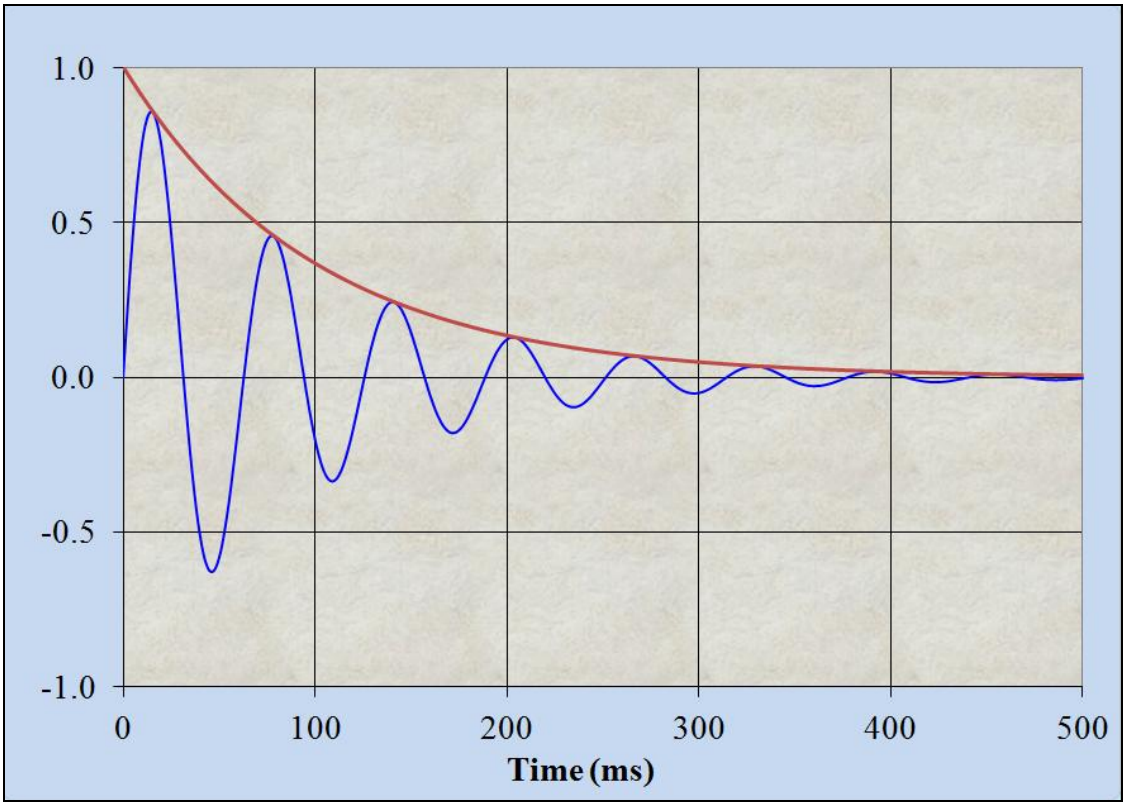


Fig. 1.1.1 An exponentially decaying sine wave (i.e. the real part)

$$V = \exp(-10t)\sin(100t)$$

$$\tau_D = 0.1s \quad \text{and} \quad \omega = 100\,\text{rads}^{-1}$$

2. A practical circuit

With good quality polymer capacitors the tan(δ) is so small that it is difficult to measure with conventional circuits. One solution is to construct a pair of integrators and an inverter with high gain blocks (HGBs) and provide positive feedback so that it is borderline stable.

The open loop gain of each HGB is so high that the closed loop phase error is negligible, compared to the effect of tan(δ) [1 and 2].

With capacitor C_3 set to zero it is easy to get the circuit oscillating with a short impulse current into one of the virtual earths. The oscillation dies away slowly. The time constant of the exponential decay is related to the average tan(δ) of the integrator capacitors. Alternatively, the addition of a small adjustable capacitor, C_3, adds a small phase lag to the inverter. This is increased until the circuit just bursts into oscillation. This condition represents a total phase shift around the loop of precisely 360 degrees and a gain (magnitude) of one at the frequency of oscillation (the "Barkhausen criterion" for oscillators [3]). The average tan(δ) can then be calculated from the resonant frequency, C_3 and R_4 as the following demonstrates.

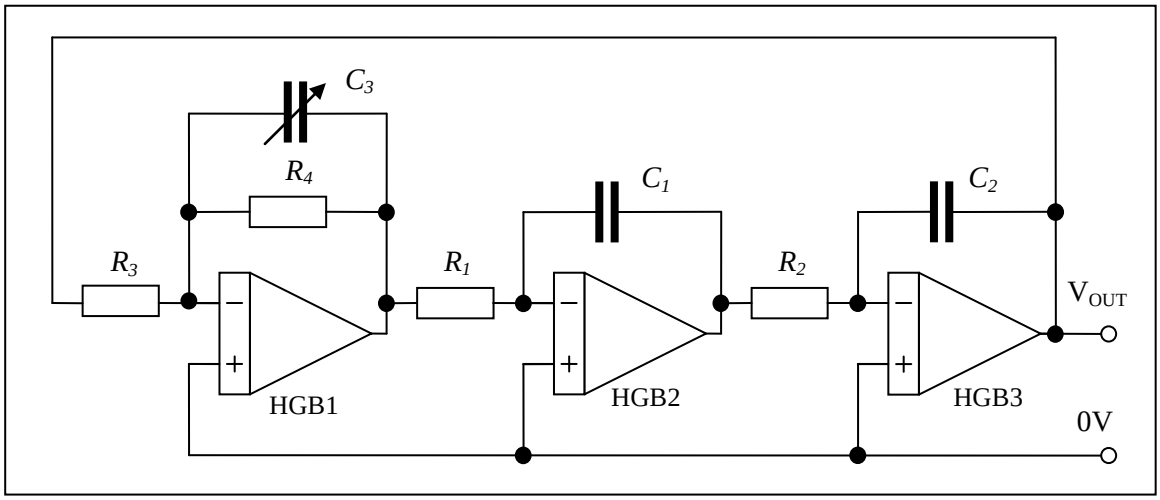


Fig. 2.1 A borderline oscillator circuit (BOC) for measuring tanδ

2.1 Analysis of the borderline oscillator circuit (BOC)

I shall assume the following: -

a) The inverter has a small (DC) gain error (use ±1% or matched resistors): gain $G \approx -R_4/R_3$
b) All resistors have negligible phase error. It is recommended that stray capacitance at the virtual earths is kept small and low to medium value resistances are used (< 10kΩ).
c) The feedback capacitors can be described as complex capacitors.

The open loop gain consists of the inverter/low pass filter (HGB1) with phase shift $\delta_3 = \omega R_4 C_3$ and two integrators (HGB2 and HGB3) with small phase errors δ_1 and δ_2 respectively [4]. The open loop transfer function is: -

$$L(s) = \frac{G}{1+sR_4C_3} \times \frac{-(1+j\delta_1)}{sR_1C_1} \times \frac{-(1+j\delta_2)}{sR_1C_2} = \frac{G(1+j\delta_1)(1+j\delta_2)}{(1+j\delta_3)s^2R_1C_1R_2C_2} \quad \text{with} \quad \delta_3 = \omega R_4 C_3$$

1. Part 4, monograph 1: "High gain blocks".
2. Part 4, monograph 3: "High accuracy amplifiers, integrators and differentiators". See section 4.
3. http://en.wikipedia.org/wiki/Barkhausen_stability_criterion
4. Part 1, monograph 6: "Low phase error capacitors and inductors".

All three phase errors are very small so that, to a good approximation: -

$$L(s) \approx \frac{G(1+j\delta_1)(1+j\delta_2)(1-j\delta_3)}{s^2 R_1 C_1 R_2 C_1} \approx \frac{G(1+j(\delta_1+\delta_2-\delta_3))}{s^2 R_1 C_1 R_2 C_1}$$

The output is connected to the input: -

$$L(s)V_{OUT} = V_{OUT} \quad \Rightarrow \quad V_{OUT} = 0 \quad \text{or} \quad L(s) = 1$$

Clearly the latter represents the case from which one can deduce: $s = j\sqrt{\frac{|G|}{R_1 C_1 R_2 C_2}} \times \sqrt{1+j(\delta_1+\delta_2-\delta_3)}$

For sustained oscillation the real part of *s* must be zero and, therefore: $\delta_3 = \delta_1 + \delta_2$

The average δ of the two capacitors is then easily calculated from the critical value of capacitance which just sustains oscillation: -

$$\delta_{AVE} = \frac{\delta_1+\delta_2}{2} = \frac{\omega_R R_4 C_3}{2} \quad \text{with} \quad \omega_R = \sqrt{\frac{|G|}{R_1 C_1 R_2 C_2}}$$

If, on the other hand, the extra capacitance is removed the result, when triggered, is an exponentially decaying sine wave, represented by a value of *s* with a (negative) real component. To a good approximation: -

$$C_3 = 0 \quad \Rightarrow \quad s \approx j\sqrt{\frac{|G|}{R_1 C_1 R_2 C_2}} \times \left(1 + j\frac{(\delta_1+\delta_2)}{2}\right)$$

One can separate out the real and imaginary parts to obtain: -

$$s = \sigma + j\omega = -\delta_{AVE}\omega_R + j\omega_R \quad \text{with} \quad \omega_R = \sqrt{\frac{|G|}{R_1 C_1 R_2 C_2}}$$

From the real part one can deduce δ_{AVE}: -

$$\sigma = -\delta_{AVE}\omega_R \quad \Rightarrow \quad \delta_{AVE} = -\frac{\sigma}{\omega_R}$$

The time constant τ_D of the decaying sine wave is related to σ. This is easily seen if one expresses the decaying sine wave in the usual form, with a numerically positive time constant: -

$$V_{OUT} = V_P e^{\sigma t} e^{j\omega t} = V_P \exp\left(-\frac{t}{\tau_D}\right)\exp(j\omega t)$$

$$\sigma = -\frac{1}{\tau_D} \quad \Rightarrow \quad \delta_{AVE} = \frac{1}{\tau_D \omega_R}$$

A practical method is counting the number, *n*, of cycles taken for the amplitude to fall by a factor 1/e: -

The period of each cycle is: $T = \frac{1}{f_R} = \frac{2\pi}{\omega_R} \quad \Rightarrow \quad \tau_D = nT = \frac{2\pi n}{\omega_R}$

From which the solution is: $\delta_{AVE} = \frac{1}{2\pi n}$

2.2 Results (sustained oscillation method)

(See log book: test data from 24 March 2011).

