Diode detectors for RF measurement
Part 1: Rectifier circuits, theory and calculation procedures.

By David W Knight[1]

Version[2] 0.09 (not yet finished), 1st Jan. 2016.

Please check the author's website to make sure you have the most recent version of this document and its accompanying files: http://www.g3ynh.info/ .

Abstract

This article addresses the subject of RF signal detection from the point of view of those who design and calibrate impedance matching bridges and other measuring instruments operating in the HF and VHF radio ranges. The principal discussion relates to the simple diode peak-detector; and covers the different possible circuit configurations, the associated theory, and the numerical methods needed for data analysis. Circuit techniques used to linearise the diode detector output are to be discussed in a separate document.

Analysis of the detector transfer characteristic for sinusoidal input shows that the AC-induced error term involves the zero-order modified Bessel function of the first kind (I_0). This result is not new, but it is often disregarded. The dynamic contribution is quite unlike the error that occurs for DC input, regardless of any compensatory modification of circuit parameters; which means that linearity correction schemes using a reference diode to produce a DC amplifier with a complementary gain law can never be perfect. It is also shown however, that the AC error is independent of frequency provided that the smoothing capacitor is 'large'. This means that the tracking detector system, which involves automatic self-calibration against a low-frequency precision rectifier, is theoretically sound.

By considering the power dissipated in the detector, it is shown that the input impedance can be calculated using first and zero-order modified Bessel functions of the first kind (I_0 and I_1). This allows the determination of detector transfer-functions that take source impedance into account. When this facility is combined with the ability to calculate the dynamic component of the peak detection error, a measurement of DC output taken with a calibrated voltmeter can be converted into a measurement of AC input without the need for an AC reference. The computation procedures required are not simple, but they are described in detail and given as Basic algorithms readily adaptable to any programming environment.

---

1 Ottery St Mary, Devon, England. http://www.g3ynh.info/

2 Revision history:
Version 0.09 (2016-01-01): Minor additions & corrections. References added. Added section 1.8. Comments on guard-ring structure in section 6.3. Re-rendered formulae using OO math.
Version 0.08 (2014-10-18): Eliminated perfect decoupling capacitors from equivalent circuits (redundant, confusing). Added section 6.1 - temperature dependence of I_S . Section 14.2 - series-parallel to series transformation, general form with macros. Section 14.3 - shunt diode rectifier with parasitics. Section 15 - content added. Problems with numerical instability still unresolved.

Diode detectors for RF measurement
Part 1: Rectifier circuits, theory and calculation procedures.

Table of Contents

Note on detector modelling routines

The program routines given in the text are available from the accompanying spreadsheet file: **det_models.ods**. To access, edit or copy the code, open the file using Apache Open Office[3] and select the top menu item: ' Tools / Macros / Organise Macros / OpenOffice Basic '. Then navigate to the library ' det_models.ods / Standard / detector_funcs '.

See the OO Basic Guide[4] for a description of the programming language. The StarOffice Basic programming guide[5] can also provide useful additional detail because it relates to the language from which OO Basic evolved.

3 http://www.openoffice.org/
4 https://wiki.openoffice.org/wiki/Documentation/BASIC_Guide
5 http://web.mit.edu/soffice_v8.0/pdfdoc/staroffice-BASIC.pdf

Introduction

The diode detector finds widespread use as a high-frequency voltmeter; its principal advantage, apart from the simplicity of the circuit, being the ability to provide a bandwidth of several hundred MHz with minimal attention to physical layout. This property means that detectors can be connected directly to the output ports of bridges and other measuring devices, thereby eliminating the need for down conversion or high-speed sampling. The essential preconditions are that the signal amplitude should be tailored to lie somewhere in the range from about 0.1 V to 10 V RMS, and that the source network should have a moderate load-driving capability.

The difficulty with the diode detector however, is that there is a big difference between using it to make a crude RF level indicator, and using it to make an accurate measuring instrument. This is primarily because the circuit behaves in an extremely non-linear manner for small signals and becomes only approximately linear as the signal voltage approaches the point at which the diode will be destroyed. Correction is not straightforward, and widely-used linearisation schemes involving a DC amplifier with a complementary gain law are not completely successful because the behaviour under AC excitation is not the same as under static conditions. The matter of turning the diode detector into an accurate broadband voltmeter is nevertheless well worth pursuing; not least because the perfect solution, an active rectifier or 'superdiode' circuit, is typically restricted to an upper frequency limit of about 50 kHz.

In this article, we will consider the diode detector on two ways: firstly, as a measuring device in its own right; and secondly, as a circuit module for inclusion in more elaborate instruments. In the latter case, we will assume that the implementation involves linearisation (or some other re-mapping) of the rectified output, using either digital or analogue methods. It is, of course, obvious that any correction process is dependent on a detailed knowledge of detector behaviour, and so the first part of the study is a necessary precursor to the second.

The archetypical diode detector is the series half-wave rectifier, either driven by a transformer, or having an RF choke as a DC short-circuit across its input. This can be analysed on the basis that the choke or transformer is perfect, and results in the principal mathematical relationship that governs the detector's dynamic behaviour. The solution for the DC output (involving the modified Bessel function of the first kind in zero order), although discussed in the academic literature[6 7 8], is evidently unfamiliar to the majority of commentators. A step-by-step analytical treatment therefore given in **section 12.1**. The insight it gives is central to the matter of voltage measurement, but it is not sufficient to solve the overall problem.

The first issue is that the prototype detector to which the analysis relates is not necessarily the best circuit to use. There are numerous rectifier configurations, with different properties and idiosyncrasies, and the variants should be considered carefully before making a choice. This is the material covered in **sections 1 - 5**. An important subtext to those considerations is that, in maximising the bandwidth of the diode voltmeter, it is a good idea to eliminate inductive devices. Doing that in a way that maximises the available input voltage range introduces resistance into the detector DC return path, and the resulting reduction in efficiency must be taken into account.

In **sections 6 - 11**, we take a look at the subject of diodes. The main point is to make the reader aware of the various types and their characteristics; although a considerable amount of background material is included, for context, and to contradict historical misinformation. From all of that, the best type of diode for a particular signal measurement applications should be clear.

The mathematical analysis leading to the DC output of the prototype detector is given in **section 12**. This is set out with all of the logical steps explained, so that it should be accessible to anyone

---

6 **Theory of the diode voltmeter**, C B Aitken, Proc. IRE, 26(7), July 1938. p859-876. [Analysis of the thermionic (valve / tube) diode detector].

7 **Polynomial economization of envelope detector static characteristics**, Z Cvetković and A Marković, IEEE Trans. on Consumer Electronics, 35(4), Nov. 1989. p876-881.

8 **Diode power detector DC operating point in six-port communications receivers**. S M Winter, H J Ehrn, A Koelpin, R Wiegel, Proc. 37th European Microwave Conf. Oct. 2007. p795-798.

with some knowledge of calculus (although it is possible to skip to the result in each subsection). The principal transfer function has already been mentioned; but in addition, it is necessary to derive expressions for the detector input impedance. Only then does precision measurement become generally possible, because knowledge of the input impedance allows the output reduction due to source loading to be determined.

Once we have an accurate knowledge of the detector transfer function, including loading effects, the detector qualifies as an absolute AC voltage measuring instrument; at least in the limited sense that an accurate DC measurement at the output can be converted into the peak value of a sinusoidal input, all without recourse to an AC voltage reference. The next problem however, is that the computational procedures are not purely analytical; which means that there is little to report in the way of handbook formulae. The solution is given in **section 13**, in the form of a library of numerical routines. This turns out to be a sizeable collection, most of which is involved with the calculation of modified Bessel functions and the circumvention of their argument-range restrictions. For the purpose of modelling detectors however, or for converting raw data into measurements, we can simply pass parameters to a main program, which uses the rest of the library as a collection of service routines.

After all of that, we still only have an accurate knowledge of the behaviour of the simple detector. Mercifully however, it transpires that we can take more realistic detectors, having source DC resistance, parasitic reactances, diode leakage and series resistance; and aggregate the parameters so that they can be passed to the simple detector programs. The necessary transformations are described in **section 14**.

Although, in **section 12**, we have described the detector transfer function in terms of modified Bessel functions, it is also possible to solve the integrals for average current and average power numerically. This allows analytically intractable variants of the of the detector problem, such as finite time constant (ripple), and diode parasitic resistance, to be included at the integration stage. In particular, it transpires that while diode parallel resistance can be separated analytically, diode series resistance (R_{ds}) cannot. Therefore, in **section 15**, we set up the numerical integration method for diodes with finite ohmic series resistance and compare it with the transformation method of **section 14**.

Note that the use of truncated series approximations is avoided in this treatment. Instead, we avail ourselves of the processing power of the modern personal computer, and use calculation methods capable of arbitrary precision within the limitations of computer floating-point arithmetic. In general, most of the routines are not noticeably slow to complete, even when implemented in interpreted Basic, except when there is a need to solve problems by numerical integration. For an alternative approach, capable of producing accurate closed-form analytical approximations of the more intractable variants of the detector problem, see the work of Xavier Le Polozec and Robert G Harrison[9] [10] [11] [12] [13] [14].

---

9 **Full nonlinear analysis of detector circuits using Ritz-Galerkin Theory**, R G Harrison, IEEE MTT-S International Microwave Symposium Digest, 1992, (Cat. No.92CH3141-9). [**Harrison 1992**]

10 **Nonsquarelaw behaviour od diode detectors analyzed by the Ritz-Galérkin method**, G Harrison, X Le Polozec, IEEE Trans. on microwave theory and techniques, 42(5) 1994. [**Harrison - Le Polozec 1994**]

11 **A full-range nonlinear diode detector model defined with the Ritz-Galerkin method**, X Le Polozec, R G Harrison. 2013, DOI: 10.13140/RG.2.1.2593.4248. Available from https://www.researchgate.net/profile/Xavier_Le_Polozec (Accessed 31st Jan 2015).

12 **Input impedance of series Schottky diode detector at low and high power**, X Le Polozec. DOI: 10.13140/RG.2.1.4530.9600. Available via ResearchGate, as above.

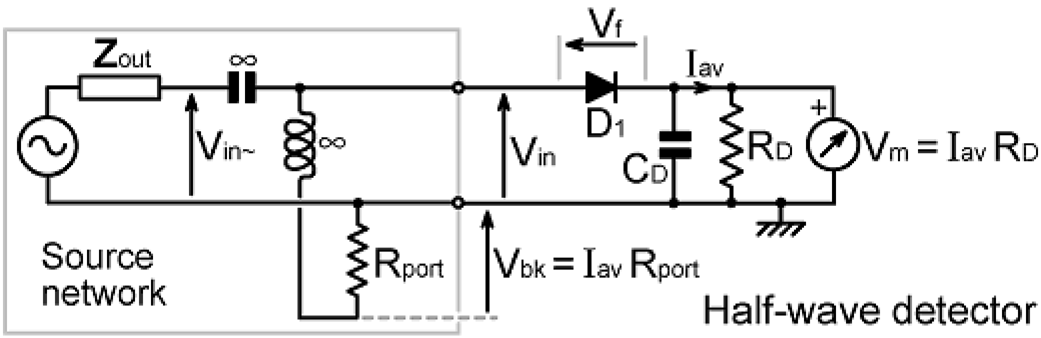
13 **Power conversion efficiency and sensitivity of diode detector calculated with the Ritz-Galerkin method**, X Le Polozec, R G Harrison, 2015, DOI: 10.13140/RG.2.1.5044.8808. Available via ResearchGate, as above.

14 **A full-range nonlinear nodel of the Latour's detector defined with the Ritz-Galerkin method**, X Le Polozec, DOI : 10.13140/RG.2.1.3753.3600. Available via ResearchGate, as above.

1. Series-diode half-wave rectifier

The simplest and best known RF detector circuit is the half-wave series-diode rectifier. Despite its familiarity however, the exact details of its operation are often poorly understood; not least because it is frequently subject to the casual presumption that the DC resistance looking back into driving network is irrelevant. It must be understood that the detector will not work if the source is a DC open circuit (although there may be a weak response due to diode reverse leakage current); and the AM demodulator on which the text-book explanation is usually based is driven by an IF or RF transformer, which is effectively a DC short-circuit. In voltage-sampling applications, there is often a significant DC resistance, and this reduces the measured output. Therefore, in the circuit below, the source network is depicted in a manner that gives an analytical separation between the DC resistance (R_{port}) and the output impedance ($\mathbf{Z}_{out}$). Of course, in practice, the electrical components involved will be common to both quantities, but separation for the purpose of predicting the circuit behaviour is usually trivial. The symbol R_{port} is used incidentally, because it is often the DC resistance of an RF sampling port on an item of test equipment, such as an impedance bridge.

1.1 Circuit behaviour and basic design principles



The source produces an on-load voltage $V_{in\sim}$ on its side of a hypothetical perfect coupling capacitor, and this has no DC component. $V_{in\sim}$ is also the voltage that will be measured using an AC-coupled probe at the input to an actual detector. It will be approximately sinusoidal provided that the input resistance of the detector is large relative to the magnitude of the output impedance of the generator, $|\mathbf{Z}_{out}|$. A DC offset is however added to the source voltage on the detector side of the hypothetical capacitor, so that the actual input voltage is:

$$V_{in} = V_{in\sim} - V_{bk}$$

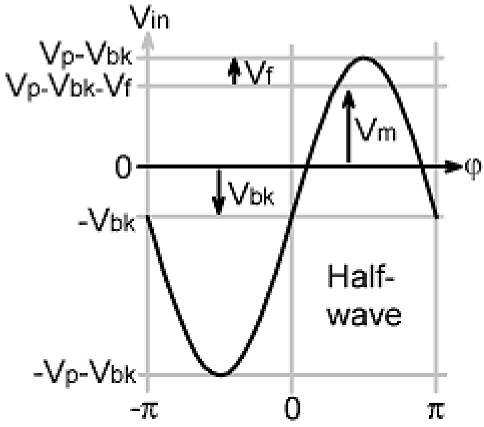
where, presuming that the diode polarity is chosen to give a positive output; V_{bk}, the *backoff voltage*, is negative relative to the ground rail shown. The backoff voltage arises because the average diode current (I_{av}) that gives rise to the DC output (measurement) voltage V_m must also flow through R_{port}. Thus the DC signal produced by the rectification process is shared between the output load resistance R_D and the port resistance in proportion to the relative values of those resistances. This means that the backoff voltage reduces V_m, for which reason R_{port} is shown below the ground rail in the diagram. V_{bk} can be measured at the input terminal by placing an RF choke in series with a high-input-resistance multimeter (assuming that the choke is properly chosen to have a high impedance at the generator frequency). In some AM radio receivers also, an RF-decoupled resistance is deliberately introduced in series with the final IF transformer output winding, so that the backoff voltage can be used for automatic gain control (AGC).

To calculate the detector output for a sinusoidal input, we can define:

$V_{in\sim} = V_p \sin\varphi$

where V_p is the peak voltage, and $\varphi = 2\pi f\, t$ is the time-varying phase angle. By this definition, $V_{in\sim}$ is the RMS of a sinusoidal voltage, so that:

$V_p = V_{in\sim} \times\sqrt{2}$



The output of the detector is then:

$V_m = V_p - V_f - V_{bk}$

where V_f is the effective diode forward voltage drop under dynamic conditions. For a semiconductor diode, V_f varies approximately logarithmically with the average forward current, but it also has a contribution given by the logarithm of the the peak-to-average current ratio. The problem of calculating V_f is investigated in detail in **section 12**; but as a rule of thumb, it is worth remembering that in signal detection circuits involving average diode currents in the region of 100 µA to 1 mA it is about 0.2 V for germanium point-contact diodes, about 0.35 V for high-inverse silicon-metal Schottky diodes (such as the 1N5711), and about 0.6 V for silicon P-N junction diodes (see **section 6.3**).

Another way of considering the detector is to say that the total detected voltage is the sum of the output and the backoff voltage. Thus:

$V_m + V_{bk} = V_p - V_f = \sqrt{2}\, V_{in\sim} - V_f$

When the port resistance is zero, the input voltage no longer has a DC component, and this becomes:

$V_m = V_p - V_f = \sqrt{2}\, V_{in} - V_f$

which is the more familiar expression for the output of a half-wave detector. Note that the output in this case can approach the peak value of the input voltage, provided that V_{in} is large relative to V_f. For this reason, rectifying detectors are often referred to as a peak detectors, even though the error in the measurement is large in the absence of some form of linearity correction.

In the circuit diagram given earlier, the output load resistance R_D is shown as being separate from the meter, and the meter is assumed to be a perfect voltmeter, i.e., of infinite input resistance. In practice, R_D is often the finite input resistance of a voltage measuring or sampling device, or the parallel combination of an input resistance and a load resistor. In a passive circuit, R_D might be the resistance of a voltmeter constructed by placing a resistor in series with a moving-coil microammeter. For example, a 100 µA meter padded-up to 100 kΩ by means of a series resistor makes a 10 V FSD (full-scale deflection) voltmeter.

The smoothing capacitor C_D should be chosen so that the time-constant $C_D \times R_D$ is at least 10 times the period (1/f) of the lowest frequency at which measurements will be made. If the lowest frequency is to be, say, 1 MHz and R_D = 10 kΩ, then we want $C_D \times R_D$ to be greater than 10 µs, i.e., C_D >1 nF. If a meter is used as the indicator, it makes no practical difference if the capacitor is somewhat larger than the minimum required, and so 0.1 μF (ceramic disk) is a typical choice; but very large capacitors will damp the meter response in a manner that depends on the driving network output impedance. A slow response can be desirable if the signal is jittery, but in some applications, such as finding the null point when adjusting a measuring bridge, a fast response is needed. 'Fast', on a human timescale, implies a time constant of less than about 10 ms (i.e., < 1 μF for a 10 kΩ load, < 0.1 μF for a 100 kΩ load, etc.).

The effect of the time constant on output frequency response when the detector is used for AM demodulation is discussed in **section 1.8**.

1.2 Diode peak inverse voltage

The minimum safe reverse voltage rating for the detector diode (V_{RM}) is most easily determined by starting with $R_{port} = 0$. In that case, for large inputs, C_D is charged to a constant voltage approaching $V_{in}\sqrt{2}$, whereas the most negative instantaneous voltage appearing at the detector input is approximately $-V_{in}\sqrt{2}$. Hence the diode must have a V_{RM} of at least $2\sqrt{2}$ (i.e., 2.82) times the maximum possible RMS input voltage.

Half-wave rectifier: Diode $V_{RM} > V_{in}(\max) \times 2\sqrt{2}$

Now if we allow R_{port} to be > 0, the output voltage V_m is reduced by an amount V_{bk}, but then a quantity $-V_{bk}$ is added to the most negative excursion of the input waveform. Hence the peak inverse voltage to which the diode is subjected is not affected by the port resistance.

1.3 Input chokes

The loss of sensitivity that results from having a finite port resistance can sometimes be a drawback, and the traditional solution to this problem (assuming that transformer coupling is not to be used) is to place an RF choke across the input to the detector. The choke should have a reactance that is high in comparison to the detector load resistance at the lowest frequency of operation. In HF radio applications, for example, 3.5 mH chokes are often used, and these have a nominal reactance of about +40 kΩ at 1.8 MHz. The practicalities of obtaining such a large inductance will however result in substantial self-capacitance; and this will resonate with the inductance at some frequency, and make the reactance of the choke capacitive at frequencies above that[15]. This leads some commentators to assume that RF chokes should not be used above the self-resonance frequency (SRF), but this rule is misleading. The SRF is actually the fundamental parallel resonance of the inductor, and corresponds to the frequency at which the impedance becomes extremely large in magnitude. Hence, from an AC point-of-view, the choke disappears at the SRF and so does not shunt the network. It also matters little whether the off-resonance reactance of the choke is positive or negative, the necessary criterion being only that it should have a large impedance magnitude in comparison to the other impedances in the network. Things go horribly wrong however at the first series-resonant frequency. This occurs at approximately twice the SRF, and corresponds to the point at which the length of the wire in the coil is one electrical wavelength. At this frequency, and for a range either side of it depending on the Q, the choke acts as a short-circuit and the output of the network is seriously reduced.

The potentially idiosyncratic behaviour of the DC bypass choke is not necessarily an issue if the inductor is well designed and the frequency-range is restricted, but spurious resonances cannot always be ruled-out without practical verification. In general, insofar as the use of chokes cannot be avoided, an RF choke composed of multiple segments (pie-wound) gives the highest first series-resonance and so is to be preferred. For a given length of winding wire also, a choke with a ferrite core gives more inductance than a non-magnetically-cored coil, and so will have a higher set of resonance frequencies. Many designers however, prefer to eliminate the choke wherever possible (they are expensive, as well as troublesome), and accept the loss of sensitivity. If the source is a complete DC open-circuit, the most basic solution is to place a resistor across the port; but this shunts the RF signal and it is sometimes better to place resistance across other elements in the network. A good design outcome is that which gives a port-resistance considerably smaller than the detector load resistance.

1.4 Large signal input impedance

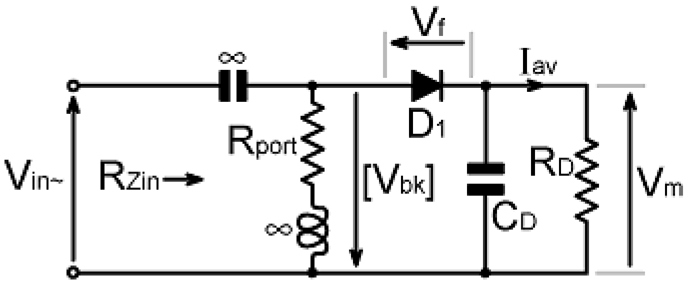
Both port resistance and diode forward voltage drop are sources of error requiring correction in precision measuring applications; but there is a third cause of error that must also be considered. This is the detector input impedance, which will obviously load the source network and cause its output to droop.

At first glance, the problem of calculating the input impedance seems intractable; because the diode does not obey Ohm's law and so the input resistance varies during the course of a cycle. For a single-frequency component of the signal however (this being what we must consider when determining impedance for the purpose of calculating the frequency response), it is sufficient to have an average picture of the input impedance. This can be obtained by considering the power delivered to the resistive elements in the network.

The AC excitation causes a direct current to circulate, and this causes the voltages V_m, V_f and V_{bk} to develop. These are related to the peak voltage by definition (because V_f is actually defined as the voltage error that occurs in peak detection). Thus:

15 **RF Chokes, their performance above and below resonance**, Courtney Hall, Ham Radio, June 1978, p40-42.

$V_p = V_{in\sim}\sqrt{2} = V_m + V_{bk} + V_f$



where $V_{in\sim}$ is the AC component of the input waveform considered in isolation from the DC offset caused by V_{bk}. Also, in the likely event that there is no physical separation between the AC source network and R_{port}, it is the voltage that would be measured using an AC-coupled probe.

Now let us define the total resistance around the loop as R_{tot}, i.e.:

$R_{tot} = R_D + R_{diode} + R_{port}$

where, to a fairly good approximation (which we will examine in detail in **section 12.3**):

$R_{diode} \approx V_f / I_{av}$

Neglecting energy loss due to the circulating harmonic currents produced by the rectification process; the total power delivered to the detector is:

$P_{in} = (V_m + V_{bk} + V_f)^2 / R_{tot}$

Thus:

$P_{in} = (V_{in\sim}\sqrt{2})^2 / R_{tot}$

i.e.:

$P_{in} = 2 V_{in}^2 / R_{tot}$

Now, if the input impedance is defined as R_{Zin}, we also have:

$P_{in} = V_{in}^2 / R_{Zin}$

so that:

$R_{Zin} = R_{tot} / 2$

Thus, assuming that harmonic generation is a lesser aspect of the behaviour of the lightly-loaded (voltage sampling) RF detector[16], the input resistance for AC signals is half the total DC resistance. This result might seem surprising, but it can be understood by thinking of the detector as a kind of voltage transformer. The RMS average of a DC signal is the same as its ordinary average, and so in the process of converting from AC to DC, the RMS level has been transformed-up by a factor of $\sqrt{2}$. Thus the detector is a 1:$\sqrt{2}$ voltage transformer, and hence a 1:$\sqrt{2}^2$ (i.e., 1:2) impedance transformer.

---

16 This assumption does not cause significant discrepancies in data analysis when making measurements using lightly-loaded detectors. Also, the AC component of the effective diode forward voltage drop, as derived in section **12**, must account for the dissipation of most of the harmonic energy.