$R_3 = R_4$ = 4k7 ±5% (G = -1 ± 10%) C_3 = 140pF ±5% Measured frequency: 3330Hz

Check the frequency. Both $R_1 = R_2 = R_I$ = 4k7 ±5%; $C_1 = C_2 = C_I$ = 10nF ±10% (Polyester): -

$$f_R = \frac{1}{2\pi R_I C_I} = \frac{1}{2\pi \times 4.7 \times 10^3 \times 10^{-8}} = 3386Hz$$

From above: $\delta_{AVE} \approx \pi f_R R_4 C_3 = 3.14 \times 3330 \times 4.7 \times 10^3 \times 140 \times 10^{-12} = 6.9 \times 10^{-3} \pm 10\%$

Clearly it would be possible to reduce the uncertainty by measuring all component values with greater precision.

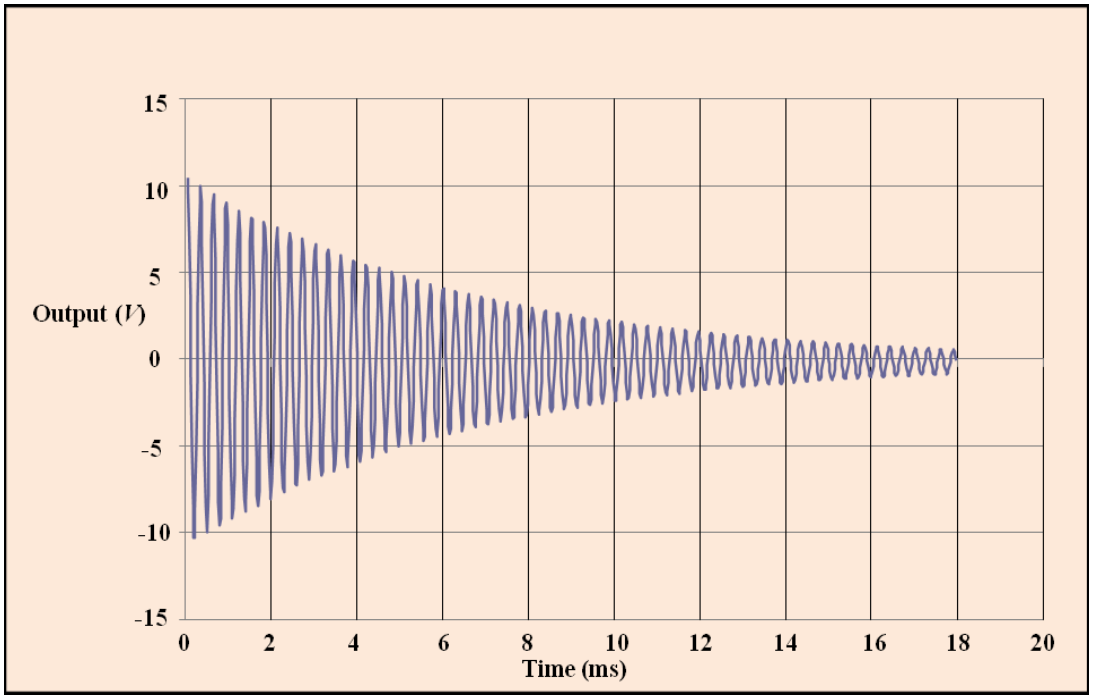


Fig. 2.2.1 Borderline oscillator with C_3 = 0

2.3 Results (decaying sinusoid method)

From fig. 2.1.1 count from where the amplitude is 10V (peak) to 3.7V (peak): n = 25±1

From above: $\delta_{AVE} = \frac{1}{2\pi n}$

The result is: $\delta_{AVE} \approx \frac{1}{50\pi} = 6.4 \times 10^{-3} \pm 4\%$

3. Some nice maths

Whereas the method above is simple and valid for low values of tan(δ) an exact analysis is possible if one converts the transfer functions into polar form. An alternative representation of the impedance of a real capacitor is, for example: -

$$Z_C = \frac{1}{j\omega C} + R_S = \frac{1}{j\omega C_I}\left(1 + j\omega R_S C_I\right) = \frac{\left(1 + j\tan(\delta)\right)}{j\omega C_I}$$

$$\text{Now} \quad 1 + j\tan(\delta) = \frac{\cos(\delta) + j\sin(\delta)}{\cos(\delta)} = \frac{1}{\cos(\delta)}\exp(j\delta) \quad \Rightarrow \quad Z_C = \frac{1}{j\omega C_I \cos(\delta)}\exp(j\delta)$$

The transfer function for each integrator in polar form is, therefore, of the form: -

$$T_I(\omega) = -\frac{Z_C}{R_I} = \frac{-1}{\omega R_I C_I \cos\delta}\exp(j\delta)$$

Similarly, the inverter/low pass filter has the following transfer function (by definition *G* is negative): -

$$T_{LP}(\omega) = \frac{G}{1 + j\omega R_4 C_3} = \frac{G}{1 + j\tan\delta_F} = G\cos\delta_F \exp(-j\delta_F)$$

The loop gain is: -

$$L(\omega) = T_{I1} \times T_{I2} \times T_{LP} = \frac{G\cos\delta_F}{\omega^2 R_1 R_2 C_1 C_2 \cos\delta_1 \cos\delta_2}\exp\left(j\left[\delta_1 + \delta_2 - \delta_F\right]\right)$$

For borderline/sustained oscillation the net phase shift is zero (modulo 360 degrees) and, therefore: -

$$\delta_F = \delta_1 + \delta_2$$

The average phase error due to the integrator capacitors is, therefore: -

$$\delta_{AVE} = \frac{\delta_1 + \delta_2}{2} = \frac{1}{2}\arctan(\omega_R R_4 C_3)$$

For the circuit to oscillate the magnitude of $L(\omega)$ must be one at the resonant frequency: -

$$|L(\omega)| = 1 \quad \Rightarrow \quad \omega_R = \sqrt{\frac{|G|\cos\delta_F}{R_1 R_2 C_1 C_2 \cos\delta_1 \cos\delta_2}}$$

If one assumes that the capacitors have the same phase error, δ_C, then $\delta_F = 2\delta_C$ and $\cos\delta_F = \cos 2\delta_C$

From a well known trigonometric identity $\cos 2\delta_C = \cos^2\delta_C - \sin^2\delta_C$ and so the exact solution, based on the assumptions, simplifies to: -

$$\omega_R = \sqrt{\frac{|G|}{R_1 R_2 C_1 C_2}\left(1 - \tan^2\delta_C\right)}$$

4. A prototype

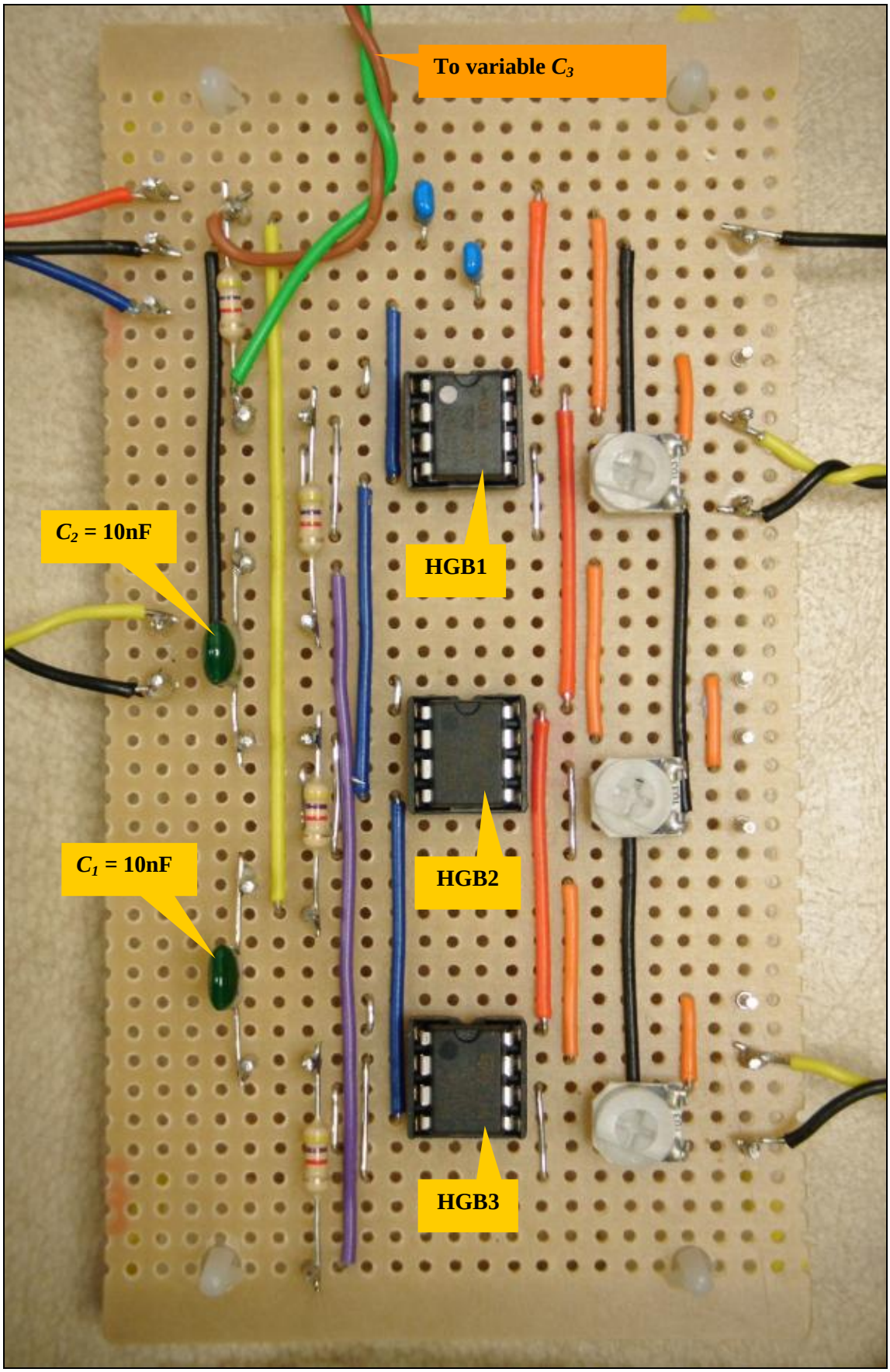


Fig. 4.1 Strip-board prototype borderline oscillator circuit

A two-stage HGB (type 2) based on a LF353 dual op-amp is sufficient for this applications. Each was adjusted for stability as an inverting amplifier (Gain = -1) with a high frequency square wave test signal. The trimpot was adjusted for ≈10% overshoot (ample stability at about 25% (k = 0.25)).