It should be noted however that the detector input impedance is highly variable, depending on drive level. This is made obvious by substituting for R_{tot} in the expression above:

$$R_{Zin} = (R_D + R_{diode} + R_{port}) / 2$$

The variable term is R_{diode}, which is not a resistance in the normal sense, but a quantity related to the average power dissipation in the diode (this is obtained by integrating the product of instantaneous diode voltage and current over a cycle of the input waveform). If the detector input voltage is large in comparison to V_f however, then we have:

$$R_{Zin} \approx (R_D + R_{port}) / 2$$

This is the limiting input impedance under large signal conditions (neglecting parasitic reactances); and if R_{port} is zero due to the use of a choke or a coupling transformer, it reduces to:

$$R_{Zin} \approx R_D / 2$$

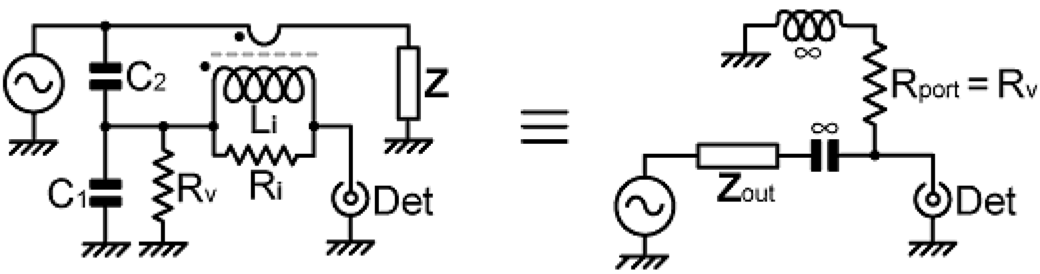
The limiting AC input resistance is a reasonable measure of the input impedance when the detector is driven hard, for best linearity. It is therefore useful for gaining an idea of the worst-case source loading. As the drive level is reduced however, due to the non-linear relationship between voltage and current, R_{diode} will increase and eventually come to dominate. Thus the input impedance increases as the drive is reduced, becoming very large (practically an open-circuit) when $V_{in} = 0$. In order to model the small-signal input impedance accurately, it is first necessary to develop a theory that describes the dynamic forward conduction characteristic. That matter is discussed from **section 12** onwards.

The limiting 1:2 impedance transformation rule applies to all non-voltage-multiplying detectors. In the case of the bridge rectifier (**section 4**), the limit is harder to approach because the detector places two diode forward voltage drops in series with the output. In the case of the bi-phase full-wave rectifier (**section 5**) the limit is slightly easier to approach, because the average diode current is shared between two diodes.

Note that an idiosyncrasy of half-wave rectification is that, since R_{port} is part of the source network, the rectification process causes power to be dissipated in the source network. This is usually of little consequence in signal-processing circuits; but it can be (or should be) an important consideration in the design of power-supplies.

1.5 Separation of port resistance and source impedance

There is no hard and fast rule for the analytical separation of the generator output impedance and the port resistance, save to say that is will usually be obvious. Take, for example, the circuit shown below, which is Douma's bridge[17], widely used for monitoring HF transmission lines and measuring reflection coefficient.



In the circuit on the left, the generator is typically a radio transmitter, and the impedance **Z** is an antenna system. The bridge combines a voltage sample, obtained from the capacitive potential divider C_1, C_2, R_V, and the output of a transformer in series with the line, which produces a voltage proportional to the current. The circuit parameters are calculated so that the bridge gives no output when the impedance **Z** is equal to the design load resistance of the transmitter. This balance condition can be maintained over a relatively wide frequency range by correct choice of the resistance R_v, which compensates for the falling reactance of the current-transformer secondary winding at low frequencies. R_v is typically a few kΩ. Thus, since the DC resistance of the current transformer secondary winding will be very small by comparison, it should be obvious that R_{port} is equal to R_v.

The output impedance at the generator frequency requires a little more consideration, but not much. Generally, the bridge is designed so that it abstracts only a very small amount of energy from the line. Thus the generator will hardly notice its presence, and its output will not droop significantly as a result of the additional loading. We can therefore assume that the output impedance of the generator is effectively zero. With that assumption; the output impedance of the voltage sampling network (by Thévenin's theorem) is given by the parallel combination of its resistance and reactances. The total output impedance is then obtained by placing the parallel combination of the current transformer load resistance and secondary reactance in series with the output impedance of the voltage sampling network. Thus[18]:

$$\mathbf{Z}_{out} = (R_V \,//\, \mathbf{j}X_{C1+C2}) + (R_i \,//\, \mathbf{j}X_{Li})$$

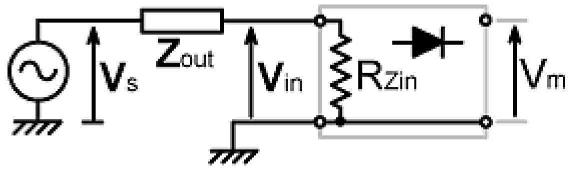
This can be converted to R+**j**X form using the standard parallel-to-series transformations[19].

17 **Directional apparatus for use with high-frequency transmission lines**. T Douma, US Pat. No. 2808566, 1957 (filed 1953).
18 The parallel impedance (//) operator is defined so that: **a** // **b** = **a b** / (**a** + **b**) , where both **a** and **b** can be complex.
19 See for example: **AC electrical theory**. D W Knight.. http://www.g3ynh.info/zdocs/AC_theory/ . Section 18.

1.6 Actual input voltage

Frequently, the RF design considerations for an item of test equipment lead to the relationship between some quantity to be measured and the open circuit (off-load) voltage of an output port. Since the port has a finite output impedance however, and the detector has a finite AC input resistance; the detector will not see that voltage. Instead, the detector input voltage is the output of a potential divider formed by $\mathbf{Z}_{out}$ and R_{Zin}. Thus, if $\mathbf{V}_S$ is the open-circuit source voltage:

$$\mathbf{V}_{in} = \frac{\mathbf{V}_S\ R_{Zin}}{(\mathbf{Z}_{out} + R_{Zin})}$$



The voltages and impedances in this expression are, of course, phasors; but since the detector does not conserve phase information, all phasors that appear as factors can be replaced by their magnitudes. Thus, if we use the convention that a voltage not written in bold is a magnitude (i.e.: $V = |\mathbf{V}|$):

$$V_{in} = \frac{V_S\ R_{Zin}}{|\mathbf{Z}_{out} + R_{Zin}|}$$

Now, if we have computed $\mathbf{Z}_{out}$ in the R+**j**X (series) form, so that (say):

$\mathbf{Z}_{out} = R_S + \mathbf{j}X_S$

then we can add the input resistance to it and work out the magnitude of the denominator in the expression for V_{in}. Thus:

$$V_{in} = \frac{V_S\ R_{Zin}}{\sqrt{(R_S + R_{Zin})^2 + X_S^2}}$$

This equation is primarily a statement of the obvious, in that if we want to avoid error due to source loading, it will be necessary to make R_{Zin} as large as possible and $|\mathbf{Z}_{out}|$ as small as possible. Such choices are not always available however; there being, for example, an issue when using a Douma bridge to drive a moving-coil panel meter, because the bridge has a large capacitive reactance component in its output impedance at low frequencies, and the meter will typically require 100 µA of drive for full-scale deflection (FSD). The only small consolation is that the source loading will be reduced at low drive levels, because of the increase in the detector input resistance (as discussed earlier). This unloading effect will partly offset the inherent non-linearity of the diode detector, it might even be used as the basis for partial linearity compensation if the source impedance is mainly resistive; but in general it will be frequency-dependent and hard to exploit to any advantage.

We have, in the preceding discussion, uncovered a number of potential sources of error affecting the use of the simple half-wave rectifier for absolute RF voltage measurement. They are, so far, quantifiable, and largely correctable; but they demonstrate that the behaviour of the detector is not quite as simple as many circuit designers seem to assume.

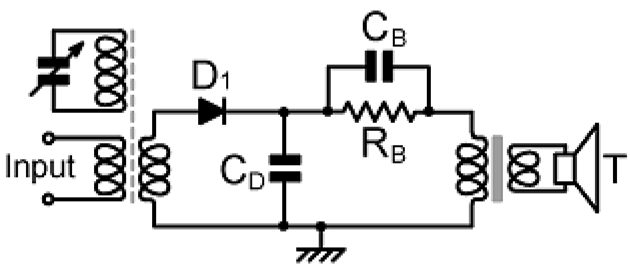
1.7 Inductively loaded AM detector

Although the principal subject of this article is AC voltage measurement; introducing the relationship between detector load and input impedance provides an opportunity to identify a widespread design error that occurs when a detector used as an AM demodulator is operated with an inductive load. This mistake typically occurs when a magnetic transducer, such as a telephone earpiece or a pair of headphones, is driven by a detector, either directly or through a step-down transformer. The issue is that the stated 'impedance' of the transducer is not the same as its DC resistance. Typically, the nominal impedance will be the impedance magnitude at some mid-band audio frequency, such as 1 kHz, whereas the DC resistance is likely to be smaller by a factor of between about 1/5 and 1/20.

Thus, for example, if we were to connect a pair of 10 kΩ high-impedance headphones to a detector, it might turn out that the DC resistance of the 'phones is about 800 Ω; in which case the large signal input impedance of the detector will be about 400 Ω instead of 5 kΩ. The effect, assuming a finite source impedance, will be to introduce excessive audio distortion by preventing the detector from moving out of its threshold region; and if the driving network is resonant, it will radically reduce the selectivity for strong signals by reducing the circuit Q.

The defect is present in most traditional crystal-set designs[20]. It results in a peculiar behaviour whereby there is relative silence between stations, but when a reasonably strong distant station is tuned in, it is heard mixed with a strong local station.

The solution is to place a parallel RC network in series with the inductive load, as shown in the diagram below. The network is known as a 'benny', after its inventor Ben Tongue[21]. Usual design practice is to use the resistor R_B to pad the transducer resistance to equal its nominal impedance value, thereby establishing the detector large-signal RF input impedance. The coupling capacitor C_B is then chosen to have a reactance magnitude that is small relative to the transducer impedance (say 1/10th) at the lowest required audio frequency.



Notice here incidentally, that the transducer T is shown with a step-down transformer, but a high-Z transducer could just as well be used without a transformer. Also, the tuned input transformer is shown with separate windings for the resonator and the detector, whereas they can be combined. This sidesteps the question of whether or not it is a good idea to tap the detector takeoff point down the tank coil to preserve selectivity. The answer is that it is not necessary to trade signal voltage for selectivity if a benny is used, because the detector input impedance, and hence the loaded Q of the filter, can be controlled by R_B.

---

20 See, for example, **The boy's book of crystal sets**, W J May, Bernards Radio Manuals 1954. http://www.rexresearch.com/xtlradio/boysbook.pdf (accessed 23rd Aug. 2014)

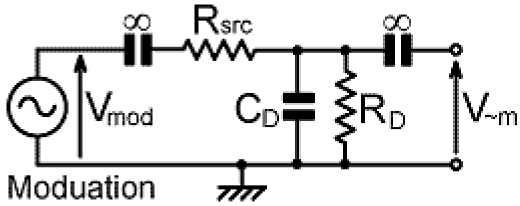
21 **Crystal radio set design**. Ben Tongue. http://bentongue.com/ (accessed 27th Dec. 2015) [Ben Hapgood Tongue, one of the founders of Blonder-Tongue Labs., died on 4th July 2015, at the age of 90]

1.8 Demodulated signal frequency response

Since the diode rectifier is the ubiquitous AM radio demodulator or 'envelope detector', it follows that if the signal generator is amplitude modulated (e.g., with a 1 kHz sine wave), then the indicator can be a transducer (earphone, etc.), or an audio amplifier with a loudspeaker. Audible output can be useful, for example, when nulling an RF bridge, because the ear can distinguish signal from noise, whereas a meter cannot.

The demodulated signal frequency response of a diode detector is dictated by the time constant, $C_D \times R_D$. We can show how this comes about by determining the frequency response function. The full derivation is strictly outside the realm of linear circuit analysis, but the result can nevertheless be deduced in a fairly straightforward way.

We start by representing the network as a simple potential divider, involving an as yet undefined source resistance, R_{src}. The equivalent circuit is shown on the right, where the inclusion of perfect coupling capacitors serves to remind us that we are dealing strictly with the AC component of the measurable output.



The relative frequency response is obtained by taking the magnitude of the ratio of the output at frequency f ($\mathbf{V}_{\sim m(f)}$ say) and the output at the frequency of maximum response, which, since we are dealing with a roll-off, we will call $V_{\sim m(0)}$. It is, of course, assumed that the process by which the signal was modulated has a completely flat frequency response. Starting with the output relative to the modulation component we have:

$$\frac{\mathbf{V}_{\sim m(f)}}{V_{mod}} = \frac{R_D // \mathbf{j} X_{CD}}{R_{src} + (R_D // \mathbf{j} X_{CD})}$$

where $\mathbf{V}_{\sim m(f)}$ is written in bold because its phase differs from that of V_{mod}. This can be rearranged:

$$\frac{\mathbf{V}_{\sim m(f)}}{V_{mod}} = \frac{1}{\left(\dfrac{R_{src}}{R_D // \mathbf{j} X_{CD}} + 1\right)} = \frac{1}{\left(\dfrac{R_{src}(R_D + \mathbf{j} X_{CD})}{\mathbf{j} R_D X_{CD}} + 1\right)} = \frac{1}{\left(\dfrac{R_{src}}{\mathbf{j} X_{CD}} + \dfrac{R_{src}}{R_D} + 1\right)}$$

At low modulation frequencies, where there is no roll off, the capacitive reactance $X_{CD} \rightarrow -\infty$ and we have:

$$\frac{V_{\sim m(0)}}{V_{mod}} = \frac{R_D}{R_{src} + R_D} = \frac{1}{\left(\dfrac{R_{src}}{R_D} + 1\right)}$$

Now taking the ratio:

$$\frac{\mathbf{V}_{\sim m(f)}}{V_{\sim m(0)}} = \frac{\left(\dfrac{R_{src}}{R_D} + 1\right)}{\left(\dfrac{R_{src}}{\mathbf{j} X_{CD}} + \dfrac{R_{src}}{R_D} + 1\right)} = \frac{1}{\left(\dfrac{R_{src} / \mathbf{j} X_{CD}}{(R_{src} + R_D)/R_D} + 1\right)} = \frac{1}{\left(\dfrac{R_{src} R_D}{\mathbf{j} X_{CD} (R_{src} + R_D)} + 1\right)}$$

Hence:

$$\frac{\mathbf{V}_{\sim m(f)}}{V_{\sim m(0)}} = \frac{1}{\left(\frac{R_{src}//R_D}{\mathbf{j}\,X_{CD}}+1\right)}$$

For the loudness response however, which is independent of phase, we want the magnitude, i.e., we square the real and imaginary parts separately and take the square root:

$$\left|\frac{\mathbf{V}_{\sim m(f)}}{V_{\sim m(0)}}\right| = \frac{1}{\sqrt{\left(\frac{R_{src}//R_D}{X_{CD}}\right)^2+1}}$$

where $R_{src}//R_D$ is the parallel combination of the detector load resistance and the source resistance.

So far so good, but we cannot use this relationship unless we know the source resistance. This is where linear analysis fails. The input signal to the detector has no component at the modulation frequency. Instead, the demodulated signal is created in the diode junction by mixing between the carrier and its sidebands. Since the smoothing capacitor C_D has a low impedance magnitude at radio frequencies, and the load resistance merely sets the DC operating point, the load impedance at the modulation frequency has no effect on the demodulation process, at least in first order. This means that R_{src} is effectively infinite. The demodulated signal therefore appears to originate from a current source, which can be thought of as a generator with a series resistance so large that the load impedance makes no difference to the current supplied. Thus, we put $R_{src}\rightarrow\infty$, and the relative voltage frequency response of the detector (also bearing in mind that $X_C = -1/2\pi fC$) becomes:

$$\left|\frac{\mathbf{V}_{\sim m(f)}}{V_{\sim m(0)}}\right| = \sqrt{\frac{1}{(2\pi f\, C_D R_D)^2+1}}$$

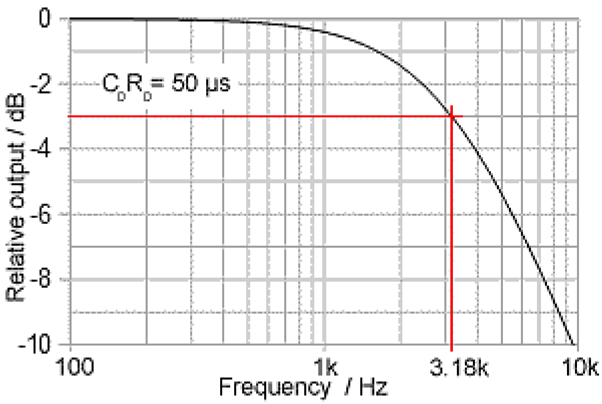
The upper demodulated-signal bandwidth limit, casually known as the 'bandwidth' and also variously known as the 'turnover frequency', the 'cutoff frequency' or the '-3 dB point', occurs when the output power has fallen to half the value it would have at low frequencies; i.e. (noting that the square of a voltage is proportional to power):

$$\left|\frac{\mathbf{V}^2_{\sim m(f)}}{V^2_{\sim m(0)}}\right| = \frac{1}{2} = \frac{1}{(2\pi f_{-3dB} C_D R_D)^2+1}$$

which, after rearrangement gives:

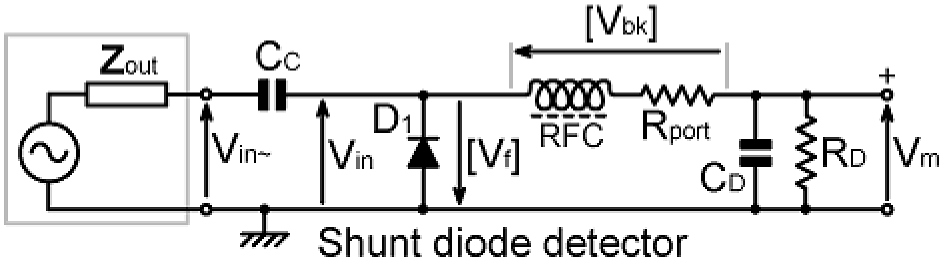
$f_{-3dB} = 1 / (2\pi C_D R_D)$

The log-log graph on the right shows the audio frequency response of a detector with a time constant of 50 µs. That gives the -3dB point at 3.18 kHz, which is a fairly typical choice for radio communications receivers and for systems using modulation signals in the vicinity of 1 kHz.



2. Shunt-diode rectifier

An alternative to the standard half-wave detector configuration is the shunt diode rectifier, shown below. This is the prototype of the indirectly-grounded detector circuits favoured by the Collins Radio Company in the 1950s, and popularised in SWR and impedance-measuring circuit construction articles by Collins design engineer Warren Bruene[22] and others[23].

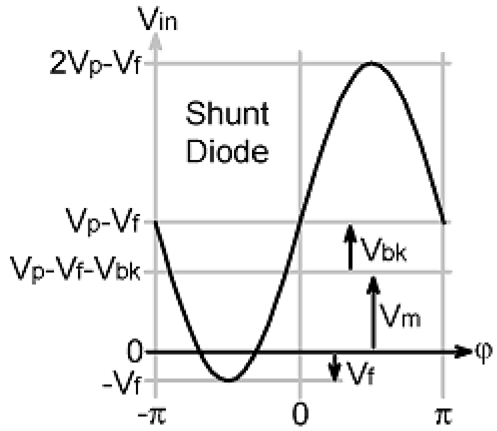


Note that a voltage placed in square brackets refers to the DC component of a waveform. As in the previous section, $V_{in\sim}$ represents the AC component of the input voltage, whereas the actual input, V_{in}, acquires a DC offset due to the rectifying action of the diode. As before also, we will define:

$V_{in\sim} = V_p \sin\varphi$

where V_p is the peak voltage, and $\varphi = 2\pi f\, t$, so that:

$V_p = V_{in\sim} \sqrt{2}$



The circuit appears, at first glance, to place a diode across the detector port; but it does no such thing because of the action of the coupling capacitor C_C. When $V_{in\sim}$ first goes negative (relative to the polarities shown on the circuit diagram), D_1 conducts and clamps one end of C_C nearly to zero, except for a shortfall due to the diode forward voltage V_f. This means that the most negative excursion of V_{in} is clamped to $-V_f$. Then, since the charge in C_C remains substantially constant on the timescale of the AC signal, the near-sinusoidal voltage at the cathode of D_1 averages about $+V_p - V_f$. The filter comprising the RF choke, R_{port} and C_D removes the AC component to give a measured voltage of $V_p - V_{bk} - V_f$, which is the same as for the ordinary half-wave detector. Overall, the main theoretical difference between the ordinary detector and the shunt diode detector is the order in which V_p, V_{bk} and V_f are added together.

Note that, as far as is possible, the nomenclature has been made consistent with that used for the ordinary detector. Thus R_{port}, although it is no longer part of the source network, has exactly the same effect on the DC output as does the DC resistance of the source network in the preceding section.

For the filter that removes the AC component, we have the choice of using a choke, a resistance, or a combination of the two (such as a realistic choke). As before, the use of a choke minimises V_{bk} but introduces the possibility of parasitic resonances, and so if some reduction in sensitivity is tolerable, the troublesome choke can be eliminated.

22 **An Inside Picture of Directional Wattmeters**, Warren B Bruene, QST, April 1959, p24-28
23 **In-Line RF Power Metering**, Doug DeMaw, QST, Dec. 1969, p11-16.

Input impedance
The AC input resistance of the shunt detector obeys the same 1:2 impedance transformation rule as the simple detector, i.e.;

$R_{Zin} = (R_D + R_{diode} + R_{port}) / 2$

The filter components however are effectively in parallel with that, and the reactance of the coupling capacitor is effectively in series; so that the total (neglecting parasitics) is:

$\mathbf{Z}_{in} = \mathbf{j}X_{CC} + (R_{port} + \mathbf{j}X_{RFC}) // [(R_D + R_{diode} + R_{port}) / 2]$

If the coupling capacitor is very large, and the choke (as is often the case) is omitted, this reduces to:

$Z_{in} = R_{Zin} = R_{port} // [(R_D + R_{diode} + R_{port}) / 2]$

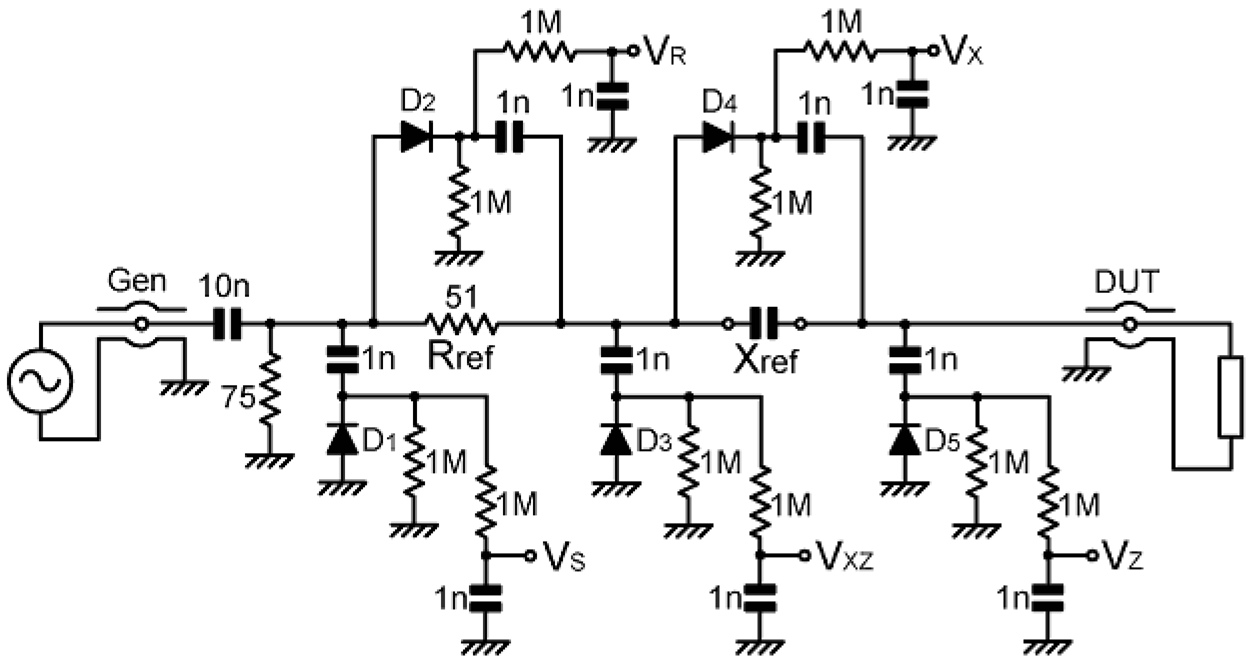
The coupling capacitor can also be physically omitted if the driving network is open-circuit to DC. If there is *any* DC path across the detector port however, C_C is essential; and if needed it should be chosen to have a reactance that is always small in comparison to R_{Zin}. If the lowest frequency of operation is say 1 MHz, and R_{Zin} is about 100 kΩ for large inputs, then the voltage drop across the capacitor might be considered to be negligible if its reactance magnitude is (say) always less than 1% of R_{Zin}, i.e., less than 1 kΩ at 1 MHz. Since $C = 1/2\pi f |X_c|$, we can then calculate that C_C should be at least 159 pF. It makes no practical difference if the capacitor is somewhat larger than the minimum required, and so 1 nF or greater would be a sensible choice. Note however, that the detector response will become sluggish if the capacitor is very large and the return path has a relatively high resistance, because it will take time for the circuit to reach equilibrium.

Diode PIV
Since the voltage at the cathode of the diode averages at roughly $+\{V_{in}\} \sqrt{2}$ (assuming that the diode polarity is chosen to give a positive output), the instantaneous peak inverse voltage sustained by the diode will be $\{V_{in}\}2\sqrt{2}$. Hence, as in the case of the simple half-wave rectifier, the V_{RM} for the diode must be at least $2\sqrt{2}$ times the maximum possible RMS input voltage.

Shunt diode rectifier: Diode $V_{RM} > V_{in}(\text{max}) \times 2\sqrt{2}$

2.1 Sampling floating voltages

One of the great advantages of the shunt-diode detector is its suitability for sampling voltages across network components that are floating with respect to ground. A practical example is given in the diagram below, which depicts a circuit for determining complex impedance (**Z** = R+**j**X) using only scalar voltage measurements. This technique was originally described by Heathkit design engineer Doyle Strandlund[24], but the implementation shown below is due to M E Gruchalla[25].