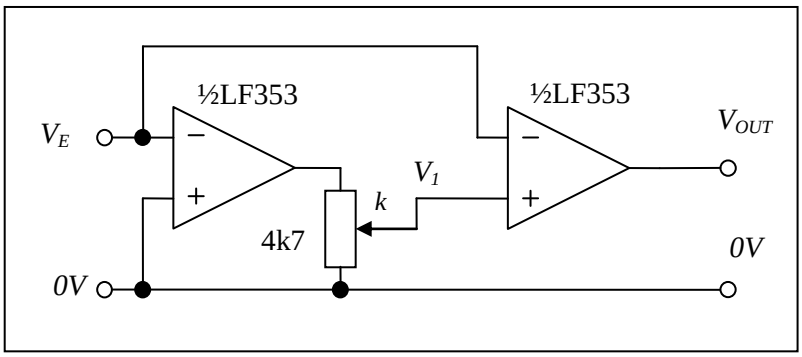


Fig. 4.2 A two-stage HGB (type 2)

I shall assume that the operational amplifiers are ideal integrators with a gain bandwidth product (GBWP) of $f_B = \dfrac{1}{2\pi\tau_B} \approx 4MHz$ for an LF353 [1].

For the circuit fig. 4.2: $V_1 = -\dfrac{k}{\tau_B s}V_E$ with $k = 0.25$ and $V_{OUT} = \dfrac{1}{\tau_B s}(V_1 - V_E)$

In a form normalised to the OPI time constant $(s = j\omega\tau_B/k)$ $\Rightarrow$ $H(s) = -\dfrac{V_{OUT}}{V_E} = \dfrac{1}{ks}\left(1+\dfrac{1}{s}\right)$

This is a type 2 HGB with: $\tau_2 = \tau_B \approx 4\times10^{-8}s$ and $\tau_1 = \tau_B/k \approx 1.6\times10^{-7}s$

The closed loop transfer function is [2]: $T_N(s) = -\dfrac{Z_2}{Z_1}D_N(s)$

Where the "*D*" factor for the amplifier/filter is [3]: - $\tau_I = R_I C_I \approx 4.7\times10^{-5}s$

$$D_2(s) \approx 1 + \alpha\beta\left(\frac{f}{f_1}\right)^2 - j\alpha\beta\left(\frac{f}{f_1}\right)^3 \quad \text{with} \quad \alpha = \frac{\tau_2}{\tau_1} = k = 0.25 \quad \text{and} \quad \beta = 1 + \frac{R_2}{R_1} = 2$$

The frequency f_1 corresponds to the time constant τ_1 (i.e. 1MHz). For the integrators it is very similar [4]: -

$$D_2(f) \approx 1 + \left(\frac{f}{f_1}\right)^2 - j\left\{\left(\frac{f}{f_1}\right)^3 + \delta\left(\frac{f}{f_1}\right)\right\} \quad \text{with} \quad \delta = \alpha\frac{\tau_1}{\tau_I} << 1$$

At 3kHz the ratios are: $f/f_1 \approx 3\times10^{-3}$ and $\delta = \alpha\dfrac{\tau_1}{\tau_I} \approx 2.1\times10^{-4}$ so that the errors are negligible [5]

1. Part 4, monograph 1: "High gain blocks". See section 2.2
2. Part 4, monograph 3: High accuracy inverting amps, integrators and differentiators". See section 2.
3. Ibid. See section 3.
4. Ibid. See section 4.
5. Ibid. See section 4.2

5. A voltage controlled high Q filter/oscillator

The borderline oscillator circuit could form the basis of a very high Q band-pass filter or accurate two-phase oscillator. Such a circuit could prove useful, for example, as part of a very sensitive null detector for the most demanding bridge applications. The usual approach is to incorporate notch filters at 50Hz and 150Hz and a band-pass filter at the bridge operating frequency (nominally 25Hz or 75Hz, also known as the "carrier" frequency). The effectiveness of the filtering is limited by the fact that the supply frequency varies, albeit by a small amount, limiting the quality factor of the notch filters. The carrier frequency is usually phase locked to the supply and also varies, limiting the quality factor of the band pass filter. The following provides a different approach – a band pass filter with adjustable frequency and quality factor, which can be operated at a much higher quality factor. The bridge balancing algorithm is more complicated but higher levels of interference could be tolerated. The basic idea is to use a pair of multipliers to adjust the feedback and control the resonant frequency and quality factor/gain in a way that is largely independent. The multipliers could be analogue or R-2R type multiplying digital to analogue converters. If you are interested in constructing/testing a prototype please contact the author.

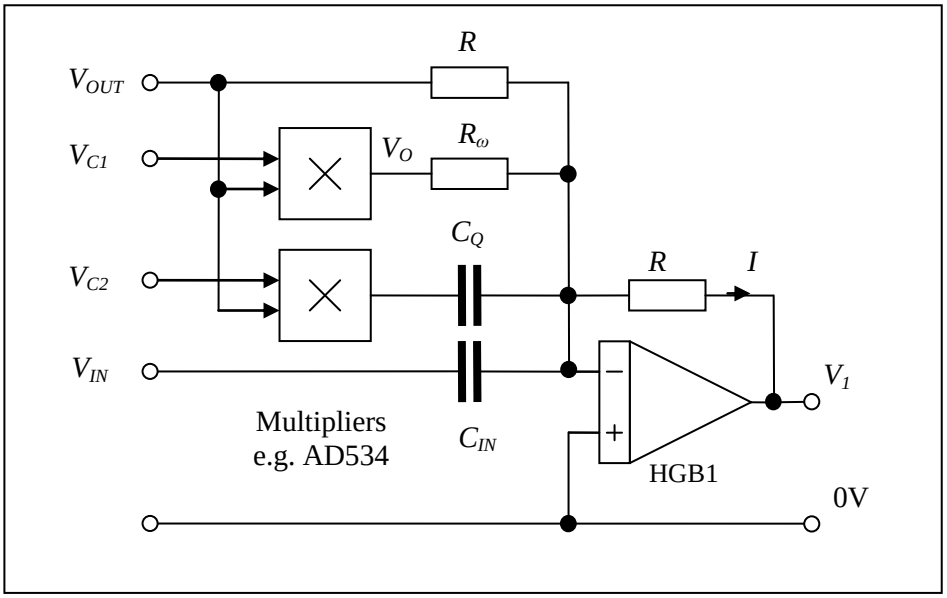


Fig. 5.1 Modified inverter stage for voltage control of resonant frequency and quality factor

5.1 Analysis of the modified inverter

The basic principle of the circuit is to vary the gain and phase shift of the inverter by a small amount. An additional input is also provided in the case of the band-pass filter. The result is a voltage controlled band pass filter, as the following analysis shows: -

The output of the analogue multiplier is typically: $V_O = \frac{V_{C1} \times V_{OUT}}{10}$ etc.

According to Ohm's law and Kirchhoff's law in the complex representation $(s = j\omega)$: -

$$I = \frac{V_{OUT}}{R} + \frac{V_{C1}}{10}\frac{V_{OUT}}{R_\omega} + \frac{V_{C2}}{10}V_{OUT}sC_Q + V_{IN}sC_{IN} = -\frac{V_1}{R}$$

Multiply by R and collect terms: -

$$V_{OUT}\left(1 + \frac{V_{C1}}{10V}\frac{R}{R_\omega} + \frac{V_{C2}}{10V}sRC_Q\right) + V_1 = -V_{IN}sRC_{IN}$$

Assume two identical integrators so that: -

$$V_{OUT} = V_1 \times \left(\frac{1 + sR_S C_I}{sR_I C_I} \right)^2 \qquad \Rightarrow \qquad V_1 = V_{OUT} \times \left(\frac{sR_I C_I}{1 + sR_S C_I} \right)^2$$

With which one can eliminate V_1: -

$$V_{OUT} \left(1 + \frac{V_{C1}}{10} \frac{R}{R_\omega} + \frac{V_{C2}}{10} sRC_Q + \frac{s^2 R_I^2 C_I^2}{(1 + sR_S C_I)^2} \right) + V_{IN} sRC_{IN} = 0$$

Simplify the algebra by defining parameters: -

$$\alpha = \frac{V_{C1}}{10} \frac{R}{R_\omega} \qquad \beta = \frac{V_{C2}}{10} RC_Q \qquad \tau_I = R_I C_I \qquad \tau_{IN} = RC_{IN} \qquad \text{and} \qquad \tau_\delta = R_S C_I = \frac{\tan\delta}{\omega}$$

$$V_{OUT} \left(1 + \alpha + \beta s + \frac{\tau_I^2 s^2}{(1 + \tau_\delta s)^2} \right) + V_{IN} \tau_{IN} s = 0$$

Multiply by $(1 + \tau_\delta s)^2$ and rearrange for the transfer function: -

$$T(s) = \frac{V_{OUT}}{V_{IN}} = -\frac{\tau_{IN} s (1 + \tau_\delta s)^2}{\tau_I^2 s^2 + (1 + \alpha + \beta s)(1 + \tau_\delta s)^2}$$

For very low frequencies and around the resonant frequency this is very nearly the ideal transfer function for a second order band pass filter, which one can see more clearly with the following approximations. I shall assume small tanδ $(|\tau_\delta s| << 1)$ and a small frequency adjustment range $(|\alpha| << 1)$ but allow the possibility of low quality factor $(|\tau_Q s| \approx 1)$: -

$$|\tau_\delta s| << 1 \Rightarrow \qquad (1 + \tau_\delta s)^2 \approx 1 + 2\tau_\delta s$$

$$|\tau_Q s| \approx 1 \qquad \text{and} \qquad |\alpha| << 1 \Rightarrow \qquad (1 + \alpha + \beta s)(1 + 2\tau_\delta s) \approx 1 + \alpha + (\beta + 2\tau_\delta) s + 2\tau_Q \tau_\delta s^2$$