This circuit produces five DC voltages, which (assuming perfect linear rectification) are proportional to the magnitudes of the AC voltages across the source, the reference resistor R_{ref}, the reference reactance X_{ref}, the load **Z**, and the junction between the two reference elements and ground (the output voltages are all referred to ground). The full set of voltages is such that, even though the individual measurements contain no phase information, they can be combined in various ways to determine any of the impedance or admittance-related attributes of the load, **Z** (including the sign of the reactance or susceptance)[26].

The diodes D_1, D_3 and D_5 all have their anodes connected to ground, and so the corresponding detectors conform to the prototype circuit given earlier, except that they are intended to feed operational amplifiers having high input resistance and so use resistive filters and dispense with the chokes. Also notice that there is a 1 MΩ resistor across each of the diodes, the reasons for its presence in each case being firstly; that the detector must have a load if it is not to have an infinite time constant (and connecting it across the diode avoids loading the output); and secondly; that the op. amp. will require a few nA of input bias current, and it is a very bad idea to get that via the diode. Note that the voltages produced by these three detectors will be positive with respect to ground.

The detectors using the diodes D_2 and D_4 have the same design considerations as the others, except that they measure AC voltages that float relative to ground. This is possible because the

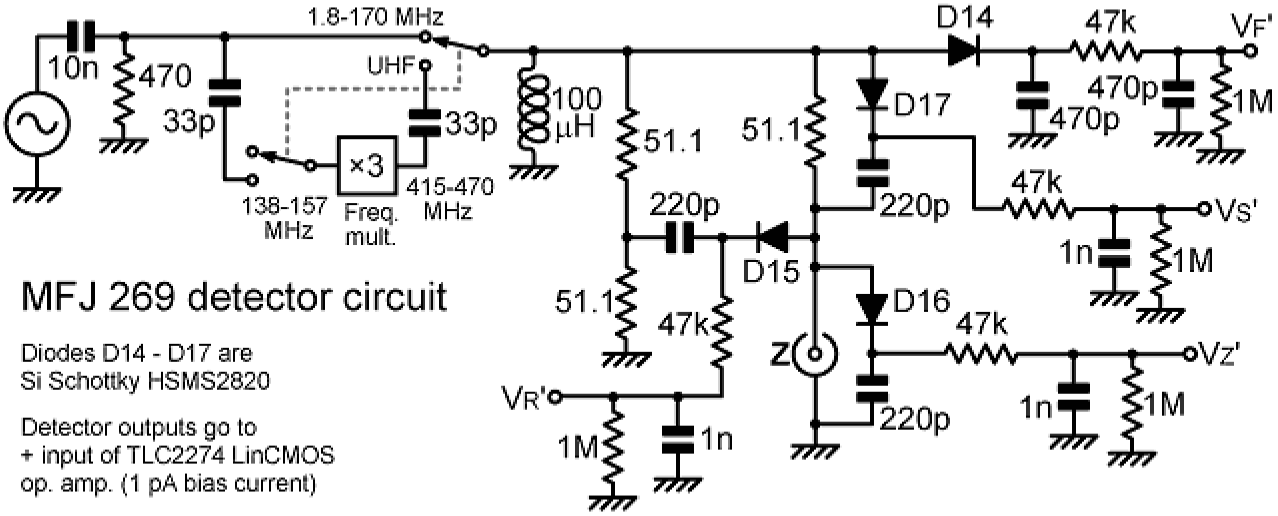
24 **Amateur measurement of R+jX**, Doyle Strandlund, QST, June 1965, p24-27. [An ingeneous graphical method for measuring complex impedance using only scalar measurements].

25 **Complex Impedance Measurement Using only Scalar Voltage Measurements**, M E Gruchalla, Communications Quarterly, Oct. 1998, p33-43. [Analytical approach to Strandlund's method and other improvements].

26 For a complete analysis, see: **Impedance and admittance measurement using scalar voltage samples**. D W Knight. http://www.g3ynh.info/zdocs/bridges/

measuring circuit is isolated from the outside world by coupling capacitors, which means that the AC signals are guaranteed to have no DC component. This allows the D_2 detector to get its ground reference via the 75 Ω input shunt resistor, while D_4 gets its reference via the 75 Ω resistor and the 51 Ω reference resistor in series. Since the detector currents are only of the order of a few µA, the anodes of D_2 and D_4 are effectively grounded from the DC point-of-view, and so the detectors give positive outputs referenced to ground.

Another example of the use of shunt-diode detectors is given in the circuit below. This shows the detector arrangement used in the MFJ269 antenna analyser[27]. A greatly simplified version of the driving arrangement is also included to show that the circuit has complete DC isolation from the RF source (and also, that it works at UHF).

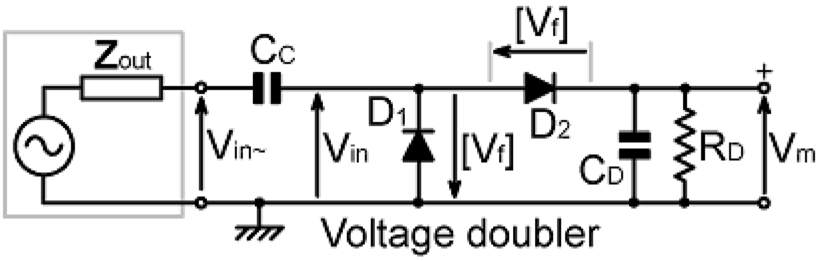


The MFJ engineers chose to use conventional half-wave detectors for AC voltages referenced to ground, and shunt detectors for the floating voltages. In this case, the filter resistor is 47 kΩ, about 1000× the resistance needed to balance the bridge, but also small enough to minimise the voltage drop caused by the 1 MΩ load resistors. The DC signals are applied to CMOS op. amps., which have an input leakage current of around 1 pA (somewhat less than might be caused by a fingerprint on the circuit board), and so there is no significant loading apart from the resistor. The MF269 uses static diode linearity compensation , and 12-bit A-D conversion; and the result is a remarkably accurate instrument, able to measure impedance-related quantities within an order-of-magnitude of 50 Ω to better than ±2%.

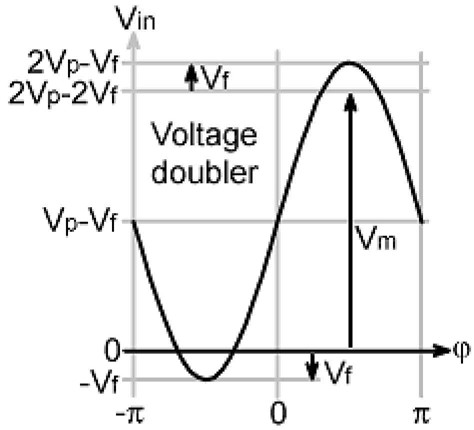
27 http://www.mfjenterprises.com/

3. Voltage-doubler rectifier

The circuit shown below[28] is obtained by the simple expedient of replacing the filter components of the shunt-diode detector with a second diode. The effect, at least nominally, is to double the voltage output in comparison to a half-wave rectifier, to eliminate the port resistance and hence the backoff voltage, and to remove the need for an RF choke. The circuit can also be used for the detection of floating AC voltages provided that the conditions outlined in **section 2.1** are met.. Such are its advantages that we might wonder why it is not universally adopted; but it turns out that its principal strength is also its principal drawback in voltage-sampling applications. The issue is that its effectiveness at converting AC input into DC output means that it has a surprisingly low input impedance.



As before, we will define $V_p = V_{in\sim}\sqrt{2}$, where $V_{in\sim}$ is the AC component of the input voltage. The operation of the circuit is then as follows: When V_{in} first goes negative, D_1 conducts and clamps one end of the coupling capacitor C_C to $-V_f$. This causes the capacitor to be charged to $V_p - V_f$. The charge in C_C then remains substantially constant on the timescale of the AC signal, and so the voltage across C_C is placed in series with V_{in}, ultimately causing the smoothing capacitor C_D to be charged, via D_2, to twice the peak AC input minus two diode forward voltages. Note that since the diodes operate alternately, this detector conducts on both positive and negative half-cycles of V_{in}, which means that it is actually a type of full-wave rectifier. Also notice that the detector has an output shortfall of two diode forward voltages; but since it also give twice the output of a half-wave detector, the overall degree of non-linearity is (at least nominally) the same as for the half-wave detector .



Diode PIV

Due to the clamping action of D_1, the voltage at the anode of D_2 cannot fall substantially below zero. Hence, since the voltage across the smoothing capacitor C_D is very nearly $+V_{in}2\sqrt{2}$, V_{RM} for D_2 must be at least $2\sqrt{2}$ times the maximum possible RMS value of V_{in}. In the case of D_1, since its effect is to clamp the most negative excursion of the input voltage approximately to ground, thereby charging C_C to $V_{in}\sqrt{2}$, the voltage across C_C will be placed in series with $V_{in\sim}$ when it makes its positive excursion. Hence V_{RM} for D_1 must also be at least $2\sqrt{2}$ times the maximum possible RMS value of V_{in}.

Voltage doubler: Both diodes, $V_{RM} > V_{in}(\max) \times 2\sqrt{2}$

28 A floating input bridge configuration, known as the 'Latour doubler', is also possible. See, for example: '**High Efficiency Low Power Rectifier Design using Zero Bias Schottky Diodes**', A Mabrouki, M Latrach, V Lorrain. 2015. https://hal.archives-ouvertes.fr/hal-01131545, also, '**Full-range nonlinear model of Latour's detector** . . .', Xavier Le Polozec, cited earlier.

3.1 Doubler input impedance

In order to determine the input impedance of the voltage doubler, we can use an argument similar to that given in **section 1.4**.

Provided that the loading on the source is not so great as to cause substantial harmonic distortion, most of the harmonic energy generated by the rectification process will be dissipated in the dynamic component of the diode forward voltage (**section 12**). In that case, the DC power dissipated in the network will be the same as the AC input power. The total DC voltage produced is:

$$2V_p = 2\, V_{in\sim} \sqrt{2}$$

and the total DC resistance is:

$$R_{tot} = R_D + 2R_{diode}$$

Hence:

$$P_{in} = (V_{in\sim}\, 2\sqrt{2}\,)^2 / R_{tot} = 8\, V_{in\sim}{}^2 / R_{tot}$$

but, by definition:

$$P_{in} = V_{in\sim}{}^2 / R_{Zin}$$

Hence:

$$R_{Zin} = R_{tot} / 8$$

and for large signals, when $R_D >> R_{diode}$,

$$R_{Zin} \approx R_D / 8$$

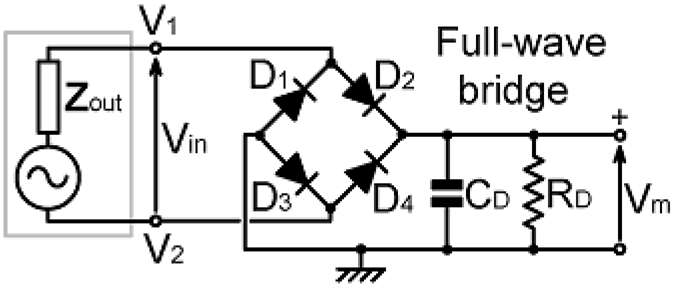
Thus, whereas the half-wave detector is a 1:2 impedance transformer, the voltage doubler is a 1:8 impedance transformer.

The disadvantage of the voltage doubler[29] for non-invasive voltage-measurement purposes can now be demonstrated by a simple example. Suppose we decide to use a voltage doubler to drive a 100 µA moving-coil meter padded to 100 kΩ to give 10V FSD. Neglecting diode resistance and parasitics, the detector input impedance will correspond to a resistance of about 12.5 kΩ. If we were to use a half-wave detector however, we would get half the output voltage, but we could easily reduce the meter padding resistance to 50 kΩ (including R_{port}) and thereby increase the sensitivity to 5 V FSD. This alternative half-wave detector will now give exactly the same meter readings as the doubler, except that its large-signal input impedance will be about 25 kΩ. We have not, incidentally, found a source of free energy by using a non-multiplying detector. We have merely avoided pumping the smoothing capacitor up to a high voltage, only to drop the voltage back down again using a resistor.

29 See also: **Crystal radio set design technical help**, Ben Tongue, http://bentongue.com/xtalset/0def_exp/0def_exp.html section 5A.

4. Bridge rectifier

A non-multiplying detector that requires no RF chokes, does not have a backoff voltage, and does not care whether or not there is a DC path through the source network, is the full-wave bridge, shown below:



This circuit is nearly universal in modern power supplies, but it has two disadvantages in signal detection applications; one being that it has two diode forward-voltages in the path to the load, and the other being that either the output or the input terminals must be allowed to float with respect to the system ground.

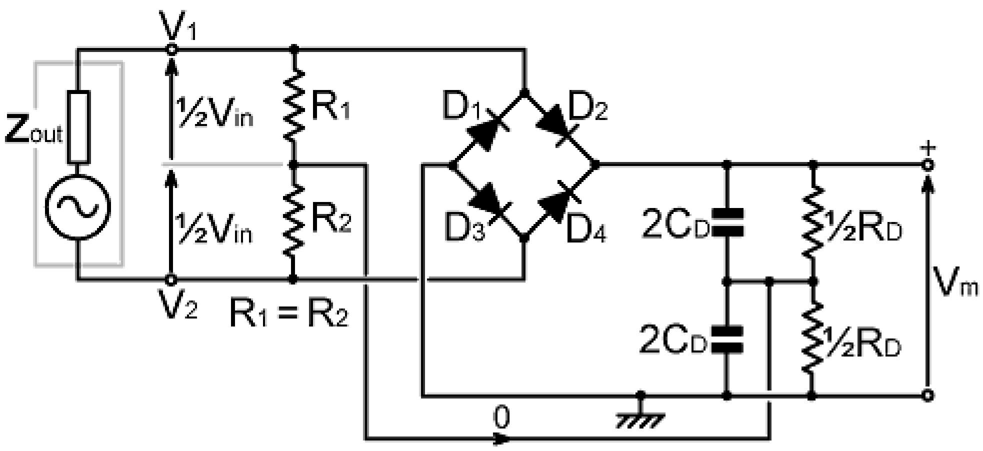
Despite the caveats, one advantage of the bridge rectifier is that the maximum inverse voltage for any of the diodes is only $\sqrt{2}$ times the RMS input voltage. The magnitude of the inverse voltage across D_2 is is prevented from rising above $V_{in}\sqrt{2}$ by the clamping action of D_1 and vice versa. The same argument applies for D_3 and D_4.

Bridge rectifier: All diodes, $V_{RM} > V_{in}(max) \times \sqrt{2}$

Principle of operation

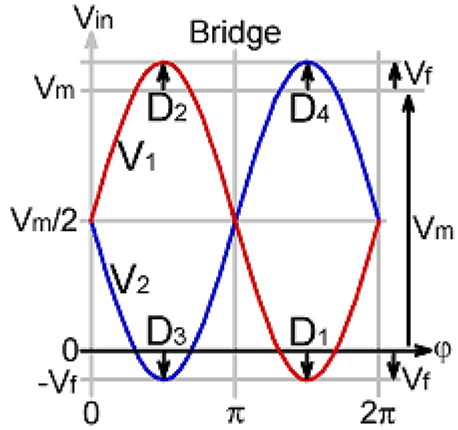
Although the bridge rectifier is very familiar, its mode of operation is surprisingly difficult to understand. This can be made apparent by asking the question: 'If the circuit above is grounded as shown, what voltage waveforms, measured relative to ground, will appear at points V_1 and V_2 ?' If you have not previously needed to solve this problem, it will probably take you a while to work out the answer.

A simple way to find the solution is to imagine adding centre taps to both the input network and the load network, as shown in the diagram below.



It is not really necessary to tap into the smoothing capacitor as well as the load resistor, but if we do so the circuit becomes reminiscent of the familiar dual-rail power supply[30]. All we have to do to turn it into the dual-rail supply is move the earth to the centre tap and feed it with a centre-tapped transformer instead of a resistive potential divider. Thus it should be obvious that if the two half-outputs are equally loaded (which they are), there will be no current flowing along the connection between the input and output centre taps. This means that, with the negative terminal of the output grounded, the alternating voltages V_1 and V_2 must be symmetric about a DC offset at half the output voltage. So now, if we let:

$V_{in} = V_p \sin\varphi$



Then:

$V_1 = (V_m + V_p \sin\varphi) / 2$

and

$V_2 = (V_m - V_p \sin\varphi) / 2$

Note that these two voltages are identical apart from a 180° phase difference. Also, the AC component in each case is half the input voltage.

Now compare the graph of the operation cycle given above with the corresponding graph for the voltage doubler in **section 3**. It appears that the bridge circuit splits the input voltage into two halves in antiphase. These two half-voltages are then applied simultaneously to a pair of voltage doublers. Since there is no overall voltage multiplication however, the input impedance is:

$R_{Zin} = (R_D + 2R_{diode}) / 2$

Ripple frequency

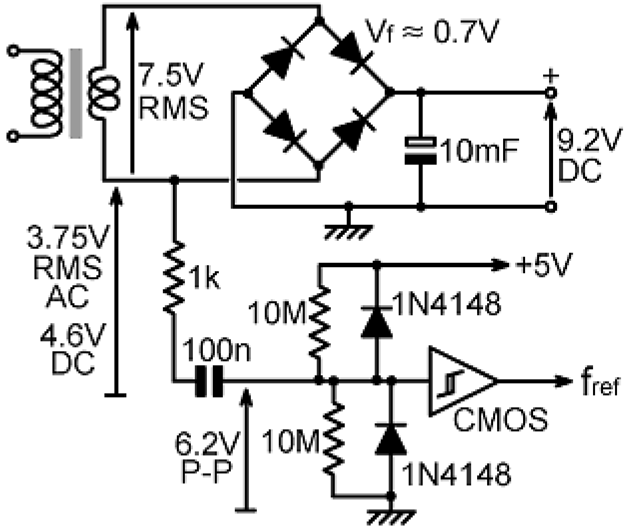
The rectifiers described in **sections 1 - 3** all top-up the smoothing capacitor once per cycle. The bridge rectifier however, tops up the capacitor twice per cycle, which means that it doubles the ripple frequency. Also, if the circuit is used without smoothing, it becomes a highly efficient frequency doubler, removing the fundamental frequency-component completely (assuming perfect symmetry). The frequency-doubled output is not sinusoidal, of course, but it can be filtered.

The frequency doubling effect is useful in mains-frequency power-supply applications because it reduces the required smoothing capacitance for a given level of ripple. It is however, usually of little importance in radio-frequency applications because smoothing capacitors are, in any case, small.

The bi-phase (half-bridge) rectifier circuit discussed in **section 5** also doubles the ripple frequency; but note that the voltage doubler (**section 3**) does not, even though it is a type of full-wave rectifier.

30 It turns into a pair of bi-phase rectifier power supplies of opposite polarity. See **section 5**.

Frequency sampling
The fact that the input terminals of a bridge rectifier, when measured relative to ground, carry an AC signal of half the floating input voltage can be exploited for the purpose of sampling the input frequency. This is the traditional method for extracting the power-line frequency from mains transformers without using a separate winding, and is used to provide a simple and accurate timing reference in devices such as electronic clocks.

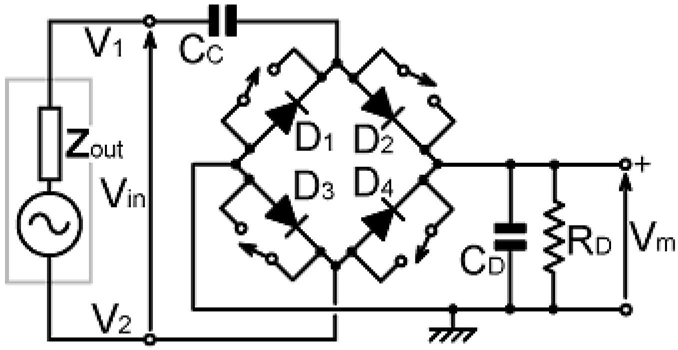


A power-line sampling circuit is shown on the right. The forward voltage for silicon P-N power diodes is about 0.7 V, so the power-supply produces a raw DC output of $7.5\sqrt{2} - 2\times0.7 = 9.2$ V, which is intended for regulation to 5 V for powering logic circuitry. Thus the voltage to ground on either side of the AC input to the rectifier is 3.75 V RMS, with a DC offset of +4.6 V. This corresponds to a peak-to-peak excursion of $3.75 \times 2\sqrt{2} = 10.6$ V, but it is fed through a resistor and a coupling capacitor to a peak clipper comprising two 1N4148 silicon P-N signal diodes. These have a forward voltage of about 0.6 V, and so the resulting waveform is limited to lie between 0.6 V above the +5 V rail and 0.6 V below ground, i.e., it is clipped to 6.2 V P-P. This is converted into a clean 5 V P-P square wave by means of a CMOS Schmidt trigger.

It should be understood, of course, that any substantial loading of the voltage at the AC input side of a bridge will upset the symmetry. For a circuit designed to produce hundreds of milli-amps, or even amps, of output current however, a sampling network such as this will make no measurable difference. Note that the 100 nF capacitor has a reactance of -32 kΩ at 50 Hz (-27 kΩ at 60 Hz), and although there will be harmonics present, it constitutes a very weak coupling. The 1 kΩ resistor gives the diodes and the Schmidt trigger some protection from any fast transients that might appear on the power-line.

Coupling capacitor

Because the bridge rectifier introduces no net DC component into the current flowing in the source network, it can be fed via a coupling capacitor (or capacitors) if so desired. In power supplies however, the use of a coupling capacitor can be dangerous unless the implications of so doing are considered in advance.



In the thoroughly nasty circuit shown on the right, a bridge is fed via a coupling capacitor, and four switches are provided for the purpose of simulating a short-circuit (which is the usual failure mode) in any one of the four diodes.

Let us first consider what happens when diode D_3 is shorted. In that case, V_2 is pulled to ground, D_1 and D_2 begin to act as a voltage doubler, and D_4 is harmlessly reverse-biased (it is assumed, incidentally, that none of the diodes have their V_{RM} ratings exceeded when the event occurs). From an external point of view, the effect is approximately to divide the input impedance by a factor of 4 (assuming a purely resistive load), and to double the output voltage. Shorting-out D_1 will, of course, have a similar effect.

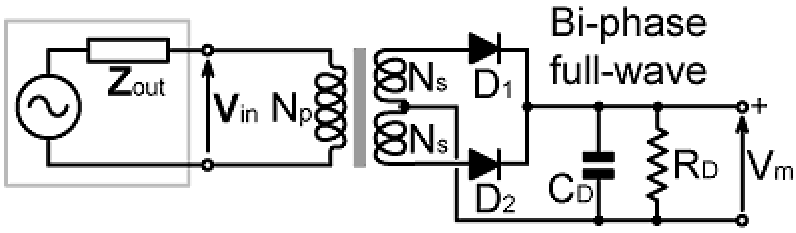
Now consider the effect of shorting-out D_2. In that case, V_1 is tied to the positive output terminal, D_3 and D_4 act as a voltage doubler producing an output that is negative relative to the positive terminal, and D_1 is permanently reverse biased. Shorting out D_4 will have a similar effect.

Thus, when the bridge is capacitively coupled, there are four ways in which the most common type of diode failure can double the output voltage and quarter the input impedance. A fuse is generally too crude a safety device to detect such an impedance change, and circuitry subjected to twice its intended operating voltage can malfunction in numerous unpleasant ways. Such might be the consequence of failing to consider this idiosyncrasy. On the other hand, a voltage doubler circuit can be provided with a half-output (×½) facility by the addition of two more diodes an a single-pole switch.

5. Bi-phase rectifier

The bi-phase (or 'half-bridge') rectifier was once widely used in the power supplies of valve (tube) equipment. As a power rectifier, it is a poor choice in comparison to a full bridge; and after selenium HV rectifiers[31] became available in 1933, its chief design merit seems to have been that it maximised the profits for owners of light-bulb factories while increasing the size and complexity of the mains transformer at the customer's expense[32 33]. The circuit creates a serious fire hazard in the event that one of the rectifiers should fail open circuit[34], because a DC component will then be present in the transformer secondary current, causing core saturation and overheating.

In signal detection applications however, the circuit has the advantage of being a full-wave rectifier with only a single diode forward-voltage drop in the path to the load. By conducting on both half-cycles of V_{in}, the circuit also halves the average diode current for a given output voltage, and so reduces the diode forward voltage drop compared to that of a half-wave rectifier. Because of the logarithmic diode V-I characteristic however, the V_f reduction is only a few %.