The transfer function is more recognisable: -

$$T(s) \approx \frac{-\tau_{IN} s (1 + 2\tau_\delta s)}{(\tau_I^2 + \beta\tau_\delta) s^2 + (\beta + 2\tau_\delta) s + 1 + \alpha}$$

The factor $(1 + 2\tau_\delta s)$ in the numerator represents a small phase shift which is negligible compared to the phase shift of the filter itself. Divide top and bottom by $(1 + \alpha)$ to obtain the standard form: -

$$T(s) \approx \frac{-\frac{\tau_{IN} s}{(1 + \alpha)}}{\frac{(\tau_I^2 + \beta\tau_\delta)}{(1 + \alpha)} s^2 + \frac{(\beta + 2\tau_\delta)}{1 + \alpha} s + 1} = \frac{-G\tau_N s}{\tau_N^2 s^2 + 2\xi\tau_N s + 1}$$

The natural frequency is: $$\omega_N = \frac{1}{\tau_N} = \sqrt{\frac{(1 + \alpha)}{(\tau_I^2 + \beta\tau_\delta)}}$$

The natural frequency is largely adjusted by control parameter α : -

$$\beta\tau_\delta << \tau_I^2 \quad \Rightarrow \quad \omega_N \approx \frac{\sqrt{(1+\alpha)}}{\tau_I}$$

For small variations the control is linear so that, to a good approximation with frequency in Hz: -

$$\alpha << 1 \quad \Rightarrow \quad f_N \approx \frac{1}{2\pi R_I C_I}\left(1+\frac{\alpha}{2}\right)$$

The deviation in natural frequency from the nominal is, therefore: -

$$\Delta f_N = \frac{V_{C1}}{20}\frac{R}{R_\omega} f_N$$

The damping ratio can be deduced from: $2\xi\tau_N = \frac{(\beta + 2\tau_\delta)}{1+\alpha}$ N.B. The quality factor is: $Q = \frac{1}{2\xi}$

This time the main control parameter is β and the effect of α is small and so, at the resonant frequency, to a good approximation: -

$$\Rightarrow \quad 2\xi \approx \frac{(\beta + 2\tau_\delta)}{\tau_N} = \frac{V_{C2}}{10} \times \frac{RC_Q}{R_I C_I} + 2\tan\delta \quad \text{noting that} \quad \frac{\tau_\delta}{\tau_N} = \omega_N R_S C_I = \tan\delta$$

If the control voltage V_{C2} is set to zero the damping ratio is simply the tanδ of the integrator capacitors and one has the borderline oscillator circuit. The control voltage can be negative and so the damping ratio can be controlled all the way down to zero (infinite *Q*) and the circuit oscillates. The upper limit is determined by the component values.

At the resonant frequency the gain depends on the damping ratio (i.e. is controlled by V_{C2}) with no upper limit: -

$$T(\omega_N) = -\frac{G}{2\xi}$$

The factor, *G*, can be deduced from the numerator: -

$$G\tau_N = \frac{\tau_{IN}}{(1+\alpha)} \quad \Rightarrow \quad G = \frac{\tau_{IN}}{\tau_N(1+\alpha)} = \frac{\sqrt{1+\alpha}}{R_I C_I} \times \frac{RC_{IN}}{(1+\alpha)} = \frac{RC_{IN}}{R_I C_I \sqrt{1+\alpha}}$$

Once again the effect of small α is small and so, to a good approximation: -

$$G \approx \frac{RC_{IN}}{R_I C_I}$$

The analysis shows that one can control the resonant frequency and damping ratio (gain) largely independently but there is a small degree of interaction possibly requiring a number of iterations. The (in-phase) test signal is applied and the output of the in-phase and quadrature detectors monitored. The damping ratio (i.e. Q) is first set to the required level of sensitivity and then the resonant frequency is adjusted so that the output of the quad detector is zero.

For a dual phase oscillator it is recommended that a continuous amplitude control is applied to the damping ratio. Consult the author for further advice.

6. Compensating for tanδ

It should be possible, in principle, to compensate for the phase error of a capacitor at least to the level of it being reproducible and stable. This is another possible project – contact the author if you require further advice.

In this case it is more convenient to model the capacitor with a large parallel resistance.

The impedance of the parallel combination is: $$Z = \frac{R_P \times 1/sC_I}{R_P + 1/sC_I} = \frac{1}{sC_I}\left(\frac{1}{1+1/sR_PC_I}\right) = \frac{1}{sC_I}\left(\frac{1+j/\omega R_PC_I}{1+1/\omega^2 R_P^2C_I^2}\right)$$

From which the precise result is: $$\tan\delta = \frac{1}{\omega R_P C_I}$$

The phase error is constant over the low to medium frequency range so that the equivalent parallel resistance is inversely proportional to frequency: -

$$R_P = \frac{1}{\omega C_I \tan\delta}$$

One can, therefore, compensate for the equivalent resistance by injecting a current through a compensating resistor with an inverting amplifier.

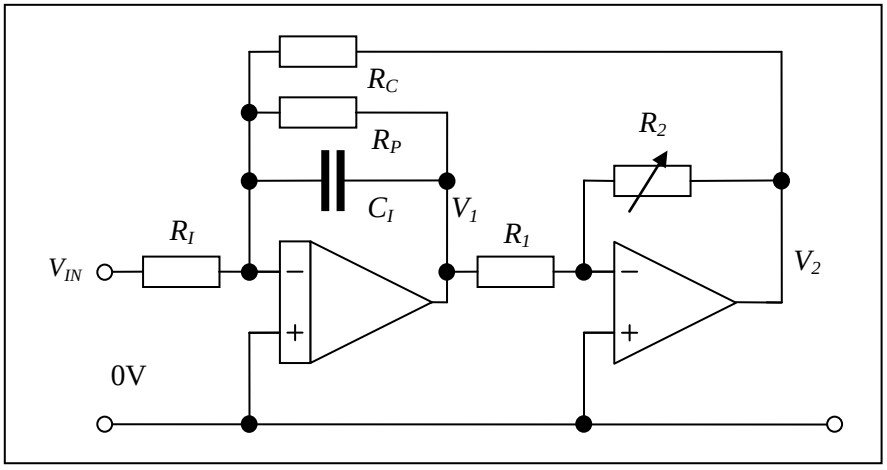


Fig. 6.1 Compensating for tanδ

The gain of the amplifier needs to be varied as the operating frequency is changed.

$$I = \frac{V_1}{R_P} + \frac{V_2}{R_C} = 0 \quad \Rightarrow \quad R_C = -R_P\frac{V_2}{V_1} = -\frac{V_2/V_1}{\omega C_I \tan\delta}$$

Once a suitably large compensating resistor is chosen the gain of the inverting amplifier can be calculated: -

$$|G| = \frac{R_2}{R_1} = \omega R_C C_I \tan\delta$$

In practice one could adjust the compensation to achieve sustained oscillation in a borderline oscillator configuration (see fig. 2.1 with C_3 = 0) after which one can be confident that each integrator provides precisely 90 degrees phase shift at that frequency.

A simulated large capacitor circuit

1. Introduction

This ingenious circuit was developed by JDY in the 1980s, based, probably, on an original idea by Robert Cutkosky of NBS [1]. JDY gave it this name because it behaves, at the operating frequency, like a very large capacitor. A better model, over the frequency range of interest, is a large capacitor in parallel with a resistor. The main application is to improve the stability of an actively driven multi-stage IVD or ratio transformer: -

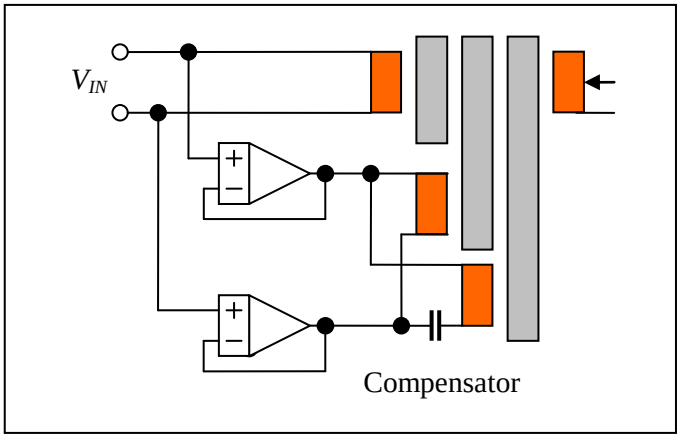


Fig. 1.1 A simulated capacitor application (An F18 type ratio transformer [2]).

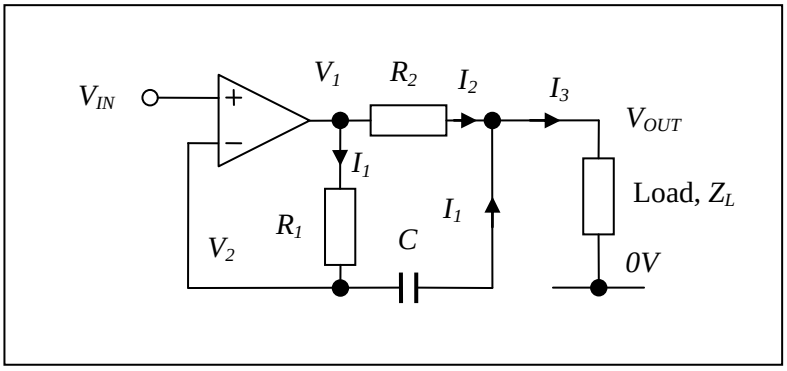


Fig. 1.2 A simulated large capacitor circuit

At very low frequency the capacitor has high impedance, compared to R_1, and the circuit is a voltage follower with 100% feedback via R_1 ($V_1 = V_{IN}$). Resistor, R_2, is then in series with the output and the load impedance.

At high frequency the capacitor has low impedance, compared to R_1, and provides 100% feedback from V_{OUT}. This time the follower action is to make $V_{OUT} = V_{IN}$.

Over the full range of frequency the circuit behaves like series impedance, Z_E: the resistor, R_2, in parallel with a large capacitor, C_E: -

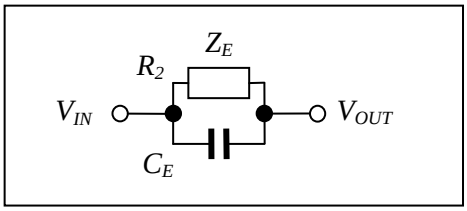


Fig.1.3 The equivalent circuit

1. National Bureau of Standards (NBS) is now known as The National Institute for Science and Technology (NIST). Ref. required.
2. Part3, monograph 7: “An F18 type ratio transformer bridge”.

It is argued elsewhere that, for stability, the resistance, R_2, needs to be of the same order (very approximately) as the maximum source resistance and the natural (cross-over) frequency of the compensator needs to be the same as the natural frequency of the ratio transformer [1]. Typical values are, therefore: -

$$R_2 \approx 100\Omega \quad f_N \approx 0.16Hz \left(\omega_N \approx 1rads^{-1}\right) \Rightarrow \left(R_1 + R_2\right)C \approx 1 \Rightarrow R_1 \approx 1M\Omega \text{ and } C \approx 1\mu F$$

The equivalent capacitance is: $C_E \approx 10mF$

This is far too large for a component capacitor, hence the need for a simulated capacitor.

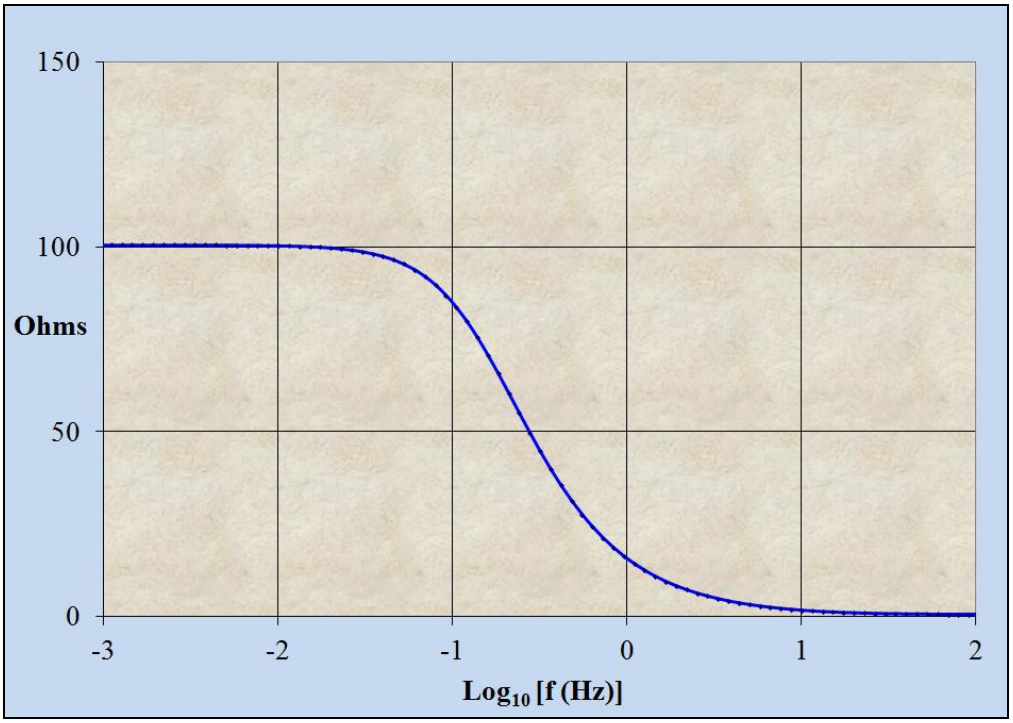


Fig. 2.1 Magnitude of a typical compensator vs frequency

2. Basic analysis

If one assumes an ideal op-amp the action of feedback is to make the inverting input the same as the non-inverting input. Also, according to Ohm's and Kirchhoff's laws, referring to fig. 1.2, in the complex representation $\left(s = j\omega\right)$: -

$$V_2 = V_{IN} \Rightarrow I_1 = \frac{V_1 - V_{IN}}{R_1} = sC\left(V_{IN} - V_{OUT}\right) \quad I_2 = \frac{V_1 - V_{OUT}}{R_2} \text{ and } I_1 + I_2 = I_3 = \frac{V_{OUT}}{Z_L}$$

Combine the three and switch to admittances where possible to simplify the algebra (maximum double-decker equations):-

$$A_L V_{OUT} = sC\left(V_{IN} - V_{OUT}\right) + A_2\left(V_1 - V_{OUT}\right) \text{ with } A_L = \frac{1}{Z_L} \text{ and } A_2 = \frac{1}{R_2}$$

We can eliminate V_1 using the first equation with: $A_1 V_1 = A_1 V_{IN} + sCV_{IN} - sCV_{OUT}$

$$\Rightarrow A_L V_{OUT} = sCV_{IN} - sCV_{OUT} + A_2 V_{IN} + A_2 R_1 sCV_{IN} - A_2 R_1 sCV_{OUT} - A_2 V_{OUT}$$

1. Part 3, monograph 4: "Three-stage RTs" See section 3.

Re-arrange to find: $$V_{OUT}\left(A_L + sC + A_2R_1sC + A_2\right) = V_{IN}\left(sC + A_2 + A_2R_1sC\right)$$

We are expecting an equation of the form: $$\frac{V_{IN} - V_{OUT}}{I_3} = Z_E \quad \text{or} \quad \frac{V_{IN} - V_{OUT}}{V_{OUT}} = \frac{A_L}{A_E}$$

Where Z_E is the equivalent impedance and $A_E = \frac{1}{Z_E}$ is the equivalent admittance of the circuit. With a little algebra: -

$$\frac{V_{IN}}{V_{OUT}} - 1 = \frac{A_L}{A_E} = \frac{A_L}{sC + A_2 + A_2R_1sC}$$

From which it is easy to see that the equivalent admittance is: -

$$A_E = A_2 + \left(1 + A_2R_1\right)sC$$

This is the admittance of R_2 in parallel with a capacitor, C_E, with: $C_E = \left(1 + \frac{R_1}{R_2}\right)C$ QED.

The reciprocal is: -

$$Z_E = \frac{R_2}{1 + \left(R_1 + R_2\right)sC} \quad \text{with natural frequency} \quad f_N \approx \frac{1}{2\pi\left(R_1 + R_2\right)C}$$

3. Errors due to limited gain-bandwidth product

The accuracy of the circuit depends on the gain-bandwidth product of the op-amp. The analysis is a bit complicated and is relegated to appendix 1. The full expression for the equivalent impedance, referring to fig. 1.2 is, to a very good approximation: -

$$Z_E \approx \frac{1}{\left(1 + A_2R_1\right)sC + A_2}\left(1 + \frac{1}{H(s)}\left(1 + A_2Z_L\right)\left(1 + sR_1C\right)\right)$$

$H(s)$ is the open loop characteristic of the op-amp (or high gain block).

For the main application the load impedance is an energising winding: inductance in series with a small resistor, which is usually much smaller than R_2. A good approximation is, therefore: -

$$\frac{R_L}{R_2} << 1 \;\Rightarrow\; 1 + A_2Z_L = 1 + \frac{R_L + sL}{R_2} \approx 1 + \tau_L s \quad \text{with} \quad \tau_L = \frac{L}{R_2}$$

Similarly, the time constant $\tau_N = R_1C$ corresponds, very nearly, to the natural frequency: $f_N \approx \frac{1}{2\pi\left(R_1 + R_2\right)C}$

We are primarily interested in the low frequency characteristics, including the maximum "DC" gain of the op-amp. The low frequency model (high gain low-pass filter) is appropriate [1]. The equation for the impedance takes the elegant form: -

$$H(s) \approx \frac{G}{1 + \tau_P s} \;\Rightarrow\; Z_E \approx \frac{1}{\left(1 + A_2R_1\right)sC + A_2}\left(1 + \frac{1}{G}\left(1 + \tau_L s\right)\left(1 + \tau_N s\right)\left(1 + \tau_P s\right)\right)$$

1. Part 4, monograph 1: "High gain blocks" See section 3.1

Typical values of the time constants (for a three-stage ratio transformer compensator) are: -

$$\text{Load: } \tau_L = \frac{L}{R_2} \approx \frac{5.4H}{100\Omega} = 54ms \qquad \text{Natural frequency: } \tau_N = R_1 C \approx 1M\Omega \times 1\mu F = 1s$$

The dominant pole of an op-amp corresponds to the frequency at which the gain starts to drop $\left(f_P \approx 50Hz\right)$: -

$$\tau_P = \frac{1}{2\pi f_P} \approx 3.2ms \quad \text{with open loop "DC" gain: } \; G > 10^5$$

The magnitude error increases with frequency starting from the lowest level determined by the op-amp "DC" gain.

Fig. 3.1 Magnitude error for typical values $\left(\log_{10}\left|\left|D(s)\right|-1\right|\right)$

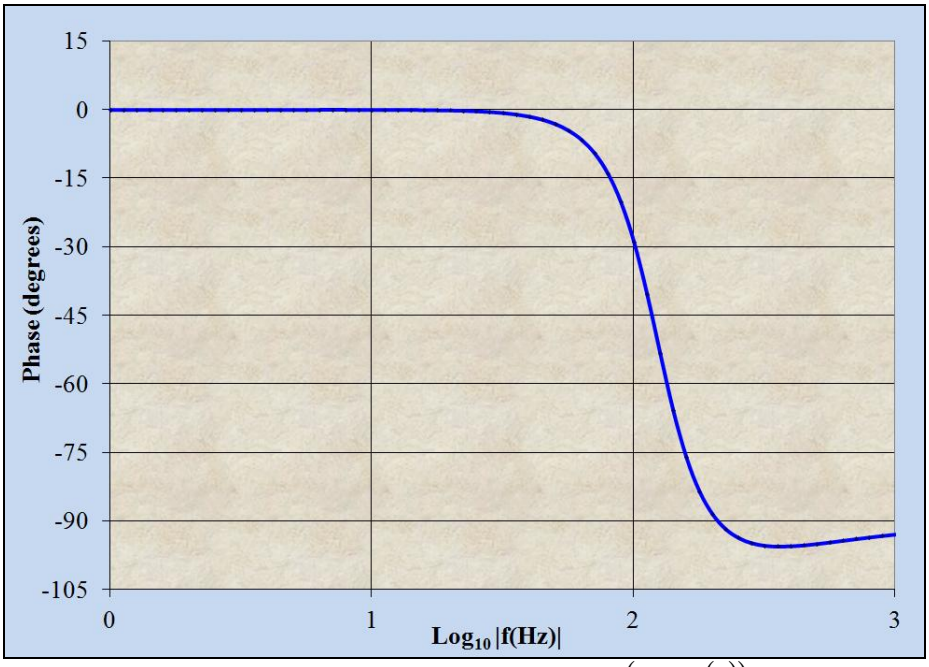


Fig. 3.2 Phase error for typical values $\left(\arg D(s)\right)$

4. Practical considerations

At high frequency, with an inductive load, the feedback factor is 100% with the usual resonant peak at the unity gain frequency [1] (See section 3.1.1). Even a small parallel load capacitance (e.g. 100pF of interwinding capacitance) would interact with the series resistance, R_2, to add extra phase shift and reduce stability margin. Fortunately this is easily fixed with a snubber (R and C in series) in parallel with the load. Typical values are 100nF and 10Ω. As frequency increases (above 160kHz) the load presents a high impedance and the snubber resistance dominates. The feedback factor drops to 10%.