A disadvantage of the circuit that it usually requires a centre-tapped coupling transformer (although it can be constructed as a pair of half-wave rectifiers driven by a dual-ended (balanced output) amplifier). The output voltage of the transformer-coupled version is given approximately by the following expression (which neglects transformer losses):

$$V_m = \sqrt{2}\, V_{in}\, N_s / N_p - V_f$$

Where N_s / N_p is the turns ratio. Correct operation requires that the transformer has sufficient primary inductance not to load the driving network significantly at the lowest frequency of operation, and that the transformer is otherwise working within its pass-band. For HF radio frequency applications, the transformer can be wound on a small ferrite toroid or two-hole ('binocular') core. Note incidentally, that if a transformer is used for the ratio arms (i.e., the voltage-dividing network) of a measuring bridge, it is possible to combine this transformer with the bi-phase detector transformer by using the fact that the bridge is a linear reciprocal network (i.e., generator and detector ports are interchangeable). The resulting configuration, a type of

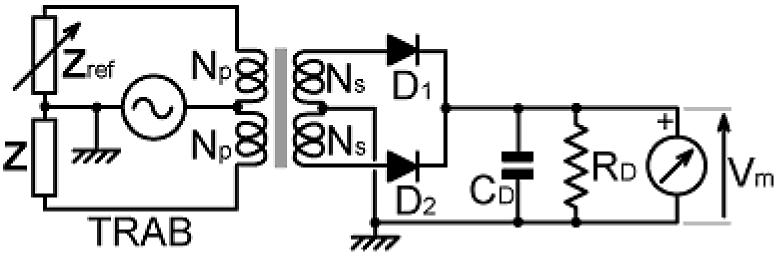
31 Selenium high-voltage rectifiers (see: http://en.wikipedia.org/wiki/Selenium_rectifier) were considerably more efficient than valves and require no heater supply, but equipment manufacturers associated with the valve cartels did not use them. The slow adoption of semiconductors in some quarters is further discussed in **section 11**.

32 As pointed out by Bob Weaver (private e-mail, 14th Aug 2014), the weight of copper required is not greatly increased when a bi-phase is used instead of a bridge, because the two secondary windings provide half the current each. The manufacturing time and the amount of insulating material however, will be increased, and the valve rectifier (in safe design practice at least) requires an additional floating heater winding.

33 The marketing of radio receivers in the 1930s, particularly in the USA, led to tube wars. Using more tubes was supposed to make a better radio set. Thermionic rectifiers were counted as tubes, even though they were nothing to do with the signal chain, but the semiconductor rectifier was not so easily turned into a marketable concept.

34 Silicon rectifiers usually fail short-circuit, which is safe with an adequately specified transformer because it will blow the primary-side fuse. Valve rectifiers come in two types; single cathode dual anode, which cannot fail with one diode open circuit; and the more usual dual diode with dual filaments in parallel, which is a fire hazard and should be replaced by a solid-state rectifier and a soft-start circuit if the equipment is to be re-certified for use. It is a good idea to replace or re-form the smoothing capacitors when changing to solid state rectification (an increase in output voltage occurs).

transformer ratio-arm bridge (TRAB) is shown below:



This circuit allows the unknown and reference impedances, the generator, and the output network, all to have one terminal connected to ground. The transformer will introduce some minor voltage shortfall, due to losses and leakage inductance; but it will not introduce non-linearity unless it is operated close to core saturation, which is highly unlikely given the trivial power requirement of a signal detection circuit. A particular disadvantage of the configuration however, is that it involves stuffing the ratio-arm transformer with a relatively large amount of wire, which might make it difficult to minimise stray capacitance between the primary and secondary windings. The ratio-arm transformer must be designed with balance as the principal criterion, and it is not necessarily a good idea to try to make it perform additional functions.

With regard to diode V_{RM}, note that the bi-phase rectifier is simply two half-wave rectifiers feeding the same smoothing capacitor. Hence the inverse voltage rating for a diode must be at least $2\sqrt{2}$ times the RMS voltage at the transformer output.

Note that the bi-phase circuit is a frequency-doubling rectifier, which can be either an advantage or a disadvantage in RF applications.

As mentioned before, the bi-phase rectifier gives an improvement in linearity over the half-wave rectifier by dividing the rectifier current between two diodes; but as will be discussed in the next and later sections, the diode forward voltage drop varies logarithmically with current in such a way that the improvement will not be particularly large, and so the additional complexity might not be warranted.

6. Diode static voltage vs. current characteristic

All of the diode detectors described in the preceding sections have at least one diode forward-voltage drop in the path to the measuring device. Under static conditions (i.e., when a DC bias is applied) the voltage across the diode (V_d) varies approximately in proportion to the logarithm of the current (I_d) passing through it. The current also depends on the temperature of the diode junction. For many types of diode, this behaviour is described to a good approximation by a modified form of the Shockley diode equation[35 36 37 38].

$$V_d = I_d R_{ds} + m (k_B T/q_e) \ln[(I_d / I_S) +1]$$

where:

I_d is the static or instantaneous forward current.

R_{ds} is the ordinary ohmic series resistance of the diode.

I_S is the junction reverse saturation leakage current (see **section 6.1**). Note that this quantity varies with temperature independently of the thermal voltage, V_T (discussed below).

m is a dimensionless correction factor between 1 and 2 (known as the 'ideality factor' or 'emission coefficient')[39].

ln is the natural logarithm operator (Log_e).

k_B is Boltzmann's constant ($k_B = 1.380662\times10^{-23}$ joules/kelvin).

q_e is the charge of an electron ($q_e = 1.6021892\times10^{-19}$ coulomb).

T is the temperature in Kelvin, K, i.e., the absolute temperature (°C +273.16).

The factor $k_B T/q_e$ has dimensions of volts and is often given the symbol V_T (the thermal voltage).

$V_T = k_B T/q_e = 25.3$ mV at 20°C.

The modified Shockley equation is then more conveniently written:

$V_d = I_d R_{ds} + m V_T \ln[(I_d / I_S) +1]$	**6.1**

When $m = 1$ and $R_{ds} = 0$, the diode forward characteristic reduces to the Shockley ideal diode equation:

35 **The theory of p-n junctions in semiconductors and p-n junction transistors**. W Shockley. Bell System Tech. Journal 28(3), 1949, p435-489. See p454, equation 3.13 [Ideal diode equation with ohmic resistance].
36 **Carrier generation and recombination in P-N junctions and P-N junction characteristics**, C T Sah, R N Noyce, W Shockley, Proc. IRE, Sept. 1957, p1228 - 1243. [Deviation from ideality].
37 **Physics of semiconductor devices**. S M Sze. Wiley, 1969. SBN 471 84290 7 [2nd edition 1981 and 3rd edition 2006 also exist]. Chapter 3, part 4 (p96-102). Current-voltage characteristics of PN junction diodes.
38 **The art of electronics**, P Horowitz and W Hill. 2nd edition 1989, Cambridge Univ. Press. ISBN 0-521-37095-7. Ebers-Moll model for transistors. p79-80.
39 Most of the original articles use n for the emission coefficient, but here we will reserve n for other purposes.

$V_d = (k_B T/q_e) \ln[(I_d / I_S) +1]$

which, using the notation $\exp(x) = e^x$, is more usually written:

$I_d = I_S [\exp(V_d q_e / k_B T) - 1]$

i.e.,

$I_d = I_S [\exp(V_d / V_T) - 1]$

The Shockley equation was derived for a P-N junction, but also applies to point-contact and semiconductor-on-metal (Schottky or 'hot carrier' [40]) diodes when I_S and m are chosen appropriately.

For an overview of the theory of metal-semiconductor junctions, see the review by Rhoderick[41]. He points out some limitations of the near-ideal diode equation close to the threshold of forward conduction, particularly in that it should strictly take the form:

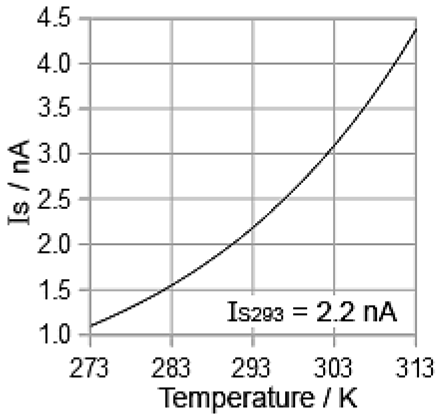
$I_d = I_S \exp(V_d / mV_T) \{ 1 - \exp(- V_d / V_T) \}$

This does not quite turn into the accepted form of the modified Shockley equation because there is no emission coefficient (m) in the second term. It makes no difference for large diode currents, but it does imply a small finite conduction threshold.

Bear in mind that the effective diode forward voltage drop under dynamic (i.e. AC) conditions, is greater than the static voltage drop. That issue is examined in **section 12**. The static characteristic is nevertheless an indicator of dynamic behaviour, and so governs the choice of diode. Also note that the diode equation as given does not describe the reverse-breakdown (avalanche) region of the characteristic; i.e., the diode is assumed to be operating within its specified V_{RM} limit.

6.1 Variation of saturable leakage current with temperature

The reverse saturation leakage current of a diode, I_S , doubles for about every 20 K (20ºC) of temperature rise[42]. A simulation is shown on the right (see spreadsheet **det_models.ods** sheet 7). This represents substantial variability, which will need to be taken into account when the diode forward voltage is significant in comparison to the peak signal voltage.



If a dependent quantity doubles for every application of a particular constant increment of an independent variable, then that behaviour constitutes exponential growth. Consequently, if we have a measurement of I_S at a particular temperature, we can find its value at some other nearby temperature by noting that the logarithm of I_S divided by its value at the original temperature will make a straight line graph. Thus, if we have a

40 **Using the hot carrier diode as a detector**. M Crane, H O Sorensen, HP Journal. 17(4) Dec. 1965.
41 **Metal-semiconductor contacts**. E H Rhoderick, IEE Proc, Vol 129, pt. 1, No. 1. Feb. 1982. p1-14.
42 **Characteristics and applications of diode detectors**. Ron Pratt. HP RF & Microwave Symposium, ≥1978. http://www.hparchive.com/seminar_notes/Pratt_Diode_detectors.pdf

measurement at (say) 293.16 K (i.e., 20ºC), and we call I_S at this temperature I_{S293}, then we can write:

$\ln(I_S / I_{S293}) = aT + b$ . . . . . (**6.2**)

where T is the temperature, a is the gradient of the line, and b sets the point at which the line crosses the axis. Now note that when $I_S = I_{S293}$, $\ln(I_S / I_{S293}) = 0$. Thus:

$0 = 293a + b$

i.e.:

$b = -293a$

Also note that if the saturation leakage current doubles for a 20 K rise in temperature, then:

$\ln(2) = 313a + b = 313a - 293a = 20a$

i.e.:

$a = \ln(2) / 20$

Feeding the expressions for a and b back into equation (**6.2**) gives:

$\ln(I_S / I_{S293}) = T [\ln(2) / 20] - 293 [\ln(2) / 20]$

i.e.:

$\ln(I_S / I_{S293}) = \ln(2) (T - 293) / 20$

Exponentiating both sides and rearranging then gives:

$I_S = I_{S293} \exp\{ \ln(2) (T - 293) / 20 \}$	**6.3**

Note that this expression contains the assumption that I_S doubles for a 20 K temperature rise. It will suffice if no other information is available[43]; but if two measurements of I_S are made at different temperatures (T_0 and T_1 say), then a more accurate form can be written:

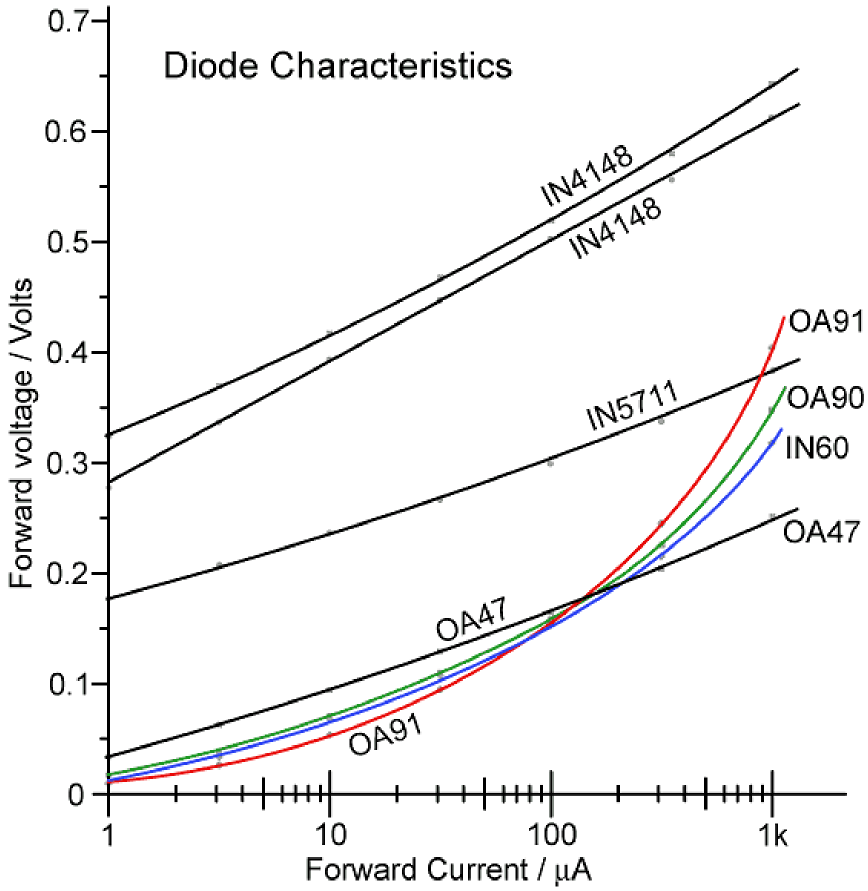
$I_S = I_{ST0} \exp\{ \ln(I_{ST1} / I_{ST0}) (T - T_0) / (T_1 - T_0) \}$	**6.4**

The large temperature variation of the diode characteristic, although quantifiable, suggests that detector diodes used for precision voltage measurement applications should ideally be operated in a temperature controlled environment. The detector, for example, might be placed inside a small temperature regulated chamber, similar to (or actually) a crystal oven.

43 **'Nonsquarelaw behavious of diode detectors** . . ', [Harrison - Le Polozec 1994], cited earlier. See equation (2).

6.3 Diode measurements

The graph below shows the measured forward voltage vs. current characteristics of a variety of small signal diodes.



The 1N4148 is a silicon P-N junction diode. The 1N5711 (Agilent 5082-2800) is a high-inverse silicon Schottky diode (i.e., silicon-metal junction diode). The OA47 is an archaic germanium gold-bonded diode, and the rest are germanium point-contact diodes. Two 1N4148 diodes from different manufacturers were measured merely to illustrate the point that silicon P-N diodes have the highest forward voltage and are therefore a poor choice for low voltage detectors. The 1N5711 curve is the average of results from four diodes (all from the same batch) that had practically identical characteristics. The OA47 curve is the average for four diodes from two manufacturers, all having similar characteristics. The OA90, OA91, and 1N60 curves are from single examples, and are therefore not necessarily representative of the type. All data were recorded at an ambient temperature of 21°C.

The data indicate that the IN4148, the 1N5711, and the OA47, all obey a logarithmic V/I relationship reasonably well, whereas the germanium point-contact characteristics show considerable curvature due to high ohmic resistance. With regard to the forward voltage drop however, the germanium diodes are all superior to the IN5711 in the 1 to 100 µA range, and the preference for the latter may merely reflect the fact that many semiconductor manufacturers no longer fabricate germanium. Silicon Schottky diodes, such as the 1N5711 and 1N6263 have better high-frequency performance than germanium diodes, but germanium diodes work well at VHF and are therefore often adequate for HF applications. Among the germanium diodes, there is little difference between the gold-bonded and standard varieties in the 1 µA to 100 µA range, but the OA47 is the best choice for currents up to 1 mA. We should observe however, that detector diodes only conduct significantly on the peak of the applied waveform, and so the instantaneous current is much higher than the average current, the difference being about an order of magnitude (see section **12**). Therefore, in selecting diodes for average currents in the region of 1 µA to 100 µA, we should consider the steady-state voltage drop in the region 10 µA to 1 mA; in which case the germanium gold-bonded diode offers the lowest forward-drop among the diodes compared. Note however, that one of the consequences of the diode equation is that low forward voltage-drop is associated with high reverse saturation leakage current. Also, there will be a non-saturable reverse leakage current, roughly equivalent to a resistance in parallel with the diode, and germanium diodes are worse than modern silicon diodes in that respect. Hence, if reverse leakage is an issue, high-inverse silicon Schottky diodes are to be preferred. To put this matter in perspective however, the reverse leakage current of an OA47 diode was measured as follows:

Vr / Volts	1.0	2.0	3.0	5.0	10.0	15.0	20.0
Ir / µA	1.1	1.3	1.4	2.1	4.1	6.3	8.8

T = 21°C

The leakage current is approximately linear in the 5 V - 20 V range and can be modelled by assuming a parallel resistance of about 2 MΩ. Such a defect has little effect on the operation of a detector loaded with a 10 kΩ - 100 kΩ resistance, but will be deleterious in the high-impedance detector circuits discussed earlier. Some users of Ge diodes in high-impedance detectors actually use them without a DC return path (using the diode leakage for that purpose); but such circuits are unpredictably dependent on the characteristics of the particular diode used.

Some final points in favour of the silicon Schottky diodes are that germanium diodes show a wider spread of characteristics, and that the OA47 and the later AAY-series Ge-Au diodes are obsolete. Hence the Si Schottky diodes are definitely preferable in applications requiring diode matching, precise calibration, or availability through normal commercial channels. The 1N5711 in particular also, has a high reverse breakdown voltage for a silicon device of its class, its V_{RM} of 70 V making it suitable for half-wave detectors of up to 24.7 V DC output. An OA47 half-wave detector has a maximum DC output of 10.6 V if V_{RM} is not to be exceeded.

Note that the silicon Schottky diode can only achieve a high peak inverse voltage if it is fabricated using a guard-ring structure[44]. The guard-ring prevents breakdown due to the high electric field strengths that occur at the edge of the metallised area. The downside however is that it increases the junction capacitance (2 pF for a 1N5711), and it creates a spurious P-N junction that can conduct slightly at very high forward currents.

For small signal applications, simple Schottky diodes (e.g., Agilent HSMS-2850) are to be preferred. Without the guard-ring, the junction capacitance is about 0.3 pF. V_{RM} is then however in the 2 V to 6 V range[45].

44 see: http://www.radio-electronics.com/info/data/semicond/schottky_diode/technology-structure-operation.php Accessed 22nd Dec. 2015.

45 Chin-leong Lim 9W2LC, Private e-mail correspondence, 4th - 5th March. 2015. Agilent HSMS-285x series data sheet.

7. Signal diode data

The table below shows some collected diode data. The information is mostly historical, but the point is to be aware of the typical forward voltages, reverse breakdown voltages and reverse leakage currents associated with the various types. Also note that Ge diodes are nowadays usually NOS or salvaged, and so old spec. sheets are not necessarily redundant.

For small-signal RF rectifier voltmeters, diodes should be chosen for low forward voltage drop and low junction capacitance (which is given in brackets where known). An example Pico-amp diode (PAD) is also included. These are not used as detectors, but are connected anode to cathode across diodes placed in op-amp feedback loops, to prevent wild excursions in the wrong output polarity range.

<table>
<tr><th rowspan="2">Type</th><th rowspan="2">Description</th><th rowspan="2">$V_{r\,max}$ / V</th><th rowspan="2">$I_{f\,max}$ / mA</th><th rowspan="2">$I_{f\,av}$ / mA</th><th colspan="2">Typ V_f @ I_f</th><th colspan="2">Typ I_r @ $V_r^{\dagger}$</th><th rowspan="2">Data</th></tr>
<tr><th>/ V</th><th>/ mA</th><th>/ µA</th><th>/ V</th></tr>
<tr><td>1N4148</td><td>Si P-N junction (4 pF @ 0 V)</td><td>75</td><td>200</td><td>-</td><td>1.0^{max}</td><td>10</td><td>0.025</td><td>20</td><td>a</td></tr>
<tr><td>1N5711</td><td>Si Schottky (2.0 pF)</td><td>70</td><td>-</td><td>-</td><td>0.41^{max}</td><td>1</td><td>0.2</td><td>50</td><td>b</td></tr>
<tr><td>1N5712</td><td>Si Schottky (1.2 pF)</td><td>20</td><td>-</td><td>-</td><td>0.55^{max}</td><td>1</td><td>0.15</td><td>16</td><td>b</td></tr>
<tr><td>1N6263</td><td>Si Schottky (2.2 pF)</td><td>60</td><td>50</td><td>15</td><td>0.41^{max}</td><td>1</td><td>0.2</td><td>50</td><td>f</td></tr>
<tr><td>AA119</td><td>Ge point contact *</td><td>45</td><td>100</td><td>35</td><td>2.6</td><td>30</td><td>170</td><td>45</td><td>c, d</td></tr>
<tr><td>AAY30</td><td>Ge Au-bonded. High speed</td><td>30</td><td>400</td><td>-</td><td>0.88</td><td>150</td><td>8.0</td><td>30</td><td>c, d</td></tr>
<tr><td>AAY32</td><td>Ge Au-bonded. High speed</td><td>30</td><td>150</td><td>-</td><td>0.60^{max}</td><td>30</td><td>11</td><td>30</td><td>c, d</td></tr>
<tr><td>AAY33</td><td>Ge Au-bonded. High speed</td><td>12</td><td>240</td><td>-</td><td>0.5^{max}</td><td>30</td><td>15</td><td>12</td><td>c, d</td></tr>
<tr><td>AAZ15</td><td>Ge Au-bonded. High voltage</td><td>100</td><td>250</td><td>-</td><td>0.8</td><td>250</td><td>16</td><td>100</td><td>c, d</td></tr>
<tr><td>AAZ17</td><td>Ge Au-bonded. Gen. purpose</td><td>75</td><td>250</td><td>-</td><td>0.8</td><td>250</td><td>16</td><td>75</td><td>c, d</td></tr>
<tr><td>BAT81</td><td rowspan="3">Si Schottky (1.6 pF)</td><td>40</td><td rowspan="3">30</td><td rowspan="3">-</td><td rowspan="3">0.33
0.41
1.0</td><td rowspan="3">0.1
1
15</td><td rowspan="3">0.2</td><td rowspan="3">30</td><td rowspan="3">a</td></tr>
<tr><td>BAT82</td><td>50</td></tr>
<tr><td>BAT83</td><td>60</td></tr>
<tr><td>OA47</td><td>Ge Au-bonded. Gen. purpose</td><td>30</td><td>150</td><td>-</td><td>0.54</td><td>30</td><td>10</td><td>30</td><td>c, d</td></tr>
<tr><td>OA90</td><td>Ge point contact *</td><td>30</td><td>45</td><td>10</td><td>2.0</td><td>30</td><td>300</td><td>30</td><td>c, d</td></tr>
<tr><td>OA91</td><td>Ge point contact</td><td>115</td><td>150</td><td>50</td><td>2.1</td><td>30</td><td>75</td><td>100</td><td>c, d</td></tr>
<tr><td>OA95</td><td>Ge point contact</td><td>115</td><td>150</td><td>50</td><td>1.85</td><td>30</td><td>80</td><td>100</td><td>c, d</td></tr>
<tr><td>PAD5</td><td>Si PN low leakage (0.5 pF)</td><td>45</td><td>-</td><td>-</td><td>0.8</td><td>5</td><td>5 pA</td><td>20</td><td>e</td></tr>
</table>

* Very high reverse leakage current. Generally best avoided, but leaky diodes will work in high impedance detector circuits with no external DC return[46]. Not for precision measurement applications.

Data sources

a) Philips Components Quick Reference Guide 1990.

b) Agilent 1N5711, 1N5712, 5082-2300 Series, 5082-2800 Series, 5082-2900 Schottky Barrier Diodes for General Purpose Applications. These have a guard-ring structure, which increases reverse breakdown voltage but also adds capacitance.

c) Mullard Semiconductors, 1974/5

46 **Hi Fi Detector for AM broadcast**. Robert Batey. http://www.g3ynh.info/circuits/hi-fi_am.html. See. TRF circuit with regeneration.

d) Mullard Industrial Semiconductors, Quick Reference Guide 1969/70
e) Linear systems datasheet. www.linearsystems.com/
f) ST Microelectronics datasheet www.st.com/

† Typical reverse leakage current at a given voltage gives a way of estimating the effective parallel resistance. Provided that the reverse saturation leakage current is much smaller than the total reverse leakage current:

$R_{dp} = V_r / I_r$

If the saturation leakage current is high, it should be subtracted from the total to find the ohmic contribution, i.e.;

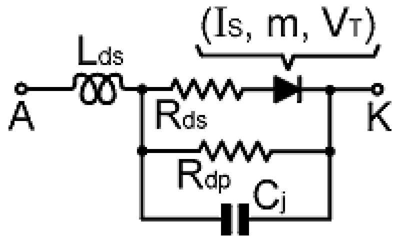
$R_{dp} = V_r / (I_r - I_S)$

Point contact junctions

Note that point-contact diodes are sometimes Schottky diodes and sometimes not. If the point-contact results in a semiconductor-metal junction, then the diode might be reasonably classified as a Schottky type. Often however, there is an initial burn-in step in manufacture, which causes some of the metal to diffuse into the semiconductor crystal and form a P-N junction[47]. Given this ambiguity, the term 'Schottky diode' is usually reserved to mean diodes having a semiconductor-metal junction formed by evaporation of metal on to the crystal surface.