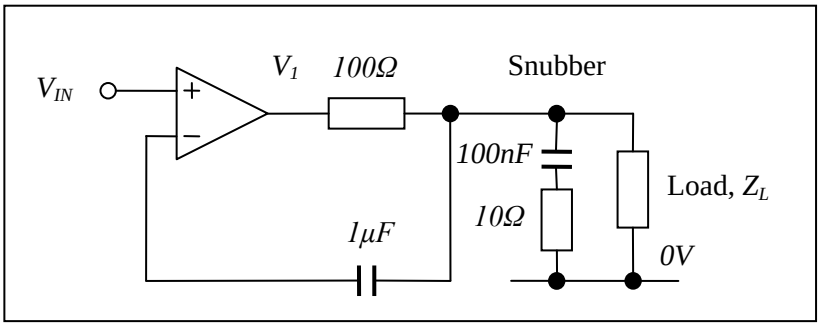


Fig. 4.1 High frequency model (typical values)

At lower frequency the snubber capacitance dominates and the load looks partly capacitive. The equivalent capacitance, C_E, is much larger than the load capacitance and the result is a slightly reduced energising voltage. A small error in the energising voltage has a much reduced effect on the main ratio accuracy, especially with a three-stage transformer [2].

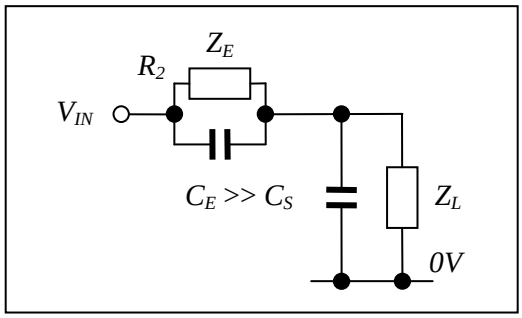


Fig. 4.2 Low frequency model

At very low frequency the parallel resistance, R_2, dominates but the load looks like a very small resistor. The snubber capacitance then presents very high impedance and, again, the effect is negligible.

The complete circuit does not warrant detailed modelling as the frequencies of interest, and impedances, are so widely separated. See appendix 2.

One may also ask if it is worth increasing the open loop gain with, for example, a two-stage high gain block (e.g. a type 3 HGB would provide a high input impedance) [1] (see section 4.3). In practice any errors (magnitude and phase or noise) introduced via energising windings are much reduced and have negligible effect on the accuracy of the ratio windings [2]. If necessary one could employ a composite op-amp with, for example, a matched BJT or JFET pair [3]. The extra stage not only boosts open loop gain at low frequency but also improves noise performance.

In practice accuracy is limited by common mode rejection ratio and, for very high accuracy, it is necessary to bootstrap the power supply.

1. Part 4, monograph 1: “High gain blocks”.
2. Part 3, monograph 3: “Two-stage IVDs and RTs”. See section 4.2.4
3. Part 5, monographs 2 &3: “Low noise BJT preamps” and “Low noise JFET preamps”

Appendix 1: Error analysis (due to limited gain-bandwidth product)

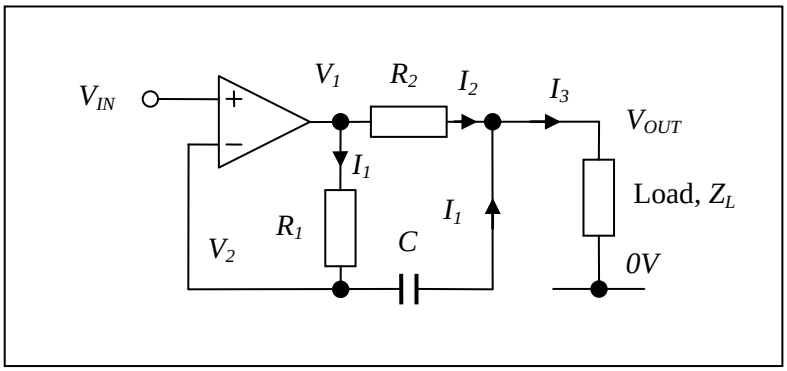


Fig. 1.2 A simulated large capacitor circuit (repeated for convenience)

According to Ohm's and Kirchhoff's laws, in the complex representation $(s = j\omega)$: -

$$I_1 = A_1(V_1 - V_2) = sC(V_2 - V_{OUT}) \quad I_2 = A_2(V_1 - V_{OUT}) \quad \text{and} \quad I_1 + I_2 = I_3 = A_L V_{OUT}$$

$$\Rightarrow \quad A_L V_{OUT} = sCV_2 - sCV_{OUT} + A_2 V_1 - A_2 V_{OUT}$$

The output of the op-amp is: $V_1 = H(s)(V_{IN} - V_2)$

$$\Rightarrow \quad A_L V_{OUT} = sCV_2 - sCV_{OUT} + \{A_2 H(s) V_{IN} - A_2 H(s) V_2\} - A_2 V_{OUT}$$

The changes are shown in curly brackets for easy checking.

We can eliminate V_2 with: $I_1 = \{A_1 H(s) V_{IN} - A_1 H(s) V_2\} - A_1 V_2 = sCV_2 - sCV_{OUT}$

$$\Rightarrow \quad V_2(sC + A_1 H(s) + A_1) = A_1 H(s) V_{IN} + sCV_{OUT}$$

Introduce impedance parameter Z to simplify the algebra: $Z = \dfrac{1}{sC + A_1 H(s) + A_1}$

$$\Rightarrow \quad V_2 = A_1 H(s) Z V_{IN} + sCZ V_{OUT}$$

$$\Rightarrow \quad \begin{aligned} A_L V_{OUT} &= \{sCA_1 H(s) Z V_{IN} + s^2 C^2 Z V_{OUT}\} \\ &- sCV_{OUT} + A_2 H(s) V_{IN} \{- A_2 A_1 H^2(s) Z V_{IN} - A_2 H(s) sCZ V_{OUT}\} - A_2 V_{OUT} \end{aligned}$$

Re-arrange to find: -

$$\frac{V_{OUT}}{V_{IN}}\left(A_L - s^2 C^2 Z + sC + A_2 H(s) sCZ + A_2\right) = sCA_1 H(s) Z + A_2 H(s) - A_2 A_1 H^2(s) Z$$

$$\frac{V_{IN}}{V_{OUT}} - 1 = \frac{A_L}{A_E} = \frac{A_L - s^2 C^2 Z + sC + A_2 H(s) sCZ + A_2 - sCA_1 H(s) Z - A_2 H(s) + A_2 A_1 H^2(s) Z}{sCA_1 H(s) Z + A_2 H(s) - A_2 A_1 H^2(s) Z}$$

It is tempting here to employ the approximation $A_1 H(s) Z \approx 1$but that could be a mistake. The difference between the two largest elements is, for example: -

$$A_2 H(s) - A_2 A_1 H^2(s) Z = \left(A_2 H(s)(sC + A_1 H(s) + A_1) - A_2 A_1 H^2(s)\right) Z = A_2 H(s)(sC + A_1) Z$$

Similarly, three other terms afford some simplification: -

$$-s^2C^2Z + sC - sCA_1H(s)Z = \{-s^2C^2 + sC(sC + A_1H(s) + A_1) - sCA_1H(s)\}Z = sCA_1Z$$

With these substitutions things are looking better: -

$$\frac{A_L}{A_E} = \frac{A_L + A_2H(s)sCZ + A_2 + sCA_1Z - A_2H(s)(sC + A_1)Z}{sCA_1H(s)Z + A_2H(s)(sC + A_1)Z}$$

Another pair of terms now cancel: $A_2H(s)sCZ - A_2H(s)(sC + A_1)Z = -A_2H(s)(A_1)Z$

$$\Rightarrow \quad \frac{A_L}{A_E} = \frac{A_L + A_2 + (sCA_1)Z - A_2H(s)(A_1)Z}{sCA_1H(s)Z + A_2H(s)(sC + A_1)Z}$$

Finally one has: -

$$A_2 - A_2H(s)(A_1)Z = \{A_2(sC + A_1H(s) + A_1) - A_2H(s)(A_1)\}Z = \{A_2(sC + A_1)\}Z$$

$$\Rightarrow \quad \frac{A_L}{A_E} = \frac{A_L + A_2(sC + A_1)Z + (sCA_1)Z}{sCA_1H(s)Z + A_2H(s)(sC + A_1)Z}$$

$$\Rightarrow \quad \frac{A_L}{A_E} = \frac{A_L + (A_2 + A_1)sCZ + A_2A_1Z}{(A_2 + A_1)sCH(s)Z + A_2A_1H(s)Z}$$

All terms now add and one can safely assume the approximation. In the limit that $|H(s)|$ tends to infinity one obtains the result in section 2 - a useful check on the algebra so far: -

$$|H(s)| \to \infty \text{ then } Z \to 0 \text{ and } A_1H(s)Z = 1 \Rightarrow \frac{A_L}{A_E} = \frac{A_L}{(1 + A_2R_1)sC + A_2}$$

Take out a factor A_L from the numerator: -

$$\frac{1}{A_L} = Z_L \Rightarrow A_L + (A_2 + A_1)sCZ + A_2A_1Z \approx A_L(1 + (A_2 + A_1)sCZ_LZ + A_2A_1Z_LZ)$$

For the denominator take out a factor $A_1H(s)Z$: $((1 + A_2R_1)sC + A_2)A_1H(s)Z$