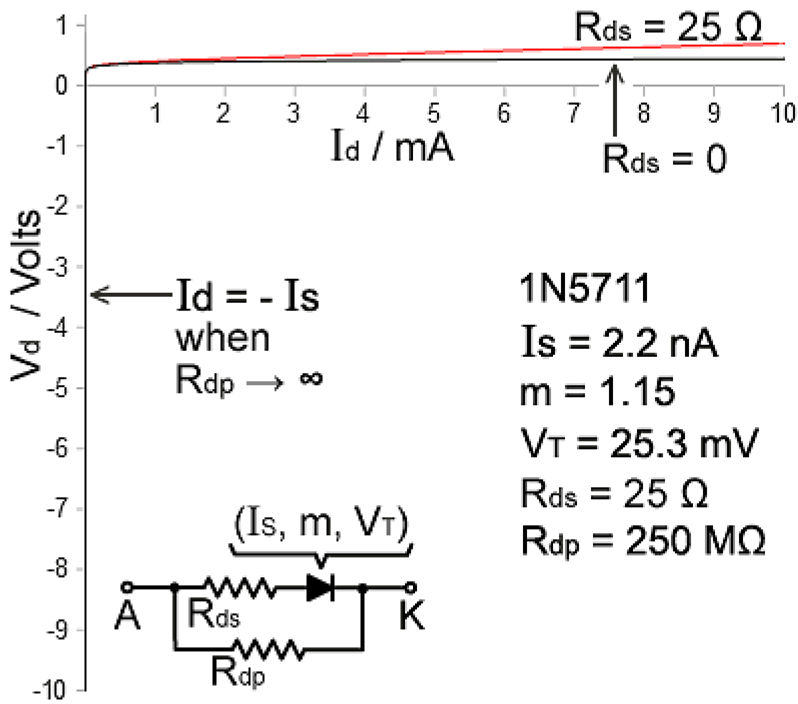
47 **Schottky barrier devices**. J C Irvin and N C Vanderwal, in **Microwave semiconductor devices and their circuit applications**, H A Watson (Ed.), McGraw-Hill 1969, LCCN 68-17197. p340-369, see p341.

8. Diode circuit model



A suitable model for in-circuit diode simulation is shown on the right. The non-linear behaviour is described by the three parameters of the modified Shockley ideal diode equation (I_s, m, V_T in equation **6.1**), and all of the other parameters can be represented as conventional circuit elements.

L_{ds} is the series partial inductance associated with every circuit component, and C_j is the junction capacitance. L_{ds} is a few nH for wire-ended diodes (the old 20 nH per inch rule provides a rough guide), but is very small for surface-mount devices (neglecting PC tracks). C_j is as per the manufacturer's literature (2 pF for the 1N5711) unless quantified experimentally. Note that a half-wave diode detector usually feeds into a smoothing capacitor. The smoothing capacitor intentionally has nearly zero reactance at the generator frequency. The capacitance shunting the diode also makes zero contribution to the DC output. Hence, at the driving frequency, C_j is effectively connected between the input-side network and ground. This can help to simplify the RF modelling problem at higher radio frequencies. Reactances, of course, can be neglected at lower frequencies.



A simulation of the DC or instantaneous characteristic of a 1N5711 Schottky diode (neglecting reactance) is shown on the right (see open document spreadsheet file **det_models.ods**, sheet 1). Voltage and current are defined as positive when the anode is positive with respect to the cathode. Two curves are shown; one with R_{ds} taken into account, and the other with $R_{ds} = 0$ (i.e., ignored). The typical R_{ds} of 25 Ω implies a substantial contribution to the forward voltage at $I_d = 10$ mA, but it is not so important with currents of a few microamps. If R_{dp} is left open circuit (i.e., ignored) the diode reaches its reverse saturation leakage current when the negative bias exceeds more than about 0.1V. If R_{dp} is included, the total current under negative bias is $-I_s + V_d / R_{dp}$ (where V_d is negative). The model assumes that the reverse breakdown voltage is never approached.

8.1 Diode stacking

The diode static current vs. voltage characteristic (neglecting parallel resistance) is given by the modified Shockley equation (**6.1**).

$$V_d = I_d R_{ds} + m V_T \ln[(I_d / I_S) +1]$$

If we put several identical diodes in series, then the forward voltage drop for a given current is multiplied by the number of diodes (n_d say). Thus

$$V_{nd} = n_d \, I_d \, R_{ds} + n_d \, m \, V_T \ln[\, (I_d / I_S) + 1 \,]$$

Thus the effect is to multiply the thermal voltage (V_T) and the series resistance by the number of diodes. The parallel resistance is, of course, similarly multiplied, as is the series partial inductance. The capacitance however is halved.

9. Back diodes

One further rectifying device that we should mention in passing is the *backward diode* or 'back-diode' [48] [49]. This is a type of tunnel diode or Esaki diode. Due to heavy doping of the semiconducting material (often germanium), the tunnel diode has a negative resistance region in its V/I characteristic, which makes it useful as an amplifier, oscillator or trigger device. The back diode is a less-heavily doped version, which means that the negative resistance characteristic largely suppressed. This gives it a characteristic more like that of a normal diode, except that it has an extremely low forward voltage drop, somewhere in the region of 90 mV for 10 mA forward current.

Unfortunately however, there is no general workaround as far as the diode equation is concerned. The back diode has an extremely low reverse breakdown voltage, and a leakage current of around 1 mA for only 500 mV of reverse bias[50]. The reason is that, relative to the normal electrode naming for a given material type, the back diode is used backwards; i.e., it conducts more in the 'reverse' direction than in the 'forward' direction. This is due to the quantum-mechanical tunnelling effect after which this class of devices is named.

The diode can therefore only be used in extremely low-impedance low-voltage circuits, which makes it unsuitable for the construction of conventional peak detectors. It does however make a good square-law detector, and can therefore be used for direct conversion of voltage readings into power readings[51]. It also has low capacitance, and is free from charge storage effects, making some types useful up to about 40 GHz.

Back diodes are still a current microwave technology[52], but they are expensive. When a square-law detector is required for more general RF applications, and presuming that active circuitry can be used, a more cost effective solution is to use a linear detector followed by an anti-logarithmic amplifier or an analogue multiplier. Another solution is to digitise the output of a linear detector and perform the arithmetic using a computer or a microcontroller.

---

48 **The art of electronics**. 2nd ed. Horowitz & Hill (previously cited). Back diode as a square-law detector, p891-892. Tunnel diode p14-15, p1060.

49 **Amplifier Handbook**. Ed. R F Shea. Mc Graw Hill 1966. ISBN 0-07-056503-1. Ch. 12. **Tunnel diodes and backward diodes**. C S Kim and J J Tiemann. Theory, characteristics and applications.

50 **Effective noise reducer and hearing protector**. R L Rod, QST, April 1978, p40. Simple audio clipper circuit using back diodes. Gives V/I characteristic for General Electric BD1.

51 **New approach to measuring SWR at high frequencies**. U L Rohde. Ham Radio. May 1979, p34-35. The Rohde & Schwarz NAUS 80 RF power meter uses back diodes as square-law detectors.

52 http://www.aeroflex.com/ams/metelics/micro-metelics-prods-TD-MTD.cfm

10. Vacuum thermionic diodes

Given that semiconductor diodes are not ideal rectifiers, some readers might wonder if thermionic diodes (valves) are capable of better performance. The answer however is a most emphatic 'no!' The forward voltage of an indirectly heated valve diode in the low-current (space-charge limited) regime[53] is given by the expression:

$V_f = (I_f / K)^{3/2} - V_c$

where K is a constant determined by the geometry of the diode and V_c is called the *contact potential*, by analogy with the potential developed by a thermocouple. The diode contact potential arises because some of the electrons ejected from the cathode arrive at the anode even when there is no bias, and so the anode becomes negatively charged. This means that the diode must be reverse biased in order to prevent it from conducting, and the amount of bias required varies with the heater temperature and the age of the valve.

In precision measurement terms, the contact potential is enormous. A sample of eight double-diode valves of type EB91 (6AL5) showed contact potentials ranging from 0.48 V to 0.96 V (with a mean of 0.73 V) when measured using a voltmeter with 10 MΩ input resistance and a stabilised 6.30 V DC heater supply (due to drift, sensible measurement was impossible using a filament transformer connected to the AC mains supply). A 100 µA moving coil meter with an internal resistance of 980 Ω was connected across a diode having an open-circuit contact potential of 0.74 V, and registered a zero-bias current of 70 µA, i.e., the diode gave a DC output of 4.8 µW due to the thermal current. When the meter was padded to 10 kΩ, to simulate the diode loading in a reasonably realistic detector circuit, the zero-bias current was 22 µA, i.e., 22% of full-scale deflection.

Without even considering the other (manifold) shortcomings of the device, we may safely conclude that, in comparison to semiconductor diodes, thermionic diodes have no merit in small-signal detecting applications. That however must leave us to wonder why they were used and recommended for that purpose in the early and mid 20th century. It certainly wasn't because the semiconductor diode had yet to be invented.

The semiconductor diode does, of course, have a thermal voltage; but it is nothing like that which obtains from heating the cathode to between 600ºC and 1000ºC.

53 **Radio Engineering**, F E Terman, McGraw-Hill, 3rd edition 1947. Section 5.5, Diodes - Space-charge effects.

11. A brief history of diode detectors

Nature abhors a vacuum tube.[54]

The vacuum thermionic diode was actually invented by Thomas Edison and described an 1884 Patent[55]. Edison was well aware of its one-way conduction property (which is known as the 'Edison effect'), but the patent relates to his use of the critical relationship between anode current and filament temperature (and hence voltage) to make an expanded-scale voltmeter and a voltage-regulator for DC power generators. J A Fleming had worked for Edison at that time, and had been involved with the experiments.

In 1904, Fleming's contract as scientific advisor to Marconi had come to an end and he was engaged in research relating to the development of high-frequency AC measuring instruments[56]. Aware that some of the types of radio detector in use at the time have a rectifying action, he decided to try using the Edison diode in conjunction with a galvanometer, and met with success. He then patented the invention[57] [58], essentially a crude diode RF voltmeter, but also somewhat overstepped the mark by laying claim upon the use of the valve for rectification in general, as well as for his particular application.

Fleming 'Oscillation Valve' 1918
(Platinum filament)

What is perhaps not obvious is that a Patent is of no intrinsic value to a private individual, and it can be a financial burden. For those without manufacturing interests, a patent is only desirable if it can produce royalties, or if it can be used as a bargaining tool. Fleming seems to have taken out the patent because he saw it as a way of getting-back his job with the Marconi company. He wrote to Marconi on several occasions about patents he had filed at around that time, but the diode appears to have been the one that did the trick. Fleming was duly re-employed, but on a contact that forced him to transfer control of all relevant patents taken out in the time between his two periods of employment. In other words, had had to give the valve patent to Marconi.

It was Marconi who turned the detector into a moderately useful demodulator by replacing the galvanometer with a telephone earpiece. Still, there was not much to listen to except the buzzing of spark transmitters, and the device was never clearly superior to other types of detector used by Marconi at the time (it is less sensitive than a coherer, but it doesn't require resetting). Also, it was doomed to become silent with the move to quasi-continuous and then continuous waves. It should, by rights, have been destined for obscurity, at the very least until the advent of amplitude modulation and broadcasting after the First World War; had not Lee de Forest had the bare-faced

---

54 Quip usually attributed to J R Pierce of Bell Labs., but attributed by Pierce to his colleague Myron Glass. http://www.smecc.org/john_r__pierce____electron_tubes.htm (accessed 27th Aug. 2014).

55 **Electrical Indicator**. T A Edison, 1884 (filed 1883), US Pat. No. 307031.

56 **Inventing the history of an invention: J A Fleming's route to the valve**. Sungook Hong, 2000. Artifacts, Vol 2, 02. http://www.artefactsconsortium.org/Publications/PDFfiles/Vol2Elect/2.01.Electronics-Hong,FlemingValveGr75ppiWEBF.pdf (accessed 27th Aug. 2014).

57 **Improvements in instruments for detecting and measuring alternating currents.** J A Fleming. 1904, British Pat. No. 24850..

58 **Instrument for converting alternating electric currents into continuous currents**. J A Fleming. 1905 US Pat. No. 803684.

effrontery to invent the triode[59].

Being the first practical electronic amplifying device, the thermionic triode really did herald the start of the vacuum-tube revolution[60]. It had two serious problems however; the first being that it had been invented by one of Marconi's rivals, and the second being that it was superficially similar to Fleming's valve. This led to years of litigation, and despite a 1915 disclaimer withdrawing invalid claims in Fleming's US patent; de Forest sustained sufficient financial damage to cause him to allow his British triode patent to lapse. That meant that the Marconi company could manufacture and sell triode valves freely in the UK, while the US market remained in a state of legal uncertainty until the Fleming patent expired in 1922. In 1943, the US Supreme Court declared Fleming's patent to have been entirely invalid for all time, due to deliberate false claims and failure to declare prior art[61]. That was too late to be of any use to the US electronics industry; and although it might have entitled de Forest to some redress, nothing came of it.

According to de Forest, the Fleming valve was nothing more than a laboratory curiosity. This is a reasonable assessment in the context of signal detection, although thermionic diodes are useful as noise sources. The publicity surrounding the legal battles relating to it however gave it a perceived merit that it does not actually posses, and this will have had an influence on circuit designers. The main reason why vacuum diode detectors were used however lies in the cosy relationship that grew up between the valve manufacturers and the equipment manufacturers, especially in the consumer sector. Often the valve makers and the set makers were one and the same; and in Britain they operated a vast cartel[62], with restrictive practices that included approved component types and sanctions against companies that tried to source parts independently. Hence a preference for the valve diode detector; although it must be understood that it was only of use in circuits having plenty of pre-amplification, such as superheterodyne radio receivers. Receivers needing sensitive low-noise detectors and mixers had to use either the triode valve, or the semiconductor crystal diode.

To be fair to the light-bulb cartel companies, it took a long time to perfect the manufacture of the crystal detector; and a couple of spare thermionic diodes could easily be included in the superhet final IF amplifier or first audio valve. That does not however explain the marketing of signal diode valves (like the 6AL5 shown earlier, which was introduced in 1945) at a time when high-inverse germanium diodes were available; or the curious practice of supplying radio sets to the British market with valve power supply rectifiers and detector diodes, and equivalent models for the North American market with semiconductor devices in those positions.

Meanwhile, in a parallel universe, the elements of the true electronics revolution were beginning to come together. The unilateral conduction of crystal-metal junctions was discovered by Ferdinand Braun[63] in 1874. The use of this effect for the detection of radio signals was then developed by Jagadis Chandra Bose from 1894 onwards[64]. Bose experimented with several types of semiconductor crystal, plotting the V-I characteristics using a galvanometer capable of registering 1 nA, and appears to have anticipated the existence of P and N-type materials. He also gave public demonstrations of radio signalling in 1895, pre-dating Marconi by two years.

---

59 **Device for amplifying feeble electrical currents**, Lee de Forest, 1907 (filed 1906), US Pat. No. 841387.

60 Although, it must be understood that de Forest's 'Audion' was too gassy to work properly as a linear amplifier, and constitutes the invention of the control-electrode rather than a fully developed technology.

61 **The genesis of the thermionic valve**. Lecture given to the IEE on the 50th adversary of the invention of the thermionic valve. G W O Howe, 1954. [Howe's lecture was not what his audience would have expected. Refer to the full transcript (http://www.g3ynh.info/valves/old/history/Howe1955_genesis.pdf), not the heavily abridged summary appearing in JIEE, March 1955, p158].

62 **Report on the supply of electronic valves and cathode ray tubes**, The Monopolies and Restrictive Practices Commission, HMSO 1956. http://webarchive.nationalarchives.gov.uk/+/http://www.competition-commission.org.uk/rep_pub/reports/1950_1959/020cathode.htm

63 http://en.wikipedia.org/wiki/Cat's-whisker_detector

64 **The work of Jagadis Chandra Bose, 100 years of mm-wave research**. D T Emerson. IEEE Microwave Symposium (MTT-S) Digest 1997, Vol. 2, p553-556.

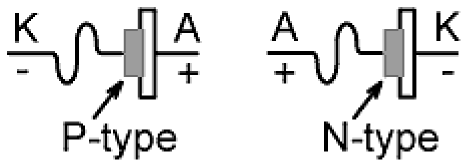
Bose's preferred detector was the galena[65]-metal junction, now recognised to be a type of Schottky barrier diode. To make a diode, the semiconductor crystal requires an area contact and a point contact, and those elements are the basis of the modern diode symbol. The springy sharpened wire used for the point contact is known colloquially as a 'cat's whisker'. Bose used the point-contact rectifier, in conjunction with waveguide and horn antennas, at frequencies up to 60 GHz. He was awarded a patent on the galena diode in 1904, having filed the specification in 1901[66]. Also, in 1906, a patent for the silicon point-contact diode was awarded to GW Pickard[67].

It is thus evident that semiconductor diodes were well known at the time when Fleming made his 'discovery', and they were considerably more sensitive than his valve even at the long-wave frequencies that Marconi was using. Both Fleming and Marconi were also known to keep a close watch on the patent and electrical literature. It is therefore inconceivable that they did not know of Bose's work, and we are left to wonder why they ignored it. A possible explanation is that Fleming was not so much looking for a detector, as looking for a novel detector that could be offered to Marconi; and that Marconi was primarily interested in inventions he could control. When commercial production of valves began in 1919, Marconi joined forces with GEC, a patent pooling arrangement was agreed, and the Marconi-Osram Valve company was formed. The Marconi company did make crystal-set radio receivers from about 1915 onwards[68 69], there being a considerable market for them; but the large valve-orientated electronics manufacturing concerns made no self-driven effort to turn crystal diodes into stable reproducible components until after the Second World War.

The modern semiconductor electronics revolution is closely associated with research into UHF and microwave radar and communications, with much of the work taking place at Bell Laboratories in the USA. There is some doubt about whether Bell Labs was first to produce a working transistor[70 71], but the company certainly made a concerted effort to investigate and perfect the detector diode from about 1934 onwards[72]. About 100 different crystalline materials were investigated by the Bell scientists, and silicon and iron pyrites[73] (fool's gold, FeS_2) were found to give the best results. Subsequent research led to the production of uniformly active rectifying surfaces on silicon crystals, eliminating the need to search for hot spots. This resulted in the first manufacturable modern diode, the 1N21 front-end mixer (used in 3 GHz radar sets) in 1942. The 10 GHz 1N23B (introduced in 1944) is similar in appearance, and is shown on the right.

65 Lead(II) sulphide, PbS. Crystals are cubic or octahedral, with a black opaque shiny appearance.
66 **Detector for electrical disturbances**. J C Bose, US Pat. No. 755840. 1904 (filed 1901).
67 **Means for receiving intelligence communicated by electric waves**. G W Pickard. US Pat. No. 836531, 1906 (filed 1906).
68 http://www.sparkmuseum.com/MARCONI.HTM
http://www.sparkmuseum.com/CRYSTAL4.HTM (Accessed 17th Aug. 2014)
69 **The Cat's Whisker. 50 Years of wireless design**. Jonathan Hill.
Oresko Books, 1978. ISBN 0-905368-47-9. p43-44.
70 **Method and apparatus for controlling electric currents**. J E Lilienfeld, US Pat. No. 1745175, 1930 (filed 1926).
Amplifier for electric currents. J E Lilienfeld, US Pat. No. 1877140, 1932 (filed 1928).
Device for controlling electric current. J E Lilienfeld, US Pat. No. 1900018, 1933 (filed 1928).
71 **The Other Transistor: early history of the metal-oxide-semiconductor field-effect transistor**, R G Arns. Engineering Sci. and Education. Journal. Oct. 1998. p233-240.
72 **Development of silicon crystal rectifiers for microwave radar receivers**, J H Scaff and R S Ohl, Bell System Technical Journal, Jan. 1947, 26(1), p1-30. http://www3.alcatel-lucent.com/bstj/
73 **Vacuum-tube and crystal rectifiers as galvanometers and voltmeters at ultra-high frequencies**. Arnold Peterson. General Radio Experimenter, May 1945.
http://www.ietlabs.com/genrad/experimenters/ (accessed 17th Aug. 2014)

Early silicon crystal specimens incidentally were categorised into two distinct types; P-type and N-type. P-type materials gave maximum conduction with the body of the crystal as anode (i.e, positive with respect to the point contact) and for N-type materials it was the other way around. This is illustrated in the diagram on the right, and it will be seen that modern circuit symbol is based on the N-type diode. The Bell researchers found that N-type crystals made the best microwave mixers; while the P-types gave better sensitivity at low voltages, which made them useful in test equipment. Nowadays, of course, we know that the two different types arise because of impurities; which can be removed by zone-refining of long monocrystalline bars, and reintroduced by doping of thin wafers; so that the material resistivity and the preferred majority carrier (holes or electrons) can be controlled.



Shown here on the right is a 1950s vintage commercial germanium diode (GEC GEX34), fused into a 4.7mm diameter glass tube; and since the cathode is marked with red paint, it can be deduced that the material is N-type. In N-type materials, the majority carriers are electrons, which have twice the intrinsic mobility of holes.

The germanium diode was originally developed as a high-inverse-voltage signal rectifier[74] ($V_{RM} > 100$ V), suitable for use as the video (final) detector in radar receivers. Its advantages over the vacuum-tube diode are nicely summarised in volume 15 of the MIT Radiation Laboratory series[75]:

- Much greater forward conductance.
- The I-V characteristic passes through the origin (practically no contact potential).
- It becomes approximately linear at low voltage.
- It has very low inter-electrode capacitance, and low capacitance to ground.
- It has no heater, and so produces no mains hum.
- It is about the same size as a small resistor, and it doesn't need a socket.

Prior to the development of zone refining techniques, early silicon diodes had high reverse leakage; and so the germanium diode was the first unambiguous direct replacement for the valve. Early commercial units made their debut in post-war TV receivers[76], since the 38 MHz final IF (inherited from radar practice, and necessitated by the need for wide signal bandwidth) was fairly near to the limit of valve diode capability.

Scientists working at Bell labs on microwave communication, and British scientists working on wartime airborne radar development, both saw the importance of semiconductor detectors early in their research programmes. Other radar researchers took a little more convincing however. In 1940, Britain desperately needed to develop its centimetric radar capabilities, but lacked the manufacturing capacity required. A decision was made therefore, that Britain's radar secrets would be shared with the USA, in return for joint R&D effort and the supply of manufactured sets. The result was the Tizard mission in August and September 1940, during which the cavity magnetron was revealed. That led to the establishment of the Radiation Laboratory (Rad. Lab.) at MIT in

74 **The origin of semiconductor research at Purdue**. http://www.physics.purdue.edu/about_us/history/semi_conductor_research.shtml (Accessed 17th Aug. 2014)

75 **Crystal Rectifiers**, H C Torrey and C A Whitmer, MIT Rad. Lab. Series Vol. 15, 1948. Ch. 12, p361. High-inverse-voltage rectifiers.

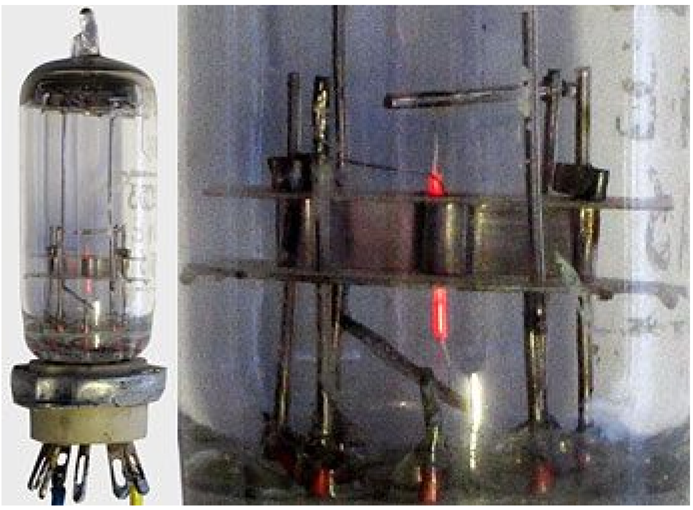
76 See for example, **Mullard germanium diodes**, Mullard Outlook, January 1954. https://sites.google.com/site/transistorhistory2/home/philips-mullard-resources (Accessed 16th Aug. 2014).

October of that year. American progress with the magnetron and its ancillary components was rapid; and in July 1941 there was an informal showdown between British and American 3 cm radar sets[77] at Leeson House, near Swanage, in Dorset, Southern England. These were bench-top tests of prototype units intended for airborne use; and involved picking-up targets looking out across the bay towards the Isle of Wight. This was the point at which the conventional vacuum tube was to be forever banished from the microwave receiver front-end.

The British and American sets gave similar performance, with the small anomaly that the British unit was the overall winner, while the Rad. Lab. set produced considerably more transmitter power. Furthermore, tests in the US had shown earlier American receivers to be superior to the British ones, and there appeared to have been a performance downturn. Finally a British receiver unit was connected to an American set; and the effective detection range increased by about a factor of three.

It was quickly discovered that the discrepancy was due to the Rad. Lab. receiver having a grounded-grid triode in the front end; whereas the British, having adopted semiconductor first mixers at the very beginning of their microwave development programme, were using silicon diodes. It turned out that the Rad. Lab. researchers had tried crystal mixers but abandoned them because they felt that vacuum tubes were superior. Somewhat later it was determined that their comparison test had been conducted using only a single diode, which had happened to be defective.

*   *   *



1A3 (CV753, DA90) Signal diode[78]
This miniature 7-pin (B7G) tube was introduced by RCA in 1943. It probably represents one of the last attempts to compete at VHF / UHF with the new germanium diodes about to come on to the market. The anode is about 2 mm tall, the anode-to-cathode capacitance is 0.6 pF, and the self-resonance frequency is about 1 GHz. The device still has a heater though; and a spurious heater-to-cathode capacitance of 0.7 pF, which inevitably restricts its application.