The impedance is, therefore: -

$$\frac{1}{A_E} = Z_E = \frac{1}{(1 + A_2R_1)sC + A_2}\left(\frac{1}{A_1H(s)Z}\right)(1 + (A_2 + A_1)sCZ_LZ + A_2A_1Z_LZ)$$

The first item is the ideal expression for the impedance and the terms in brackets are correction factors, both with magnitudes approximating 1 at low frequency. Introduce the deviation factor $D(s)$: -

$$Z_E = \frac{1}{(1 + A_2R_1)sC + A_2}D(s)$$

with: $D(s) = \frac{1}{A_1H(s)Z} + \frac{(A_2 + A_1)sCZ_L + A_2A_1Z_L}{A_1H(s)}$

From above one has: $$\frac{1}{Z} = sC + A_1 H(s) + A_1$$

$$\Rightarrow \quad D(s) = 1 + \frac{sC + A_1 + (A_2 + A_1)sCZ_L + A_2 A_1 Z_L}{A_1 H(s)}$$

In most cases one can assume a very good approximation: -

$$A_2 >> A_1 \quad \Rightarrow \quad D(s) \approx 1 + \frac{1 + sR_1C + Z_L A_2 sR_1C + Z_L A_2}{H(s)}$$

The error term factorises to something simple: -

$$D(s) \approx 1 + \frac{1}{H(s)}(1 + A_2 Z_L)(1 + sR_1C)$$

The load impedance is usually inductance in series with a winding resistance, R_L, which is much smaller than R_2. A good approximation is, therefore: -

$$\frac{R_L}{R_2} << 1 \quad \Rightarrow \quad 1 + A_2 Z_L = 1 + \frac{R_L + sL}{R_2} \approx 1 + \tau_L s \quad \text{with} \quad \tau_L = \frac{L}{R_2}$$

Similarly the time constant $\tau_N = R_1 C$ corresponds, very nearly, to the natural frequency: $f_N \approx \frac{1}{2\pi(R_1 + R_2)C}$

We are primarily interested in the low frequency characteristics, including the "DC" gain of the op-amp. The low frequency model (high gain low-pass filter) is appropriate. The equation for the impedance then takes the elegant form: -

$$H(s) \approx \frac{G}{1 + \tau_P s} \quad \Rightarrow \quad D(s) \approx \left(1 + \frac{1}{G}(1 + \tau_L s)(1 + \tau_N s)(1 + \tau_P s)\right)$$

Appendix 2: Snubber example calculations

Unity gain frequency (typically 5MHz): $\omega_B = 3.14 \times 10^7 \quad \Rightarrow \quad R_S + \frac{1}{j\omega_B C_S} \approx (10 - 0.3j)\Omega$ and the snubber resistance dominates. The impedance of the load capacitance (100pF) is much larger: $\frac{1}{\omega_B C_L} \approx 320\Omega$

At the lowest operating frequency (25Hz) the compensator capacitance (10mF) has low impedance compared to the parallel resistance (100Ω): $\omega_O = 157 \quad \Rightarrow \quad \frac{1}{\omega_O C_E} \approx 1.57\Omega$

At very low frequency $(\omega < 10)$ the snubber capacitance has very high impedance $\frac{1}{\omega C_S} > 1M\Omega$ and has negligible effect on the circuit.

At the highest operating frequency (160Hz): $\omega_O = 10^3 \quad \Rightarrow \quad \omega L \approx 5k\Omega$ and $\frac{1}{\omega C_S} \approx 10k\Omega$ is comparable but $\frac{1}{\omega C_E} \approx 0.1\Omega$ and the result is a small (10ppm in-phase) error in the energising voltage and negligible.

A simulated negative capacitor circuit

1. Introduction

An actively driven three stage ratio transformer (e.g. of the F18 type) has, potentially, a very high input impedance, capable of high accuracy even with a significant source resistance. In practice, however, the interwinding capacitance sets an upper limit which, at the very least, results in a high quadrature imbalance, even at a low operating frequency. Also, the capacitance between primary and secondary is a problem. Fortunately, for the very highest accuracy applications, there is a simple and ingenious solution: a simulated negative capacitor circuit which neutralises the interwinding capacitance over a wide range of frequency.

The basic principle is to employ positive feedback, through a matching capacitor, with a low-pass filter in the loop to ensure stability at high frequency. The cut-off frequency of the filter is chosen to be well above the operating frequency but sufficiently low to ensure adequate attenuation and stability margin at high frequency.

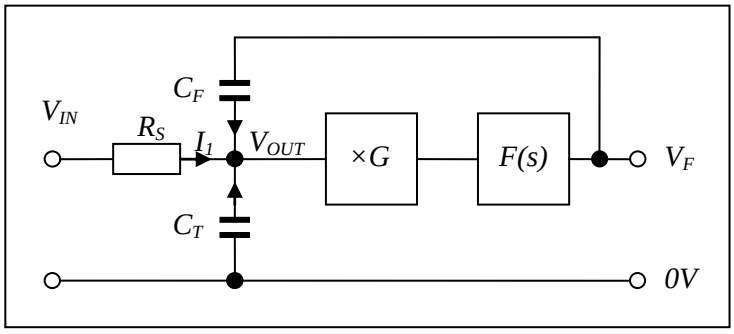


Fig. 1.1 Outline schematic of a negative capacitor

V_{OUT} = Voltage at the transformer primary. Ideally $V_{OUT} = V_{IN}$ and $I_1 = 0$

R_S = Source resistance.

C_T = Transformer interwinding capacitance.

C_F = Matching feedback capacitance $(C_F \approx C_T)$.

G = Gain of a wide-band amplifier (variable and approximately ×2).

$F(s)$ = Transfer function of a low pass filter.

1.1 The scale of the problem

A typical scenario is a source resistance of up to R_S = 100Ω and interwinding capacitance of up to C_T = 1nF. To put the latter into perspective it is equivalent to adding 10m of coax cable to the primary side of the bridge.

The operating frequency is 75Hz. Without the negative capacitor the transfer function would be a low-pass filter circuit for which the transfer function is, in the complex representation $(s = j\omega)$: -

$$T(s) = \frac{1}{1 + sR_S C_T} \approx 1 - sR_S C_T + s^2 {R_S}^2 {C_T}^2 \approx 1 - j4.7\times10^{-5} - 2\times10^{-9}$$

The in-phase error is small but the quadrature component of 47ppm could be a problem. Also, the capacitance between primary and secondary produces a similar effect. In a typical bridge the transformer ratio is adjusted for in-phase balance (a null) and the quadrature servo injects a signal to achieve a quadrature balance. The injected signal has a phase accuracy of around 2mrad (0.2%) so that a quadrature imbalance of 50ppm would result in an in-phase error of 0.1ppm. It is worth the cost of a bit of extra circuitry to neutralise this effect. It is hardly more practicable to add extra cable (or a capacitor) to the other side of the bridge.

It should be noted that the transformer interwinding capacitance also has a phase error (the insulating material is usually PTFE) and the type of feedback capacitance should be chosen to have similar dielectric properties (polystyrene is the best match). It is for the same reason that matching cables should be used on the reference and variable arms of the bridge, achieving near quadrature balance.

2. Basic circuit analysis

According to Kirchhoff's law the currents into the node add up to zero and Ohm's law provides, in the complex representation $(s = j\omega)$: -

$$I_1 + (V_F - V_{OUT})sC_F + (0 - V_{OUT})sC_T = 0$$

With $V_F = GF(s)V_{OUT}$ $\Rightarrow$ $I_1 = V_{OUT}(sC_T + sC_F - GF(s)sC_F)$

At low frequency $F(s) \approx 1$ and the gain is adjusted so that $G = \frac{C_T + C_F}{C_F}$ $\Rightarrow$ $I_1 \approx 0$ as required

The circuit works best when: $C_F \approx C_T$ $\Rightarrow$ $G \approx 2$

We can simplify the algebra, without loss of generality, by assuming $C_F = C_T$ and $G = 2$

From Ohm's law: $I_1 = \frac{V_{IN} - V_{OUT}}{R_S}$ $\Rightarrow$ $\frac{V_{IN} - V_{OUT}}{R_S} = 2sC_T V_{OUT}(1 - F(s))$

$$\Rightarrow \quad T(s) = \frac{V_{OUT}}{V_{IN}} = \frac{1}{1 + 2sR_S C_T(1 - F(s))}$$

3. A circuit with a simple first order filter

A simple first order low pass filter has a maximum phase shift of 90 degrees at high frequency and the loop gain can be made to fall below 0dB (×1) with a phase margin of close to 90 degrees.
The transfer function is of the form: -

$$F(s) = \frac{1}{1 + \tau s}$$

N.B. The time constant is typically related to a resistance and capacitance of the form: $\tau = RC$

Now $1 - F(s) = \frac{\tau s}{1 + \tau s}$ $\Rightarrow$ $T(s) = \frac{1 + \tau s}{1 + \tau s + 2R_S C_T \tau s^2}$

In normalised form $(s = j\omega/\omega_N)$: $T(s) = \frac{1 + as}{1 + as + s^2}$ with $\omega_N = \frac{1}{\sqrt{2R_S C_T \tau}}$ and $a = \tau\omega_N$

This is exactly the same as for a two-stage low pass filter and no accident as the positive feedback acts just like a bootstrap stage. The error analysis is the same. At low frequency: -

$$\omega << \omega_N \quad \Rightarrow \quad T(s) = 1 - \frac{s^2}{1 + as + s^2} \approx 1 - s^2 + as^3$$

In more convenient form with frequency in Hz: -

$$f << f_N \quad \Rightarrow \quad T(f) \approx 1 + \left(\frac{f}{f_N}\right)^2 - a\left(\frac{f}{f_N}\right)^3 \quad \text{with} \quad f_N = \frac{1}{2\pi\sqrt{2R_S C_T \tau}}$$

The phase error is small but the in-phase error is only second order with respect to frequency and is a problem, as the following demonstrates: -

3.1 Example calculation

A filter cut-off frequency of 16kHz $\left(\tau = 10^{-5}\right)$ has been found to be a reasonable compromise for operation at 75Hz.

$$f_N = \frac{1}{2\pi\sqrt{2R_S C_T \tau}} \approx 113kHz \quad \text{and} \qquad a = \tau 2\pi f_N = 7.1$$

The value of parameter *a* is quite high indicating plenty of stability margin (see the monograph "Two-stage filters" by the same author). With the negative capacitor the significant error components are: -

$$f << f_N \qquad \Rightarrow \qquad T(f) \approx 1 + \left(\frac{f}{f_N}\right)^2 - a\left(\frac{f}{f_N}\right)^3 \approx 1 - 4\times10^{-7} - j2\times10^{-9}$$

At 2ppb the quadrature error is now negligible but the in-phase error of 0.4ppm is too high for many applications. The latter is due to the phase shift of the filter as the following demonstrates.