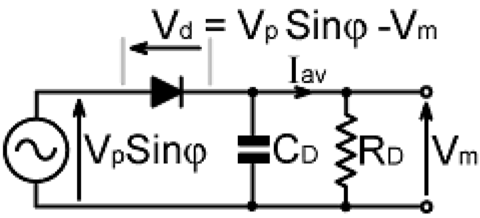
With the heater connected to a 1.5 V Lithium cell, the contact potential for the diode shown was 0.1 V when measured using a 10 MΩ DVM.

77 **The invention that changed the world. The story of radar from war to peace**. Robert Buderi. Abacus, 1996. ISBN 0 349 11068 9. see p117.
78 Technical data obtained from National Valve Museum, http://www.r-type.org/exhib/aaa0872.htm and http://www.r-type.org/exhib/aad0031.htm (accessed 17th Aug. 2014).

12. Simple diode voltmeter dynamic characteristics

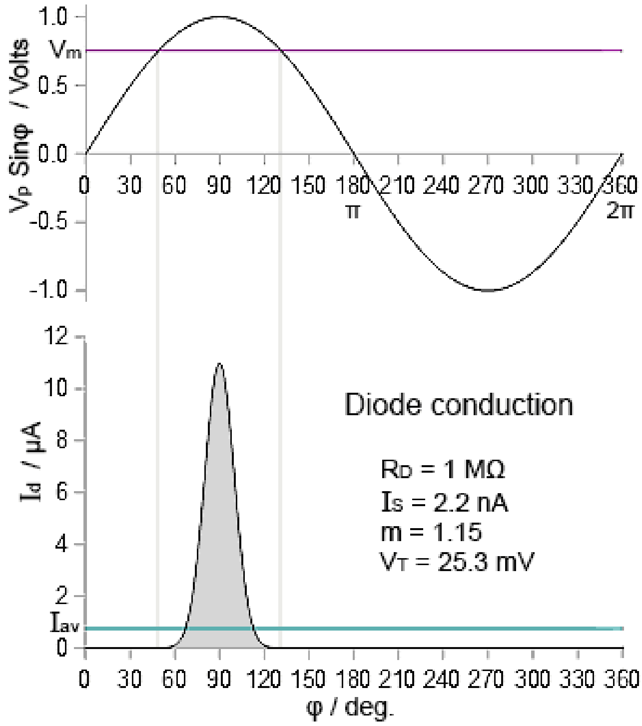
For those who wish to make accurate RF measurements using only DC measuring instruments for calibration, it should be evident that, for the majority of applications, the semiconductor-metal junction diode is the best broadband rectifying device available. It is certainly not the most linear of detectors if it is compared against active (powered) circuits, but it is the basis of circuits offering extremely wide bandwidth in conjunction with a well defined conduction characteristic. To make use of its properties for absolute voltage measurement purposes however, or to to make accurate corrections for its non-linearity, we must quantify its dynamic behaviour.

12.1 AC-DC Transfer function



We will use the simplified detector circuit shown on the right as a starting point for AC analysis[79]. This model assumes that the generator has zero output impedance and no DC resistance, and that the diode has no series or parallel resistive components (matters that will later require correction). Also, the smoothing capacitor C_D is taken to be very large, so that there is no voltage droop between the peaks of the driving waveform.

The generator produces an instantaneous output voltage of $V_p \sin\varphi$, where V_p is the peak voltage, and $\varphi = 2\pi f\, t$ is the time-varying phase angle[80]. V_m is the DC output voltage (the raw measurement), and V_d is the instantaneous voltage appearing across the diode. R_D is the total detector load resistance, including the input resistance of any subsequent circuitry.



The instantaneous current flowing through the diode is, according to the modified ideal diode equation (see section **6**):

$$I_d = I_S \left[\exp\left(\frac{V_d}{m V_T} \right) - 1 \right]$$

Expressed in terms of the time-varying input, this becomes:

$$I_d = I_S \left[\exp\left(\frac{V_p \sin\phi - V_m}{m V_T} \right) - 1 \right] \qquad (\mathbf{12.1})$$

This is shown plotted on the right for an input of 1 V peak, a load of 1 MΩ, and diode parameters as for a 1N5711 with zero junction resistance (see worksheet file **det_models.ods**, sheet 2). The forward conduction angle (i.e., the portion of the cycle during which the diode is positively biased) varies with V_p, and so note that the graphs do not simply scale according to the input. The average diode current, I_{av}, and the

79 The development given here, for the simplified case, is based on the author's attempt to recreate the derivation of formulae given in the article: **Theory of diode voltmeters and some applications**, by A E **Weller,** QEX, Jan. 1984, p7-14. In that article, unfortunately, there is an error in the solution of the main integral (equation 4), and so the subsequent analysis is flawed.

80 The problem can also be defined using the cosine, see, for example: Cvetković and Marković 1989. cited earlier.

measurable output, V_m , are the solutions to the problem we are presently investigating.

Equation (**12.1**) allows us to plot the instantaneous current for a given output voltage, but we need to determine what that output voltage is. To do that we note that:

$V_m = I_{av} R_D$

The average diode current is, of course, the area under the curve for instantaneous current (I_d) divided by the length of a cycle (i.e., 2π radians). Thus:

$$I_{av} = \frac{V_m}{R_D} = \frac{1}{2\pi}\int_0^{2\pi} I_d \, d\phi$$

It will be helpful to carry out a rearrangement and define some composite parameters in order to perform this integration. Firstly, the exponential in (**12.1**) can be separated into two parts, thus:

$$I_d = I_S\left[\exp\left(\frac{-V_m}{mV_T}\right)\exp\left(\frac{V_p \sin\phi}{mV_T}\right) - 1\right]$$

Now notice that, because we have made the smoothing capacitor C_D very large, V_m does not vary over the course of a cycle; and so the first exponential is a constant in the integration. $V_p/(mV_T)$ is also a constant, and so let us define:

$K_m = \exp\{-V_m / mV_T\}$ and $u = V_p / mV_T$

The integral then becomes:

$$I_{av} = \frac{I_S}{2\pi}\int_0^{2\pi}\left[K_m \exp(u \sin\phi) - 1\right] d\phi \qquad (\mathbf{12.2})$$

An obvious but somewhat long-winded way of performing this integration[81] is to expand the exponential as a series using:

$$\exp(x) = e^x = \sum_{i=0}^{\infty}\frac{x^i}{i!}$$ (note that i is used here to represent a real integer)

Hence, using this in (**12.2**) we get:

$$I_{av} = \frac{I_S}{2\pi}\int_0^{2\pi}\left[K_m \sum_{i=0}^{\infty}\frac{u^i \sin^i\phi}{i!} - 1\right] d\phi \qquad (\mathbf{12.3})$$

81 There is a more elegant solution to this problem involving Bessel's integral form of $J_0(x)$. We will use it to check the result. The longer method given here requires no special knowledge, and identifies the error in Weller's article.

General expressions for $\sin^i\varphi$ are given by Ryshik and Gradstein[82] (R&G). When a sine is raised to an odd power, the generating function is:

$$\sin^{2n-1}\phi = \frac{1}{2^{2n-2}}\sum_{k=0}^{n-1}(-1)^{n+k-1}\,.\,{}^{2n-1}C_k\ \sin\left[(2n-2k-1)\phi\right] \qquad \text{(R\&G 1.320.3)}$$

and when a sine is raised to an even power, the generating function is:

$$\sin^{2n}\phi = \frac{1}{2^{2n}}\left[{}^{2n}C_n + \sum_{k=0}^{n-1}(-1)^{n-k}.2\,.\,{}^{2n}C_k\ \cos\left[2(n-k)\phi\right]\right] \qquad \text{(R\&G 1.320.1)}$$

where ${}^{q}C_r$ is a binomial coefficient (pronounced "q choose r"), defined as:

$${}^{q}C_r = \frac{q!}{r!\,(q-r)!} = \frac{q(q-1)\,.....\,(q-r+1)}{r!} \qquad \text{(see R\&G, p433, or Dwight, p1)}$$

Expansions of $\sin^i\varphi$ for powers up to 7 are given by Dwight[83]. Evaluation of binomial coefficients is assisted by noting that they are the elements of Pascal's triangle. In this case however, it also pays to note the following:

- In the expansion of $\sin^{2n-1}\varphi$ (**1.320.3**), every term has the sine of an integer multiple of φ as a factor.

- In the expansion of $\sin^{2n}\varphi$ (**1.320.1**), the terms have the cosine of an integer multiple of φ as a factor, except for a single constant (${}^{2n}C_n / 2^{2n}$).

- Sinusoids (i.e., sines and cosines) that undergo an integer number of cycles over the range of an integration average to zero.

The corollary is that all of the sine and cosine terms from these expansions vanish in the integration process, leaving only the constants from the even power series. To make use of that information, we need to change the summation index in (**12.3**) because we want a series that only produces even powers. The substitution required is $i = 2n$. Putting that into (**12.3**) gives:

$$I_{av} = \frac{I_S}{2\pi}\int_0^{2\pi}\left[K_m\sum_{n=0}^{\infty}\frac{u^{2n}\,.\,{}^{2n}C_n}{(2n)!\;2^{2n}} - 1\right]d\phi$$

Now substituting for the series of binomial coefficients we obtain:

$$I_{av} = \frac{I_S}{2\pi}\int_0^{2\pi}\left[K_m\sum_{n=0}^{\infty}\frac{u^{2n}\,\cancel{(2n)!}}{\cancel{(2n)!}\;2^{2n}\;n!\,(2n-n)!} - 1\right]d\phi$$

82 **Tables of Series, Products, and Integrals**. I M **Ryshik** and I S **Gradstein**. VEB Deutscher Verlag der Wissenschaften, Berlin, 1957. p26 and 27. 1.320.1 (even), 1.320.3 (odd).

83 **Tables of Integrals and Other Mathematical Data**. H B **Dwight**. 4^{th} edn. Macmillan 1961. LCCN: 61-6419. Page 82. Formulae 404.12 - 404.17.

This can be rearranged [84] :

$$I_{av} = \frac{I_S}{2\pi} \int_0^{2\pi} \left[K_m \sum_{n=0}^{\infty} \frac{(u/2)^{2n}}{(n!)^2} - 1 \right] d\phi$$

Now recall that $u = V_p / (mV_T)$ is a constant in the integration. Therefore everything in the integral is now a constant, and the indefinite integration process is simply a matter of multiplying the whole thing by φ. Then in the upper integration limit we set φ=2π, and the the lower limit we set φ=0. Subtracting lower limit value of the integral (=0) from the upper limit value gives:

$$I_{av} = I_S \left[K_m \sum_{n=0}^{\infty} \frac{(u/2)^{2n}}{(n!)^2} - 1 \right]$$

An infinite series with the square of a factorial in the denominator is the classic signature of a Bessel function. Reference to McLachlan[85] or Dwight[86] reveals the series to be the modified Bessel function[87] of the first kind, zero order, $I_0(u)$.

$$I_0(u) = J_0(\mathbf{j}\,u) = \sum_{n=0}^{\infty} \frac{(u/2)^{2n}}{(n!)^2} \qquad \text{where } \mathbf{j} = \sqrt{-1}$$ (McLachlan p200, eq^n. 153, Dwight, p195, 813.1)

Thus the expression for the average diode current is:

$I_{av} = I_S [K_m I_0(u) - 1]$ . . . . (**12.4**)

There is, incidentally, a well known integral form for the ordinary Bessel function $J_0(x)$ that corroborates the result just given[88 89] :

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\pm \mathbf{j}\, x \sin\phi)\, d\phi$$ (McLachlan p190, eq^n. 10. Bowman p57, eq^n. 4.2)

If we put $x = \mathbf{j}u$ then $J_0(x) = I_0(u)$ and, bearing in mind that $\mathbf{j}^2 = -1$, we get:

$$I_0(u) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\pm u \sin\phi)\, d\phi$$

Substituting this into (**12.2**) gives (**12.4**) directly.

---

84 Notice here that 2n-n = n(2-1) = n , so that (2n-n)! = n! . This is where the mistake in Weller's article occurred. He did not give his working, and wrote 2! instead of n! .
85 **Bessel Functions for Engineers**. N W McLachlan. 2nd edition. Oxford, Clarendon Press 1955. p200. eq^n. 153.
86 **Dwight** (already cited), p195, 813.1.
87 Modified Bessel functions are the functions that result from evaluating an ordinary Bessel function using a purely imaginary argument.
88 **McLachlan** (already cited). p190. 2.10
89 **Introduction to Bessel Functions**. F Bowman. Dover 1958. LCCN 58-11271. SBN 486-60462-4. Chapter IV. Definite integrals. p57. eq^n. 4.2 .

We can now re-expand (12.4) by replacing the composite parameters. Thus:

$I_{av} = \frac{V_m}{R_D} = I_S \left[\exp\left(\frac{-V_m}{mV_T} \right) \ I_0 \left(\frac{V_p}{mV_T} \right) \ -1 \right]$	**12.5**

A solution for the rectified output voltage (V_m) can be obtained from this expression. Note however that there are two instances of V_m, one separable and one as an exponent. This means that there will be no closed-form analytical solution; although finding a solution by trial and error is fairly straightforward.

More instructively however, we can compare the expression with the modified Shockley equation as given at the beginning of this section:

$I_d = I_S$ [exp{ V_d / mV_T } - 1]

Here, V_d can be taken to be either the instantaneous diode voltage drop under dynamic conditions, or the diode forward voltage drop under static conditions (i.e., the error that occurs if we place it in series with a DC voltmeter). So let us define the effective diode forward voltage drop under dynamic conditions as:

$V_f = V_p - V_m$

This is, of course, the error that occurs when the diode is used as a peak detector.

In order to extract V_f from equation (12.5), we can define a new function (W_0 say) that provides a factor that converts an exponential into a modified Bessel function. This gives:

$$I_{av} = \frac{V_m}{R_D} = I_S \left[W_0 \exp\left(\frac{V_p}{mV_T} \right) \exp\left(\frac{-V_m}{mV_T} \right) - 1 \right] \qquad (12.6)$$

Combining the two exponentials then gives:

$$I_{av} = \frac{V_m}{R_D} = I_S \left[W_0 \exp\left(\frac{V_f}{mV_T} \right) - 1 \right] \qquad (12.6a)$$

Which can be rearranged:

$$\frac{V_m}{I_S R_D} + 1 = W_0 \exp\left(\frac{V_f}{mV_T} \right) - 1$$

Now taking the natural logarithm of both sides and rearranging, we get:

$$V_f = mV_T \left[\ln\left(\frac{V_m}{I_S R_D} + 1 \right) - \ln\left(W_0 \right) \right] \qquad (12.7)$$

An expression for W_0 is obtained by comparing (**12.5**) and (**12.6**):

$$I_0\left(\frac{V_p}{mV_T}\right) = W_0 \exp\left(\frac{V_p}{mV_T}\right)$$

Thus:

$$W_0 = \frac{I_0(V_p/mV_T)}{\exp(V_p/mV_T)}$$

Inserting this into (**12.7**), and noting that $-\ln(W_0) = \ln(1/W_0)$, gives:

$V_f = mV_T\left[\ln\left(\frac{V_m}{I_S R_D}+1\right)+\ln\left(\frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}\right)\right]$	**12.8**

Now bearing in mind that: $V_m = V_p - V_f$, this expression can be solved manually by entering an initial guess for $V_p - V_f$ and adjusting it until the two instances of V_f agree. Experimentation is also facilitated by noting that Bessel functions are built-in to modern spreadsheets (such as Open Office Calc). Program routines to perform this iteration process automatically must be carefully designed however, there being a problem of reliable convergence when V_p is close to zero (this matter is resolved in **section 13.3**).

If we put equation (**12.8**) in terms of the average diode current we get:

$$V_f = mV_T\left[\ln\left(\frac{I_{av}}{I_S}+1\right)+\ln\left(\frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}\right)\right] \qquad (\mathbf{12.9})$$

The first term is identical to the DC form of the diode equation except that it has an average current instead of a constant current. The second term is therefore a correction for the fact that the true current is varying over time. Thus we can write the effective diode forward voltage drop under dynamic conditions as:

$$V_f = V_{f\_} + V_{f\sim}$$

where the component associated with the continuous output current (DC) is:

$$V_{f\_} = mV_T \ln\left(\frac{I_{av}}{I_S}+1\right) \qquad (\mathbf{12.9dc})$$

and the AC contribution is:

$$V_{f\sim} = mV_T \ln\left(\frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}\right) \qquad (\mathbf{12.9ac})$$

There is, of course, a temptation to rewrite the latter expression as:

$V_{f\sim} = V_p - mV_T \ln[I_0(V_p / mV_T)]$

The reason for not using this form is that the ratio V_p/mV_T can easily amount to several hundred. Say for example that we have a peak input voltage $V_p = 10$ V and $m = 1.15$. If the temperature is 20ºC, then $V_T = 25.3$ mV, and so $V_p/mV_T = 344.2$. We then find that $e^{344} = 3.1\times10^{149}$ and $I_0(344) = 6.6\times10^{147}$. Subtracting the logarithms of such enormous numbers is likely to produce inaccurate results due to floating-point rounding error, but their ratio is only 46.5 in this case, and $V_{f\sim}$ is 112 mV. Also, as we will see shortly, this ratio is an important physical quantity. Hence it is best to compute the bracket first and then take the logarithm. This is particularly true when the argument is large because there is an asymptotic form of I_0 that involves an exponential. This is given explicitly for the zero-order function by Dwight[90] and McLachlan[91] and can be found in more general form in numerous other references.

$$I_0(u) = \frac{\exp(u)}{\sqrt{2\pi u}} \left[1 + \frac{1^2}{1!\,8u} + \frac{1^2\,3^2}{2!(8u)^2} + \frac{1^2\,3^2\,5^2}{3!(8u)^3} + \ldots\ldots \right] \qquad \text{for } u \geq 10 \qquad (\mathbf{12.10})$$

The polynomial in square brackets is not given in compact form in the books cited, but if we give it the symbol $P_{aI0}(u)$ (polynomial used in the asymptotic form of I_0), it can be written:

$$P_{aI0}(u) = 1 + \sum_{n=1}^{\infty} \frac{(2n-1)!}{(n-1)!\,2^{n-1}\,(8u)^n} \qquad \text{as } u\to\infty \;,\; P_{aI0}(u) \to 1 \qquad (\mathbf{12.10p})$$

The asymptotic form of $I_0(u)$ is good to 6 significant figures for $u \geq 10$, and can be evaluated to 5 significant figures with only 4 terms in the polynomial (i.e., terms up to $n = 3$). To put that in perspective with regard to the diode voltmeter problem, note that if $mV_T = 29.1$ mV (1N5711 @ 20ºC), then the asymptotic form applies when the peak input voltage is ≥ 0.29 V. For a silicon Schottky detector diode, the input threshold voltage for a usable output (and a starting point for accurate non-linearity correction) is somewhere around 0.35 V. Thus we find that the asymptotic form of the Bessel function is applicable even at the nominal threshold of usability of the detector.

The wide-range of applicability of the asymptotic form should not be taken to imply that we can disregard the general analytical form; particularly because the general form can (within its limitations) extend the range of our diode voltmeter theory to zero peak input. It does however allow us to characterise the large-signal limiting behaviour of the model with good precision using simple formulae.

If we substitute (**12.10**) into (**12.9ac**), with the polynomial written as $P_{aI0}(u)$, we get:

$$V_{f\sim} = mV_T \ln\{ \sqrt{2\pi u} / P_{aI0}(u) \} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \; (\mathbf{12.10ac})$$

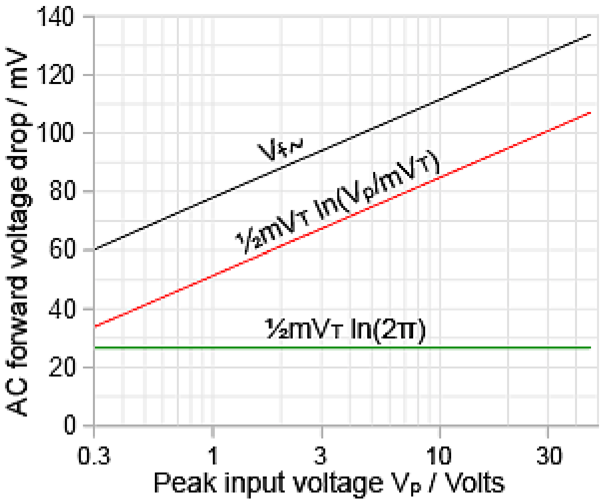
where $u = V_p / mV_T$

This can be separated into three parts, with the square root eliminated by halving the logarithm:

$V_{f\sim} = (½)mV_T \ln(2\pi) + (½)mV_T \ln(u) - mV_T \ln\{ P_{aI0}(u) \}$

90 **Dwight** (already cited). p196, 814.1.
91 **McLachlan** (already cited), p220, Table 13.

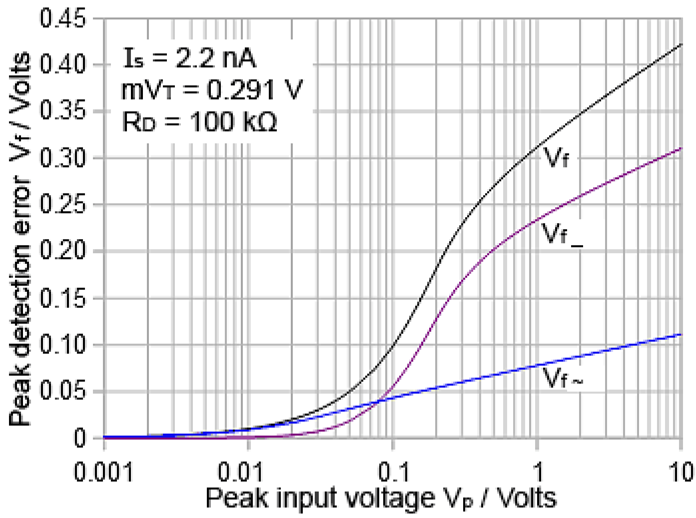
The positive terms and the total are shown plotted on the right for mV_T = 29.1 mV. The term involving the log of the polynomial is included, but it is not shown on the graph because it makes a contribution of only -0.38 mV when V_p is 0.29 V, and -10 µV when V_p is 10 V (see spreadsheet **det_models.ods**, sheet 5). The constant term ½mV_T ln(2π) is 26.7 mV, and the total varies between 60 mV and 112 mV as the peak input changes from 0.3 V to 10 V.

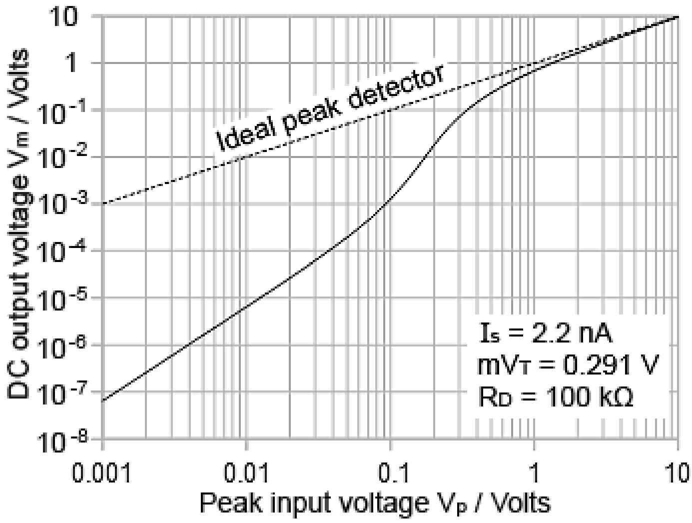


The error that would be incurred by computing the forward voltage drop using average current in place of direct current (i.e., by ignoring the dynamic correction) is fairly small for large input; around 1% for inputs of about 10 V peak and a 1 MΩ load. For precision absolute voltage measurement however, it is evident that the correction should be included.

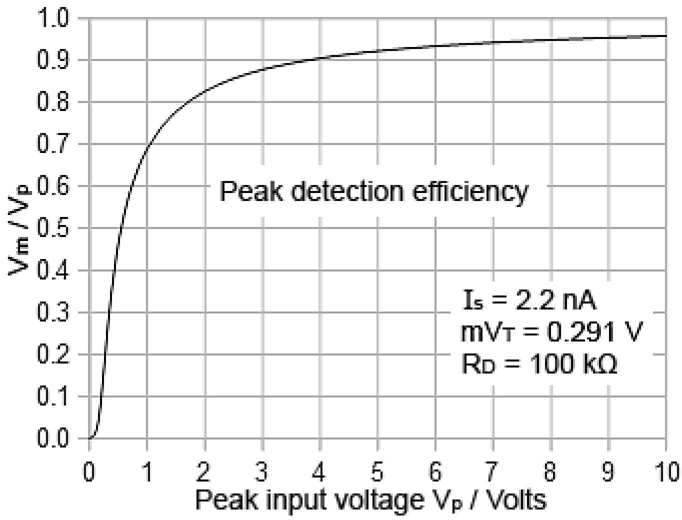
The issue of how to calculate the transfer function of a diode detector using the theory just developed is the subject of **section 13** (calculation procedures). Some representative graphs, modelling the behaviour of a 1N5711 detector with a 100 kΩ load resistance are given below. These were produced using the function program **DVfp2m**(R_D ; V_p ; mV_T ; I_S), which is described in **section 13.3** (see spreadsheet **det_models.ods**, sheet 3, for the working details).

The graph on the right shows the diode dynamic forward voltage-drop over a detector peak input voltage range from 1 mV to 10V. Also shown are the separate DC and AC contributions, the latter having been calculated using equation (**12.9ac**). Notice that the AC contribution is very different from the DC characteristic. This means that linearity compensation schemes using reference diodes in DC circuits cannot provide complete correction.





Here the dotted line shows what the output of an ideal peak detector would be. The realistic detector[92] falls somewhat short of that, giving practically no output until the peak input voltage reaches about 0.1 V.



This graph shows the ratio V_m / V_p as a function of the peak input voltage. For an ideal detector this would always be 1, so the curve is a measure of detection efficiency. Note the output onset threshold for a peak input of around 0.1 V.

92 Note that by plotting input voltage on a log. scale, the curve shown can be compared directly with graphs plotted in dB. See for example, **Harrison 1992** [cited earlier], Fig. 2., **Harrison - Le Polozec 1994** {cited earlier], figs. 3 & 4.

12.2 Peak to average current ratio

An expression for the instantaneous diode current was given earlier as equation (**12.1**)

$$I_d = I_S\left[\exp\left(\frac{V_p \sin\phi - V_m}{mV_T}\right) - 1\right]$$

By inspecting this we can see that the peak forward current occurs when $\sin\varphi = 1$, i.e., when $\varphi = +90^o$ or $\pi/2$ radians. Thus we can easily write a separate expression for the peak forward current (which we will call I_p) :

$$I_p = I_S\left[\exp\left(\frac{V_p - V_m}{mV_T}\right) - 1\right]$$

But V_p-V_m is the effective diode forward drop under dynamic conditions, to which we previously assigned the symbol V_f. This we have:

$I_p = I_S$ [exp(V_f / mV_T) - 1]

which can be rearranged:

(I_p / I_S) + 1 = exp(V_f / mV_T) . . . . (**12.11**)

Also, somewhat less trivially, we obtained an expression for the average diode current by integrating the expression for instantaneous current. This was given by equation (**12.6a**) as:

$I_{av} = I_S$ [W_0 exp(V_f / mV_T) - 1]

which can be rearranged:

(I_{av} / I_S) + 1 = W_0 exp(V_f / mV_T) . . . . (**12.12**)

The function W_0 was given earlier as:

$$W_0 = \frac{I_0(V_p/mV_T)}{\exp(V_p/mV_T)}$$

So taking the ratio (**12.11**) / (**12.12**) we get:

$$\frac{(I_p/I_s)+1}{(I_{av}/I_S)+1} = \frac{1}{W_0} = \frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}$$

Multiplying numerator and denominator of the left-hand side by I_S then gives:

$$\frac{I_p+I_s}{I_{av}+I_S} = \frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)} \qquad (\mathbf{12.13})$$

Now, bearing in mind that the reverse saturation leakage current (I_S) is small (a few nA for a silicon Schottky diode), this quantity is effectively the peak to average current ratio. It is also the ratio that governs the AC correction term in the effective diode forward voltage drop.

The error in the detector output ($V_f = V_p - V_m$) was given by equation (**12.9**) as:

$$V_f = \underbrace{mV_T \ln\left(\frac{I_{av}}{I_S}+1\right)}_{V_{f\_}} + \underbrace{mV_T \ln\left(\frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}\right)}_{V_{f\sim}}$$

This solution has given us the basis for calculating the error, and it tells us that the AC contribution is dependent only on the peak input voltage. That conclusion is somewhat opaque however, because it does not seem to offer a physical reason for the effect. All becomes clear when we use (**12.13**) as a substitution:

$$V_f = mV_T \ln\left(\frac{I_{av}}{I_S}+1\right) + mV_T \ln\left(\frac{I_p+I_S}{I_{av}+I_S}\right) \qquad (\mathbf{12.14})$$

Now it can be seen that the AC error term is governed by the peak to average current ratio; which is physically reasonable, and perhaps also obvious in retrospect.

A corollary of equation (**12.14**) is that the AC error disappears when the peak current is the same as the average current. This has a trivial meaning when both are zero; but it also implies that the error will disappear when the rate of change of the detector input voltage is so slow that the output can follow it. In this simplified derivation, we have prevented that from happening by specifying that the smoothing capacitor must be 'large'. This gives rise to a further and perhaps more interesting corollary; which is that the dynamic voltage error is independent of frequency provided that the capacitor is large. That point leads to a useful conclusion, as will now be explained.

It is possible to make extremely linear active detectors ('superdiode' circuits) by the use of high gain and negative feedback. Such detectors however have limited bandwidth. We can nevertheless calibrate a diode voltmeter against an active detector by comparing the readings when both are driven simultaneously by a signal generator working within the active detector's range. Furthermore, we can then vary the output voltage of the generator until the output of the diode detector agrees with that of an identical detector connected to an (unknown) RF voltage that we want to measure. The level adjustment of the reference generator can of course be done automatically. The reading given by the linear detector is then a measurement of the unknown RF voltage. This type of voltmeter is known as an 'amplitude tracking detector'.

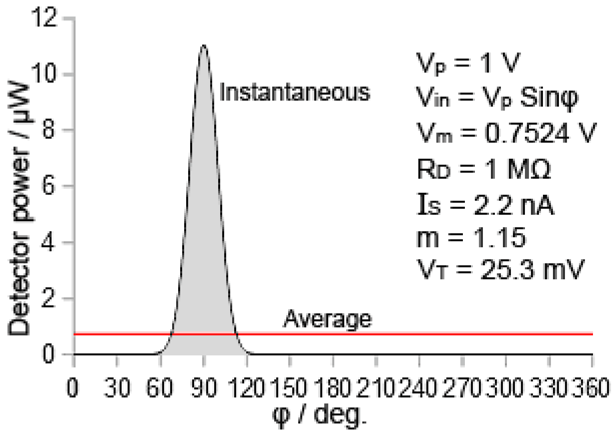
12.3 Detector power dissipation and input impedance

In section **12.1**, for the purpose of deriving the basic transfer function, it was assumed that the generator driving the detector has zero output impedance. This is unrealistic, of course, but we can adapt the model to deal with finite source impedance provided that we know the detector input impedance at the driving frequency. The actual input voltage is obtained from a potential divider formed by the source impedance and the input impedance (see section **1.6**).

A rough idea of the AC input resistance of a practical half-wave rectifier circuit was given in section **1.5**. Here we will carry out a more detailed analysis. The determination is a matter of applying the principle of conservation of energy, so that the power dissipated in the detector and its DC network is equal to that abstracted from the source.

The instantaneous power delivered to the detector is the product of the instantaneous input voltage, multiplied by the instantaneous input current. Thus:

$P_{inst} = V_{in}\, I_d = V_p\, I_d \sin\varphi$



Where, as was originally given as equation (**12.1**):

$$I_d = I_S\left[\exp\left(\frac{V_p \sin\phi - V_m}{m V_T}\right) - 1\right]$$

As before, V_p is the peak input voltage, and V_m is the detector DC output voltage (i.e., the measurable voltage). For illustration, the instantaneous power, for a 1N5711 detector with a 1 MΩ load and $V_p = 1$ V is shown in the graph above (see worksheet file **det_models.ods**, sheet 2).

The average power, from which an equivalent AC input resistance (i.e., the detector input impedance) can be calculated, is obtained by integrating the instantaneous power over a complete cycle of the input waveform and dividing by the integration range. Thus, if we also define the same composite parameters as were used in **section 12.1**, i.e.:

$K_m = \exp\{ -V_m / mV_T \}$ and $u = V_p / mV_T$

Then, assigning the symbol P_{det} to the average power (i.e., the power):

$$P_{det} = \frac{V_p I_S}{2\pi}\int_0^{2\pi} \sin\phi\left[K_m \exp(u \sin\phi) - 1\right] d\phi$$

Although this integral is not the same as in the expression for average current given in **section 12.1**, the discussion in that section is nevertheless relevant to its solution. Recall (as mentioned in the text following equation **12.3**) that all pure sinusoids average to zero over a complete cycle. Thus, if we were to multiply-out the part of the expression above that is to be integrated, we would end up with a term -sinφ, which averages to 0. Hence, the integral reduces to:

$$P_{det} = \frac{V_p I_S K_m}{2\pi}\int_0^{2\pi} \sin\phi\, \exp(u \sin\phi)\, d\phi$$

Now, as was previously done in order to solve the integral equation (**12.2**), we can expand the exponential as an infinite series. In this case however, we also need to multiply every term in the series by sinφ, which has the effect of increasing the index for $\sin^i\varphi$ by 1. Thus:

$$P_{det} = \frac{V_p I_S K_m}{2\pi} \int_0^{2\pi} \sum_{i=0}^{\infty} \frac{u^i \sin^{i+1}\phi}{i!} d\phi$$

As was also discussed previously, raising a sinusoid to an integer power results in a series of harmonics, plus a constant in those cases where the power is even. All of the harmonics, being sines and cosines of integer multiples of φ, vanish in the integration, leaving only the series of constants. Since the integral then contains no instances of φ, integration between 0 and 2π is reduced to a matter of multiplying the whole thing by 2π (which cancels the existing factor of 1/2π).

The series of constants is given by ${}^{2n}C_n / 2^{2n}$, where ${}^{2n}C_n$ is a binomial coefficient (see the discussion following **12.3**), and in this case, $2n = i+1$. To produce only the even terms, we can change the summation index to n and use the substitution: $i = 2n-1$. Note that for $i=0$, $i+1$ is odd, so the new series starts from n=1. Thus:

$$P_{det} = V_p I_S K_m \sum_{n=1}^{\infty} \frac{u^{2n-1} . {}^{2n}C_n}{(2n-1)!\, 2^{2n}}$$

The binomial coefficient is evaluated as: ${}^{2n}C_n = (2n)! / (n!)^2$. As before, the square of a factorial in the denominator should immediately alert us to the possibility that the series represents a Bessel function. Also, since the expression for the current involves a *zero-order* modified Bessel function of the first kind, and the matter of getting from a current to a power, in general, involves *raising the order* of an expression; we might reasonably be able to guess which Bessel function it is. Thus the solution in this case is a matter of proving the most obvious conjecture. Expanding the binomial coefficient gives:

$$P_{det} = V_p I_S K_m \sum_{n=1}^{\infty} \frac{u^{2n-1} (2n)!}{(2n-1)!\, 2^{2n} (n!)^2}$$

This expression can be simplified using the substitutions: $(2n)! = (2n-1)! \times 2n$ and $2^{2n} = 2 \times 2^{2n-1}$. The result is:

$$P_{det} = V_p I_S K_m \sum_{n=1}^{\infty} \frac{(u/2)^{2n-1}\, n}{(n!)^2}$$

The series summation (as expected) corresponds to the *first-order* modified Bessel function of the first kind [93], $I_1(u)$. Thus the solution for the total detector power, averaged over a complete cycle of the input waveform is:

$$P_{det} = V_p I_S K_m I_1(u)$$

After expansion of the composite parameters K_m and u, this becomes:

93 **Dwight** (already cited), p195, 813.2.

$P_{det} = V_p\, I_S\, \exp\left(\frac{-V_m}{mV_T}\right) I_1\left(\frac{V_p}{mV_T}\right)$	**12.15**

Now, if R_{Zin} represents the resistive component of the detector input impedance due to AC to DC conversion; then, since the power delivered to the detector (excluding losses due to causes other than conversion) is, by Joule's law:

$P_{det} = V_{in(RMS)}{}^2 / R_{Zin}$

and, for a sinusoidal input waveform:

$V_{in(RMS)} = V_p / \sqrt{2}$

Then

$R_{Zin} = V_p{}^2 / 2\, P_{det}$	**12.16**

Note however, that (**12.15**) is not the best method for calculating the detector power because it will run into the floating-point arithmetic upper range limit when V_p / mV_T is large. An unrestricted approach will be developed in the next two sections.

12.4 Diode power dissipation

The preceding derivation provides the basic solution to the detector input impedance problem; but in the matter of providing insight into the behaviour of the detector, it is somewhat unsatisfactory. In particular, it should be obvious that, for moderately large signal levels, the power dissipation in the detector is dominated by the dissipation in the load resistance, which is trivially obtainable as the product of the output voltage V_m and the average diode current I_{av}. The interesting information is that which governs the non-linear relationship between input voltage and input impedance, this being related to the dissipation that occurs in the diode. The diode power can be written:

$P_{diode} = P_{det} - P_{load} = P_{det} - V_m\, I_{av}$

Now recall that in section **12.1**, for the purpose of extracting the diode dynamic forward voltage, V_f, we defined a function $W_0(V_p / mV_T)$, which converts an exponential into a *zero-order* modified Bessel function of the first kind. Here, we will perform a similar trick by defining a function $W_1(V_p / mV_T)$, which converts an exponential into a *first-order* modified Bessel function of the first kind. Equation (**12.15**) then becomes:

$$P_{det} = V_p\, I_S\, \exp\left(\frac{-V_m}{mV_T}\right) W_1 \exp\left(\frac{V_p}{mV_T}\right)$$

Combining the two exponentials gives:

$$P_{det} = V_p\, I_S\, W_1 \exp\left(\frac{V_f}{mV_T}\right) \qquad (12.17)$$

and W_1 is defined as:

$$W_1 = \frac{I_1(V_p/mV_T)}{\exp(V_p/mV_T)}$$

The load power is given by $V_m \times I_{av}$, and I_{av} was given previously as equation (**12.6a**):

$$I_{av} = I_S\left[W_0 \exp\left(\frac{V_f}{mV_T}\right) - 1\right]$$

Thus the diode power is:

$$P_{diode} = V_p\, I_S\, W_1 \exp\left(\frac{V_f}{mV_T}\right) - V_m\, I_S\left[W_0 \exp\left(\frac{V_f}{mV_T}\right) - 1\right]$$

This can be rearranged to give:

$$P_{diode} = I_S\left[\exp\left(\frac{V_f}{mV_T}\right)\left(V_p\, W_1 - V_m\, W_0\right) + V_m\right] \qquad (12.18)$$

Also, a corollary of equation (**12.6a**) is that:

$$\exp\left(\frac{V_f}{mV_T}\right) = \left(\frac{I_{av}}{I_S} + 1\right)\frac{1}{W_0}$$

Substituting this into equation (**12.18**) and multiplying out gives:

$P_{diode} = (I_{av} + I_S) [(V_p W_1 / W_0) - V_m] + I_S V_m$

Where:

$W_1 / W_0 = I_1(V_p / mV_T) / I_0(V_p / mV_T)$

Thus:

$P_{diode} = (I_{av} + I_S)\left[V_p \frac{I_1(V_p/mV_T)}{I_0(V_p/mV_T)} - V_m\right] + I_S V_m$	**12.19**

This is the general expression for diode power, but it should also be noted that there exists an asymptotic form[94 95] of $I_1(u)$, which (like the asymptotic form of $I_0(u)$, equation 12.10) is applicable over the entire practical working range of an uncorrected detector (i.e., $V_p/mV_T > 10$). The asymptotic form can be written:

$$I_1(u) \approx \exp(u)\, P_{aI1}(u) / \sqrt{2\pi u} \qquad \text{. . . . . (12.20)}$$

Where $P_{aI1}(u)$ (polynomial used in the asymptotic form of I_1) is given by:

$$P_{aI1}(u) = 1 - \frac{4-1^2}{1!\,8u} - \frac{(4-1^2)(4-3^2)}{2!\,(8u)^2} - \frac{(4-1^2)(4-3^2)(4-5^2)}{3!\,(8u)^3} - \ldots\ldots \qquad (12.20p)$$

as $u \to \infty$, $P_{aI1}(u) \to 1$

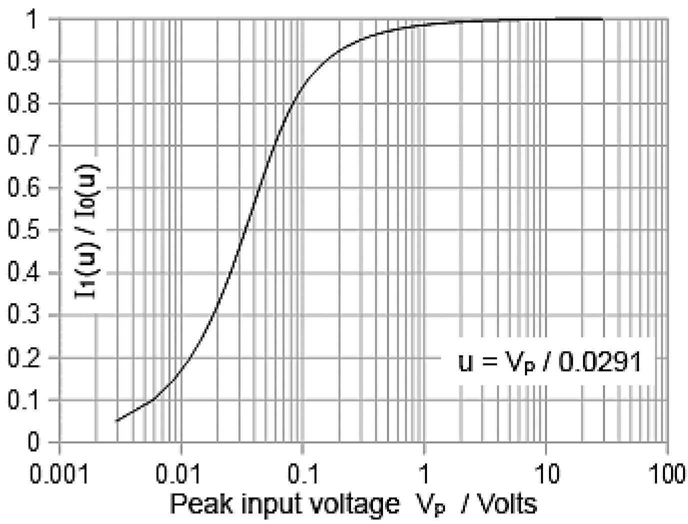
or alternatively:

$$P_{aI1}(u) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!\,(8u)^n} \prod_{i=1}^{n} \left[4-(2i-1)^2\right] \qquad (12.20s)$$

The asymptotic forms of the zero and first order modified Bessel functions are identical except for the polynomial used, which means that if we substitute them into (12.19) we end up with the ratio of the two polynomials, i.e.;

$$P_{diode} = (I_{av}+I_S)\left[V_p \frac{P_{aI1}(V_p/mV_T)}{P_{aI0}(V_p/mV_T)} - V_m\right] + I_S V_m \qquad V_p/mV_T > 10$$

The graph on the right shows the ratio of the two Bessel functions versus detector input voltage when $mV_T = 29.1$ mV (e.g., 1N5711 @ 20ºC). The calculation can be switched to use the ratio of the two polynomials when $V_p/mV_T = 10$, which corresponds to $V_p = 0.291$ V. This is not a large input voltage, but at this point the ratio has already reached 0.95, and it gets closer to 1 as the input voltage is further increased. Bearing in mind also that we are in the process of estimating the power dissipated in the diode, which is generally a small fraction of the total power delivered to the detector; it is apparent that for most purposes, it will be sufficient to assume that the ratio of the Bessel functions is 1. This leads to the approximation:



$$P_{diode} \approx (I_{av} + I_S)(V_p - V_m) + I_S V_m$$

94 **McLachlan** (already cited), p220, Table 14.
95 **Dwight** (already cited), p196, 814.2.

i.e.:

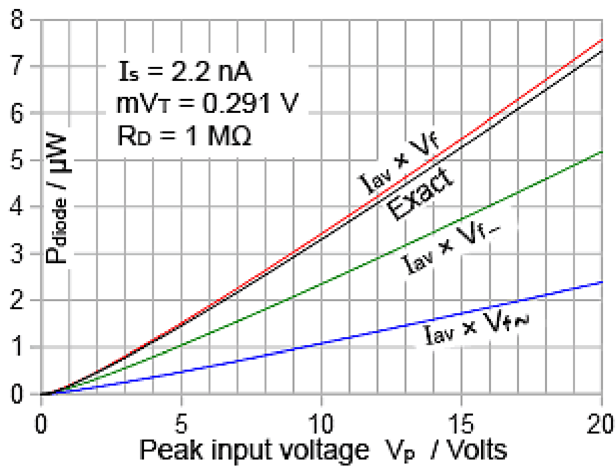
$P_{diode} \approx I_{av} (V_p - V_m) + I_S V_p - I_S V_m + I_S V_m$

and noting that V_p-V_m = V_f (the dynamic forward voltage drop) and that $I_S V_p$ is small (a few nano watts):

$P_{diode} \approx I_{av} V_f$	**12.21**

This is an extraordinary result because, although we might adopt the habit of imagining that V_f is a DC voltage present in the detector network; it is actually a transitory voltage (the peak detection error), which only exists for an instant in each cycle at the point when sinφ=1. Still, for moderately large inputs, it is apparent that we can estimate the diode power on the basis that it is an average voltage to be multiplied by the average diode current.

A graph of diode power vs. input voltage for a detector with a 1 MΩ load, mV_T = 29.1 mV and I_S = 2.2 nA is shown on the right. The curve marked 'Exact' is generated by equation (**12.19**), and the curve for for the approximation $I_{av} \times V_f$ (equation **12.21**) lies a little above it. The over-estimation occurs because the ratio of Bessel functions included in (**12.19**) has the effect of reducing the effective value of V_p slightly. The error is around 4% for reasonably large inputs (see spreadsheet **det_models.ods**, sheet 3), but since the diode power is a small proportion of the total in that case, the effect on input impedance estimation is fairly minor (see **section 12.6**).



In **section 12.1** it was shown that V_f can be split into two components, $V_{f\_}$ (DC) and $V_{f\sim}$ (AC). This separation also leads to two power components, which are shown on the graph. The curve for $I_{av} \times V_{f\_}$ shows that neglecting the dynamic (AC) component leads to a considerable underestimation of diode power.

12.5 Fast unrestricted computation of input impedance

With the use of asymptotic forms for large Bessel-function arguments, equation (**12.19**) is the best choice for the calculation of diode power. Adding back the load power to obtain the total converted power (AC to DC + harmonics) then gives:

$$P_{det} = (I_{av}+I_S)\left[V_p \frac{I_1(V_p/mV_T)}{I_0(V_p/mV_T)} - V_m\right] + I_S V_m + I_{av} V_m$$

Which turns into a remarkably compact expression:

$$P_{det} = (I_{av}+I_S)\, V_p \frac{I_1(V_p/mV_T)}{I_0(V_p/mV_T)}$$	**12.22**

Routines for the calculation of the Bessel functions I_1 and I_0 and their asymptotic forms are developed in **section 13** (calculation procedures). In **subsection 13.9** however, the various methods are combined into a single program that calculates the ratio. The program chooses initially between the small argument and asymptotic forms and then computes both the zero and first order functions from within the same program loop. This approach is probably about as efficient as it gets within a given programming environment, and is, of course, free from practical argument-range restrictions. The detector power is then given by:

$P_{det} = (I_{av} + I_S)\, V_p$ **RatioI1_0**(V_p/mV_T)

where **RatioI1_0**( ) is the function Combining this result with equation (**12.16**) gives the input impedance:

$R_{Zin} = V_p{}^2 / 2\, P_{det}$

i.e.:

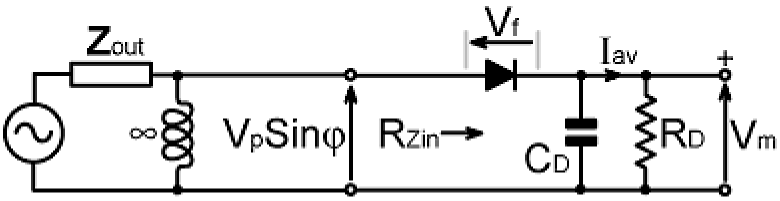
$$R_{Zin} = \frac{V_p}{2(I_{av}+I_S)} \frac{I_0(V_p/mV_T)}{I_1(V_p/mV_T)}$$	**12.23**

12.6 Using diode dynamic resistance to estimate input impedance

In **section 1.5**, the effective diode resistance under dynamic conditions, R_{diode}, was introduced as a term in the total detector DC resistance, the latter being required for the determination of input impedance. With the analysis given in **sections 12.1 to 12.4**, we can, of course, now estimate the dynamic diode resistance as V_f / I_{av}, and this allows us to derive a useful approximate expression for the input impedance.

The component of the detector input impedance due to AC-DC power conversion was given as equation (**12.16**):

$R_{Zin} = V_p{}^2 / 2\, P_{det}$



which can be rearranged:

$P_{det} = V_p{}^2 / 2\, R_{Zin}$ . . . . . (**12.24**)

Also, using the approximation (**12.21**):

$P_{diode} \approx I_{av} V_f$

we can write:

$P_{det} = P_{diode} + P_{load} \approx I_{av} V_f + I_{av} V_m = V_p I_{av}$

where

$I_{av} = V_m / R_D = (V_p - V_f) / R_D$

Therefore:

$P_{det} \approx V_p (V_p - V_f) / R_D$

i.e.:

$P_{det} \approx V_p^2 (1 - V_f / V_p) / R_D$

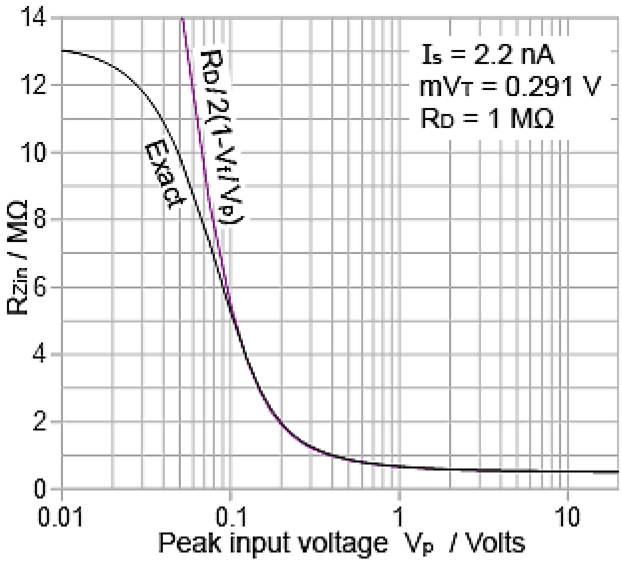
Equating this with (**12.24**) gives:

$V_p^2 (1 - V_f / V_p) / R_D \approx V_p{}^2 / 2\, R_{Zin}$

i.e.:

$R_{Zin} \approx R_D / [\, 2 (1 - V_f / V_p) \,]$	**12.25**

The graph on the right shows a comparison between this formula and the exact expression (i.e.; using **12.23**) (see spreadsheet **det_models.ods**, sheet 3). The approximation does not describe the behaviour of the input impedance at very low input voltages, but the two methods agree within 4.1% at 0.1 V, 1.2% at 1 V and 0.12% at 10 V. The simple formula also illustrates the asymptotic behaviour of the input impedance, which tends towards R_D /2 as the input becomes very large.