4. A circuit with a two-stage filter

The in-phase error of the previous example can be reduced to negligible levels by the use of a two-stage filter. This could be done with an actively driven two-stage filter with its very low output impedance. In practice, however, it is most easily achieved with a directly connected RLC filter. The output impedance is sufficiently low and substantially inductive, obviating the need for an additional op-amp stage. See the monograph "Two-stage filters", section 6 by the same author.

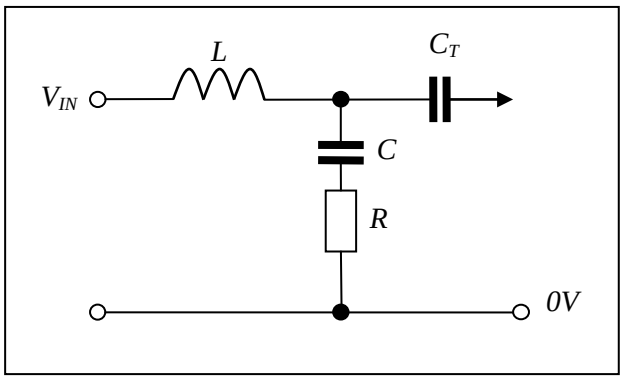


Fig. 4.1 An RLC filter connects directly to the feedback capacitor

If one neglects the loading effect of the feedback capacitor (see section 4.2) the transfer function for the filter is: -

$$F(s) = \frac{1 + sR_F C_F}{1 + sR_F C_F + s^2 C_F L_F} \qquad \text{so that:} \qquad 1 - F(s) = \frac{s^2 C_F L_F}{1 + sR_F C_F + s^2 C_F L_F}$$

From above the overall transfer function is, after multiplying top and bottom by $\left(1 + sR_F C_F + s^2 C_F L_F\right)$: -

$$T(s) = \frac{1}{1 + 2sR_S C_T (1 - F(s))} = \frac{1 + sR_F C_F + s^2 C_F L_F}{1 + sR_F C_F + s^2 C_F L_F + s^3 2R_S C_T C_F L_F}$$

This is the same as a three-stage low-pass filter which, in normalised form, $\left(s = j\,\omega/\omega_N\right)$ is: -

$$T(s) = \frac{1 + as + bs^2}{1 + as + bs^2 + s^3} \qquad \text{with} \qquad a = R_F C_F \omega_N \quad b = C_F L_F \omega_N^2 \quad \text{and} \quad \omega_N = \frac{1}{\sqrt[3]{2R_S C_T C_F L_F}}$$

At low frequency, to a very good approximation: -

$$\omega << \omega_N \Rightarrow \qquad T(s) = 1 - \frac{s^3}{1 + as + bs^2 + s^3} \approx 1 - s^3 + as^4$$

In more convenient form, with frequency in Hz: -

$$f << f_N \Rightarrow \qquad T(f) \approx 1 - j\left(\frac{f}{f_N}\right)^3 + a\left(\frac{f}{f_N}\right)^4$$

The in-phase error is now fourth order, with respect to frequency, and truly negligible.

4.1 Example calculation

As above typical values for an application are: Source resistance: $R_S = 100\Omega$ and capacitance: $C_T = 1nF$

The filter natural frequency is again 16kHz with: $L_F = 1mH$ $C_F = 100nF$ and $R_F = 150\Omega$

The overall transfer function has parameters: -

$$\omega_N = \frac{1}{\sqrt[3]{2R_S C_T C_F L_F}} = 3.68 \times 10^5 \text{ rads}^{-1} \text{ (58.6kHz)} \qquad a = R_F C_F \omega_N = 5.5 \qquad b = C_F L_F \omega_N^2 = 13.5$$

The result is a low and broad peak around the natural frequency indicating plenty of stability margin: -

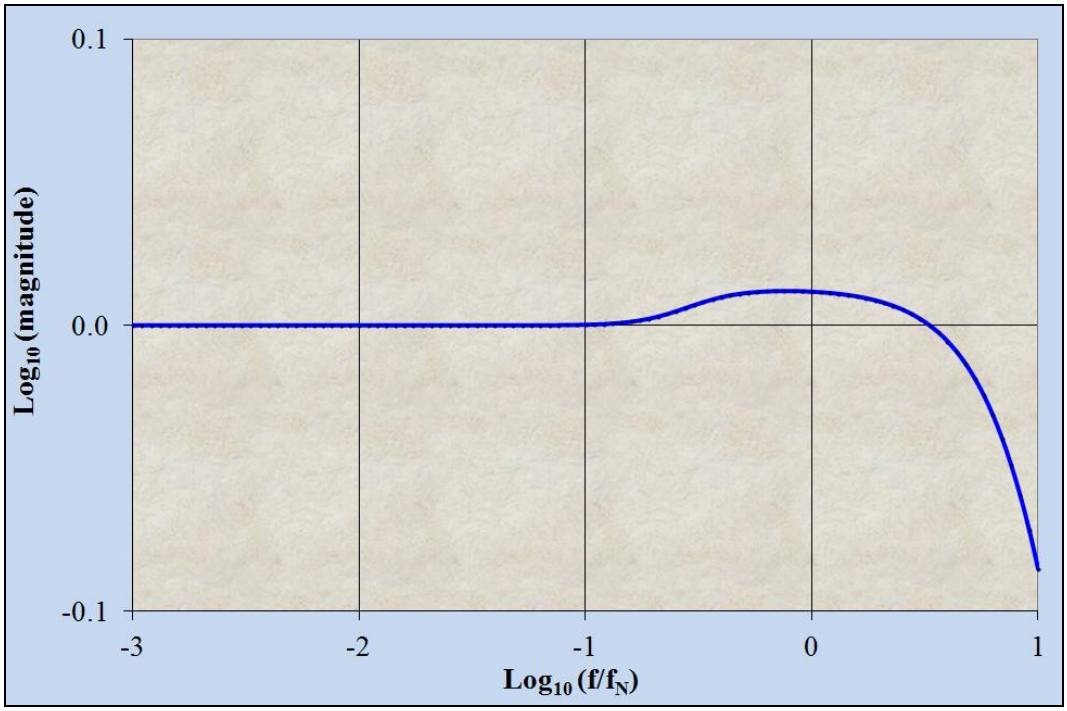


Fig. 4.1.1 Frequency response near the cut-off frequency (for dB ×20) [1]

(a = 5.5 b = 13.5)

1. Spreadsheet: “Negative cap analyser”.

4.2 A check on the output impedance

In section 4 it was assumed that the loading effect of the feedback capacitor was negligible. It is necessary to check this. The output impedance of the filter is the RC series combination in parallel with the inductance.

$$Z_{OUT}(s) = \frac{\left(R + \frac{1}{sC}\right)sL}{R + \frac{1}{sC} + sL} = \frac{(1 + sRC)sL}{1 + sRC + s^2CL}$$

At low frequency, the output impedance is, therefore, substantially the inductance: -

$$f << f_N \Rightarrow \quad Z_{OUT}(s) \approx sL_F$$

At 75Hz the impedance of the inductor and feedback capacitor are, respectively: -

$$Z_L(s) \approx j2\pi fL_F \approx 0.47\,j\Omega \qquad \text{and} \qquad Z_C(f) = -j\frac{1}{2\pi fC_T} \approx -2.1\,jM\Omega$$

The effect is to reduce the value of the capacitor by a negligible amount. Also, in order to keep the phase error due to resistance in series with the inductor to less than, say 10^{-3} radians that resistance must be less than 2kΩ, which is easily achieved.

4.3 A practical circuit

The most practicable circuit is a fully differential negative capacitor – neither side needs to be at earth potential. A pair of low cost JFET input op-amps (e.g. a TL072 dual) are perfectly adequate for this circuit. The leakage currents (a few pA) and their random fluctuations, flowing through the source resistance, have negligible effect. The interwinding capacitance can vary greatly and so the feedback capacitors are "select on test" to the nearest preferred value. The variable resistor allows one to adjust the differential gain for precise neutralisation.

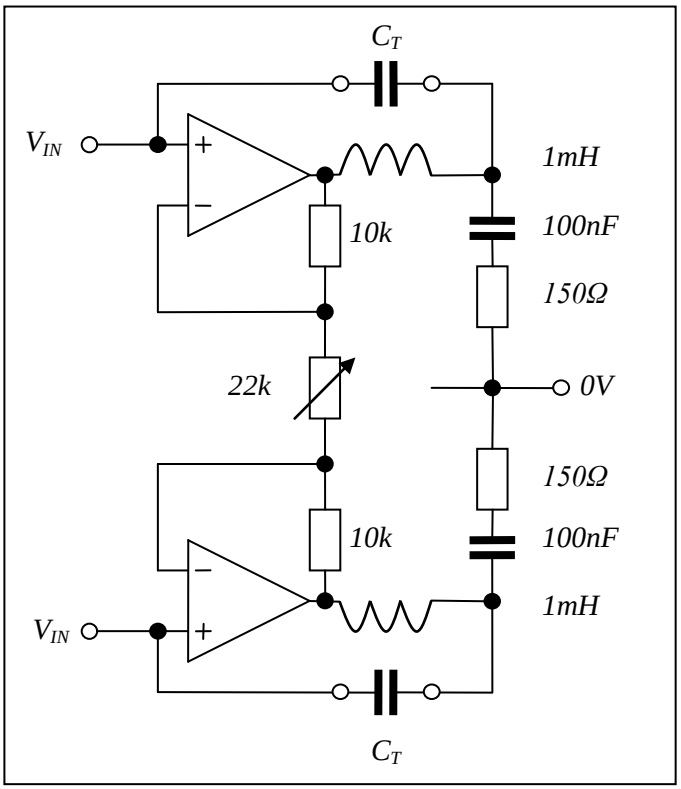


Fig. 4.3.1 A differential negative capacitor circuit.