We can, incidentally (according to the 2:1 impedance transformation rule) also define R_{Zin} as:

$R_{Zin} = (R_D + R_{diode}) / 2$	**12.26**

where: $R_{diode} \approx V_f / I_{av}$ and $I_{av} = (V_p - V_f) / R_D$

Hence:

$R_{diode} \approx R_D V_f / (V_p - V_f)$

i.e.:

$R_{diode} \approx R_D / (V_p / V_f - 1)$	**12.27**

The exact expression for the diode dynamic resistance is obtained by rearranging (**12.26**):

$R_{diode} = 2 R_{Zin} - R_D$	**12.28**

R_{Zin} being obtained from (**12.23**). This final expression is however probably only of use for the purpose of evaluating approximations; since the point of calculating R_{diode} is to determine the input impedance, and so there is no need to do it if the input impedance is known.

13. Calculation procedures for the simple diode voltmeter

The solution to the basic diode voltmeter problem, i.e., the relationship between AC input voltage and DC output voltage, is obtainable from equation (**12.8**):

$$V_f = mV_T\left[\ln\left(\frac{V_m}{I_S\,R_D}+1\right)+\ln\left(\frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}\right)\right] \qquad V_m = V_p - V_f$$

Where V_m is the DC output (measurement) voltage, V_p is the peak input voltage, and V_f is the effective diode forward voltage drop under dynamic conditions.

The expression above is however tautologous, in that any substitution to reduce the three principal variables to two (e.g., by replacing V_m with $V_p - V_f$) results in the need to input a value of a voltage that we want to find before it will give an output of that quantity. This means that there are no closed-form analytical solutions. Nevertheless, it is a straightforward matter to set up the various parts of the equation in the columns of a spreadsheet, and then input trial solutions until both instances of the desired quantity agree. That method is, of course, tedious and unsatisfactory; and so the problem is that of how to automate the process (i.e., turn it into a set of algorithms). We will moreover want solutions in either direction, i.e.; so that we can put in a value of V_p and obtain V_f and V_m; or so that we can put in a value of V_m and obtain V_f and V_p. There is also a particular difficulty in this case, which is that the argument of a logarithm function-call must never become less than zero; and so for small values of V_m we will be performing an iterative procedure close to a point at which a runtime error will occur.

A further issue is that the the problem involves exponentials and modified Bessel functions of large argument. In Open Office, for example, the double-precision floating-point range is from 4.941×10^{-324} to 1.798×10^{308}, and a simulation for a 1N5711 detector operating at 20°C (for example) hits the upper limit at $V_p = 20.62$ V. The built-in spreadsheet function for I_0 is moreover, not accessible from the macro programming environment. It is therefore useful to have routines for the modified Bessel function and its asymptotic form.

13.1 Modified Bessel function, first kind, zero order

A calculation procedure involving the summation of an infinite series should, if possible, confine itself to the use of the elemental operations of addition and multiplication. This is advisable because most interpreters and compilers handle operations such as 'raising to a power' by exponentiation, even when the power is an integer. Such composite operations can incur significant rounding errors (which accumulate as the summation progresses), whereas simply multiplying a quantity by itself is less affected. Greatest computational efficiency is also obtained by exploiting any recursion relations[96] that might exist; a recursion factor being a quantity by which a given term in a series can be multiplied in order to obtain the next term. The use of elemental operations usually depends on the existence of such a factor.

The infinite series for the modified zero-order Bessel function or the first kind was introduced earlier and has the form:

96 **Practical considerations in the calculation of Kelvin functions and complete elliptic integrals**. Robert S Weaver, Oct. 2009. Available from: http://www.g3ynh.info/zdocs/magnetics/ and http://electronbunker.ca/CalcMethodsRef.html

$$I_0(x) = \sum_{n=0}^{\infty} \frac{(x/2)^{2n}}{(n!)^2}$$ McLachlan, p200, eq^n. 153. Dwight, p195, 813.1

This has a simple recursion factor, which can be spotted by writing down the first few terms:

term(0) = 1

term(1) = $(x/2)^2 / (1^2)$

term(2) = $(x/2)^2 . (x/2)^2 / (1^2 . 2^2)$

term(3) = $(x/2)^2 . (x/2)^2 . (x/2)^2 / (1^2 . 2^2 . 3^2)$

etc.

Hence the overall pattern is:

term(n) = term(n-1) × $(x/2)^2$ / (n × n)

Furthermore, the quantity $(x/2)^2$ can be pre-calculated as x x /4 .

The resulting algorithm, coded in Open Office Basic, is shown below:

```
Function I0Bessel(byval x as double)as double
'Calculates modified Bessel function, 1st kind, zero order, I0(x)
I0Bessel = 1
If x <= 0 then exit function
Dim xx4 as double, term as double, sum as double, rcf as double
Dim n as integer
term = 1
sum = term
xx4 = x*x/4
n = 0
do
  n = n+1
  rcf = xx4/(n*n)
  term = term*rcf
  sum = sum + term
loop until term/sum < 1E-12
I0Bessel = sum
end function
```

This routine works for arguments up to 708, as does the built-in spreadsheet function; and it agrees with the built-in function to at least 12 significant figures (see spreadsheet file **det_models.ods**, sheets 1 & 2).

13.2 Polynomial used in the asymptotic form, first kind, zero order

For arguments greater than about 10, the asymptotic form of the modified Bessel function $I_0(x)$ can be used. It was given earlier as equation (**12.10**)

$$I_0(x) = \frac{\exp(x)}{\sqrt{2\pi x}}\left[1 + \frac{1^2}{1!\,8x} + \frac{1^2\,3^2}{2!\,(8x)^2} + \frac{1^2\,3^2\,5^2}{3!\,(8x)^3} + \ldots\ldots\right] \qquad x \geq 10$$

Dwight, p196, 814.1.
McLachlan, p220, table 13

This form will not extend the argument range of the function itself because it is the output, rather than some internal variable, that falls outside the floating-point arithmetic range. We can however, use it in the AC part of the expression for V_f, in which case it eliminates the exponential and vastly increases the calculation range.

The AC part of V_f, originally given as equation (**12.9ac**), is:

$$V_{f\sim} = mV_T \ln\left(\frac{\exp(V_p/mV_T)}{I_0(V_p/mV_T)}\right)$$

Using the exponential form, with $x = V_p / mV_T$, this becomes (**12.10ac**):

$V_{f\sim} = mV_T \ln\{ \surd(2\pi\, x) / P_{aI0}(x) \}$

where $P_{aI0}(x)$ is the polynomial appearing in the asymptotic form, i.e.:

$$P_{aI0}(x) = 1 + \frac{1^2}{1!\,8x} + \frac{1^2\,3^2}{2!\,(8x)^2} + \frac{1^2\,3^2\,5^2}{3!\,(8x)^3} + \ldots\ldots$$

We can approach the problem of calculating this series as before:

Term(0) = 1

Term(1) = $(2\times1 - 1)^2 / (1 . 8x)$

Term(2) = $(2\times1 - 1)^2 . (2\times2 - 1)^2 / (1\times2 . 8x . 8x)$

Term(3) = $(2\times1 - 1)^2 . (2\times2 - 1)^2\ (2\times3 - 1)^2 / (1\times2\times3 . 8x . 8x . 8x)$

etc. Hence:

Term(n) = Term(n-1) × (1/8x) × (2n-1) × (2n-1) / n

An algorithm that performs the calculation is shown below. It is restricted to arguments ≥ 9.6 in order to prevent floating-pont underflow errors (in the OO Basic programming environment. Other platforms may differ). Its output ranges from 1.013 to 1 as x goes from 10 to ∞.

```
Function PaI0(byval x as double)as double
'Calculates the polynomial used in the asymptotic form of I0(x)
PaI0 = 1
If x < 9.6 then exit function          'floating-point underflow occurs for arguments < 9.6
Dim x8 as double, term as double, sum as double, rcr as double
Dim n as integer
term = 1
sum = term
x8 = 1/(8*x)
n = 0
do
  n = n+1
  rcr = x8*(2*n-1)*(2*n-1)/n
  term = term*rcr
  sum = sum + term
loop until term/sum < 1E-9
PaI0 = sum
end function
```

A comparison of calculations of $V_{f\sim}$ using the small argument formula and the polynomial formula is given in the spreadsheet file **det_models.ods**, sheet 5. The two expressions agree to within a picovolt for arguments > 60 ; except that the small argument formula, as mentioned earlier, fails for arguments > 708 ; whereas the polynomial version was tested initially with arguments up to 35481 without error. The latter corresponds to just over 1030 V peak input for the 1N5711 example, not that a small signal diode can survive this.

13.3 Determining output voltage from peak input voltage

When the input voltage V_p is given, and the exercise is to calculate the forward voltage (or detector error) V_f and the output voltage V_m, then the AC part of V_f can be determined analytically. The relevant formulae are (**12.9ac**) for values of V_p / mV_T up to (say) 60, and (**12.10ac**) for larger arguments. Routines for use in the evaluation of either form are given above. The expression for V_f then reduces to:

$$V_f = V_{f\sim} + mV_T \ln\left(\frac{V_p - V_f}{I_S R_D} + 1\right)$$

In order to solve this equation for V_f, we can start by giving different symbols to the two instances, i.e.:

$$y = V_{f\sim} + mV_T \ln\left(\frac{V_p - x}{I_S R_D} + 1\right) \qquad \text{and} \quad x = V_f \quad \text{when} \quad x = y \qquad (\mathbf{13.1})$$

If we set-up this equation in a spreadsheet, it is a simple matter to insert a value of x into the right-hand side and see what value of y comes out. The input x can then be adjusted until it agrees with y to a number of significant figures deemed sufficient, and then x (or y) is the solution. The point however, is to find a procedure that will work automatically and never fail; regardless of the peak input voltage V_p , the parameters mV_T and $I_S R_D$, and the initial, intermediate and final values of x.

A starting point for guessing x is to note that, by definition, it can never be smaller than $V_{f\sim}$. Hence it is safe to begin by setting $x = V_{f\sim}$. The simplest automatic strategy for finding the solution is then to make x equal to the value of y that comes out, and try again. The process is repeated until x and y converge.

In fact, in a test with $mV_T = 29.1$ mV and $I_S R_D = 2.2$ mV, the simple strategy worked well for input voltages (V_p) above about 200 mV. For smaller inputs however, it failed miserably due to overshoot. Feeding back the first value of y caused $V_p - x$ to be < -1. That caused the argument of the log function to be < 0, resulting in a program error. Furthermore, the reason for developing the program is to explore the behaviour of the detector in the poorly-characterised region of the diode forward-conduction threshold.

In order to make the procedure stable for small inputs, we need to find a more accurate way of estimating the required value of x. To do that, we can analyse the adjustment process as follows:

When $x \rightarrow x + \delta x$, $y \rightarrow y + \delta y$

A solution occurs when:

$$x + \delta x = y + \delta y$$

If this were a linear system, then the shift in y would be equal to the shift in x multiplied by the rate rate of change of y with respect to x ; i.e., for linear or infinitesimal change:

$$\delta y = \delta x \, \partial y/\partial x$$

(where $\partial y/\partial x$ is the partial derivative of y with respect to x , i.e., the function is differentiated with all other variables treated as constants). In this case, of course, the system is not linear; but the differential relationship between δy and δx will nevertheless become approximately true as we approach a solution; and it will become exact at the point at which a solution is found because the required shift is then infinitesimal. Thus, in the vicinity of a solution, we expect:

$$x + \delta x \approx y + \delta x \, \partial y/\partial x$$

This can be rearranged to give an estimate of the required adjustment:

$$\delta x \approx (x - y) / (\partial y/\partial x - 1)$$

The quantity $x - y$ moreover provides the termination criterion for the iteration process, i.e., we can exit the program loop when it becomes suitably small.

The derivative of equation (**13.1**) is easily obtained by noting that $d(\ln\{x\})/dx = 1/x$ and applying the chain rule. Thus:

$$\partial y/\partial x = mV_T \, (-1 / I_S R_D) / [1 + (V_p - x) / I_S R_D]$$

Calculating the required shift by using the derivative reduces the minimum input voltage for successful iteration by about an order of magnitude in comparison to the simple procedure. There is however, still a tendency to overshoot for very small inputs, this being due to the inaccuracy of the derivative in the first round of iteration. The problem is easily solved however, by adding an extra nested program loop. In this loop, δx is added to x, and the argument of the logarithm is tested to see if it is greater than 1. If the argument is too small, the original value of x is restored, δx is halved, and the program returns to the point at which δx is added to x. Thus δx is successively reduced until the log argument is in range.

The complete algorithm is shown below. It has been tested for values of V_p down to 1 pV (well outside any practical measurement range), at which point it is still both stable and convergent. Note that I_S is called Isat in the program, because 'is' is a Basic keyword.

```
Function DVfp2m(RD as double, Vp as double, mVT as double, Isat as double) as double
'calculates voltage error (forward drop) of simple diode peak detector.
'Calls the functions I0Bessel( ) and PaI0( ).
DVfp2m = 0
if Vp <= 0 then exit function
Dim u as double, Vfac as double, x as double, y as double, Vsr as double
Dim kd as double, arg as double, der as double, deltax as double, xold as double
'Calculate Vf~.  Use the asymptotic form for arguments >= 60.
u = Vp/mVT
if u < 60 then
   Vfac = mVT*log( exp(u) / I0Bessel(u) )
else
   Vfac = mVT*log( sqr(2*pi*u) / PaI0(u) )
end if
'Set up starting values and composite parameters
x = Vfac
Vsr = Isat*RD
arg = 1 + (Vp-x)/Vsr
kd = -mVT/Vsr
'Calculate the difference between x and y, the derivative, and the estimated shift in x
do
   y = mVT*log(arg) + Vfac
   diff = x-y
   der = kd/arg
   deltax = diff/(der-1)
'Apply shift to x and check that the log argument is valid.  If not, reduce the shift and try again.
   do
      xold = x
      x = x + deltax
      arg = 1 + (Vp-x)/Vsr
      if arg <=1 then
         deltax = deltax/2
         x = xold
      end if
  loop until arg > 1
loop until abs(diff) < 1E-12
DVfp2m = x
end function
```

13.4 Determining peak input voltage from output voltage

The computational methods described so far are useful for modelling detectors; but for practical measurement purposes, the matter of interest is that of inferring the AC input voltage (V_p) from the DC output. This problem is made difficult by the fact that we cannot pre-calculate the the AC component of the diode forward voltage ($V_{f\sim}$), because we need to know V_p in order to do it. Also, the unknown is used in the argument of a modified Bessel function, and the choice of calculation method depends on that argument.

Small voltages

To solve the peak input problem for small voltages, it is probably easiest not to use equation (**12.8**) but to go back to its precursor, equation (**12.5**):

$$I_{av} = \frac{V_m}{R_D} = I_S\left[\exp\left(\frac{-V_m}{mV_T}\right) I_0\left(\frac{V_p}{mV_T}\right) - 1\right]$$

This can be rearranged into a form that would be analytical but for want of an inverse of the modified Bessel function.

$$I_0\left(\frac{V_p}{mV_T}\right) = \left(\frac{V_m}{I_S\, R_D} + 1\right)\exp\left(\frac{V_m}{mV_T}\right)$$

The right-hand side of this expression can be calculated, and if we call it y (say), then:

$V_p = mV_T\, \mathrm{AntiI}_0(y)$

where $\mathrm{AntiI}_0(y)$ is the inverse of the modified Bessel function of the first kind, zero order.

Modified Bessel functions have a single-valued inverse because, unlike the ordinary Bessel functions, they are not undulatory. Indeed, the curves for the I_n (first kind) functions roughly resemble exponential growth. Inverse calculation routines however, do not appear to be standard library functions. That problem was solved by writing a suitable routine; which will be described in **section 13.6**.

Large voltages

A limitation of the solution given above, of course, is that it contains an exponential and will easily exceed the floating-point computation range in typical applications. To circumvent that problem, we must find an expression that eliminates the exponential. That can be done by substituting the asymptotic form of the modified Bessel function (**12.10** - **12.10ac**) into the AC part of equation (**12.8**), thereby producing the expression:

$$V_f = mV_T\left[\ln\left(\frac{V_m}{I_S\, R_D} + 1\right) + \ln\left(\frac{\sqrt{2\pi x}}{P_{aI0}(x)}\right)\right] \qquad x = V_p / mV_T \quad , \quad V_m = V_p - V_f$$

We already have a routine for the polynomial $P_{aI0}(x)$ (see **section 13.2**), and we know that it converges for arguments greater than about 10 and does not deviate greatly from 1. Therefore it can be seen that solutions based on the expression above will be viable for x between 10 and the floating-point limit.

Now, substituting for V_f and using $x = V_p/mV_T$ throughout we get:

$$x - \frac{V_m}{mV_T} = \ln\left(\frac{V_m}{I_S R_D} + 1\right) + \ln\left(\frac{\sqrt{2\pi x}}{P_{aI0}(x)}\right)$$

This can be rearranged to put all instances of x on one side:

$$\ln\left(\frac{V_m}{I_S R_D} + 1\right) + \frac{V_m}{mV_T} = x - \ln\left(\frac{\sqrt{2\pi x}}{P_{aI0}(x)}\right)$$

The left-hand side can now be calculated, and if we call it (say) y , we have:

$y = x - \ln[\ \sqrt{2\pi x} / P_{aI0}(x)\ ]$

The logarithm can also be separated into terms, giving:

$y = x - (½) \ln(2\pi) - (½) \ln(x) + \ln[\ P_{aI0}(x)\ ]$

This can be solved iteratively by making an estimate of x (x_1 say), which can be inserted into the formula above to produce an approximation to y (y_1 say). Also notice that y is a good first approximation for x because the first term on the right is much larger than all of the others. The required change in y (δy) can then be used to estimate the required change in x (δx) via the derivative dy/dx, i.e.:

$\delta y = y - y_1 \approx \delta x\ dy/dx$

Thus:

$\delta x \approx (\ y - y_1\ ) / (\ dy/dx\ )$

Note that $1/(dy/dx)$ is not the same as dx/dy because the original relationship is non-linear. The differential relationship does however become exact at the point at which the solution is found, and so $y - y_1 \rightarrow 0$ is the termination criterion for the iteration process. The derivative is easily obtained by applying the chain rule to the final term. Thus:

$dy/dx = 1 - 1/(2x) + [\ dP_{aI0}(x)/dx\ ] / P_{aI0}(x)$

One additional mathematical function is required, and that is the derivative of the polynomial. It is easily obtained by differentiating the series term by term, and a suitable calculation routine is given in **section 13.5**. Its contribution to the overall derivative is very small and negative; but it must be included if the calculated shift in x is to vanish at the point of solution.

Overall calculation

The overall calculation routine using the techniques just described is given below. Note that it returns the detector forward voltage drop (i.e., the detector error), which can be added to the output voltage to determine the peak input voltage.

For detector outputs of less than 1.5 V, the inverse modified Bessel function method is used, the main calculation being performed by a separate routine called Anti_I0(). For higher voltages, the iterative method is used, and this involves calls to calculate the polynomial and its derivative. Procedure development, prior to coding, was carried out in the spreadsheet **det_models.ods**, sheet 3. The program was tested, with $I_S R_D = 2.2$ mV and $mV_T = 29.1$ mV, for V_m values between 1 pV and 100 kV. The changeover point at $V_m = 1.5$ V will not lead to out-of-range polynomial or polynomial derivative arguments when used in conjunction with modern semiconductor diodes, but it can easily be changed to deal with unusual diodes or large diode stacks if necessary.

```
Function DVfm2p(RD as double, Vm as double, mVT as double, Isat as double) as double
'calculates voltage error of simple diode peak detector from the output voltage Vm.
'Calls the functions: Anti_I0( ), PaI0( ), DerPaI0( )
DVfm2p = 0
if Vm <= 0 then exit function
Dim y as double, x as double, Vp as double
if Vm < 1.5 then
  y = (1+Vm/(Isat*RD))*exp(Vm/mVT)
  x = Anti_I0(y)
else
dim y1 as double, diff as double, deltax as double, poly as double, deriv as double
  y = Vm/mVT +log(1+Vm/(Isat*RD))
  x = y
  do
    poly = PaI0(x)
    y1 = x - log(2*pi)/2 -log(x)/2 +log(poly)
    diff = y - y1
    deriv = 1 - 1/(2*x) + DerPaI0(x)/poly
    deltax = diff / deriv
    x = x + deltax
  loop until abs(diff) < 1E-9
end if
Vp = mVT*x
DVfm2p = Vp-Vm
end function
```

13.5 Derivative of the asymptotic form polynomial, first kind, zero order

The polynomial used in the asymptotic (large argument) form of the zero order, first kind, modified Bessel function (see **section 13.2**) is given by the series:

$$P_{aI0}(x) = 1 + \frac{1^2}{1!\,8x} + \frac{1^2\,3^2}{2!\,(8x)^2} + \frac{1^2\,3^2\,5^2}{3!\,(8x)^3} + \ldots\ldots$$

Differentiating this, term by term, gives:

$$dP_{aI0}(x)/dx = -\frac{1\,.\,1^2}{1\,.\,8\,x^2} - \frac{2\,.\,1^2\,3^2}{1\,.\,2\,.\,8^2\,x^3} - \frac{3\,.\,1^2\,3^2\,5^2}{1\,.\,2\,.\,3\,.\,8^3\,x^4} - \frac{4\,.\,1^2\,3^2\,5^2\,7^2}{1\,.\,2\,.\,3\,.\,4\,.\,8^4\,x^5} - \ldots\ldots$$

Thus:

Term(0) = $-1/8x^2$

Term(1) = $(-1/8x^2)\ 1^2\ 3^2 / 1\ .\ 8x$

Term(2) = Term(1) $\times\ 5^2 / 2\ .\ 8x$

Term (n) = Term(n-1) $\times\ (2n+1)^2 / 8x\ n$

The calculation procedure is given below. Note that in order to avoid underflow errors, the termination criterion is set to 10^{-9} for arguments above 13.4, and 10^{-6} for arguments between 9.7 and 13.4. This applies to OO Basic. Other programming environments may differ.

```
Function DerPaI0(byval x as double)as double
'Calculates the derivative PaI0'(x)
DerPaI0 = 0
If x < 9.7 then exit function
Dim x8 as double, term as double, sum as double, rcf as double, test as double
test = 1E-9
If x < 13.4 then test = 1E-6
term = -1/(8*x*x)
sum = term
x8 = 1/(8*x)
Dim n as integer
n = 0
do
  n = n+1
  rcf = x8*(2*n+1)*(2*n+1)/n
  term = term*rcf
  sum = sum + term
loop until term/sum < test
DerPaI0 = sum
end function
```

13.6 Inverse modified Bessel function, first kind, zero order

As mentioned earlier, the modified Bessel functions all have a unique inverse; and as has been found here, it is sometimes useful to be able to determine it. No strenuous effort has been made to find the most efficient method for the inverse of the first kind, zero order, but the following approach works well.

If $y = I_0(x)$, then let us define $x = \text{AntiI}_0(y)$ and find its value iteratively.

Given y, we first need to make an estimate of x (x_1 say). The estimate does not have to be particularly accurate, provided that the iteration procedure converges from its starting point. Using x_1 we can calculate the corresponding value of y (y_1 say) using the routine for $I_0(x)$ that we already have, i.e.;

$$y_1 = I_0(x_1)$$

We can now estimate the required shift in x using the procedure described in the previous section, i.e.:

$$\delta x \approx (y - y_1) / (dy/dx)$$

where dy/dx is, in this case, the first derivative of the modified Bessel function, first kind, *zero order*; which just happens to be the modified Bessel function of the first kind, *first order*, i.e.:

$$dy/dx = dI_0(x)/dx = I_1(x)$$

A routine for calculating this function will be given in the next section. δx is added to x_1 and a new value of y_1 is calculated. This iteration process is continued until δx becomes very small.

Prior to coding, the procedure was set up for manual implementation in the spreadsheet **det_models.ods**, sheet 4. It was found that for values of y up to about 10, the routine would converge rapidly if the initial estimate for x was always set to 1. For very large values of y however, this leads to much unnecessary calculation; and as the floating-point limit is approached it can result in overshoot and runtime error.

An improved estimation procedure for large x values was obtained by noting that the polynomial used in the asymptotic form of $I_0(x)$ never deviates greatly from 1. Therefore a reasonably good estimate for large arguments is given by omitting it. Thus, for $x > 10$:

$$y = I_0(x) \approx \exp(x) / \sqrt{2\pi x}$$

Unfortunately, this cannot be rearranged to give x on its own, but if we set a value of 10 to the instance of x that occurs in the square root bracket we get:

$$\exp(x) \approx y \sqrt{20\pi}$$

i.e.:

$$x \approx \ln(y) + (½) \ln(20\pi)$$

It was found that for very large values of y, approaching the floating-point limit, this simple formula estimates x to an accuracy of about 0.3%. Hence the calculated δx is extremely accurate, and a termination criterion of $|\delta x| < 1\times10^{-12}$ has the iteration completed within two or three cycles. This high accuracy holds until y is less than about 400 ($x < 7.93$), but the iteration procedure will still terminate correctly if it is used down to $y = 1$. For smaller inputs the estimate asymptotes to $x \approx \ln(20\pi)/2 = 2.07$. Since x lies between 0 and 1.81 for $y < 2$, it was felt that setting $x = 1$ for $y < 2$ would probably eliminate at least one cycle of iteration in the lower input ranges.

The resulting program routine is shown below. It is accurate to at least 12 decimal places, and in the OO Basic programming environment, using the companion functions **I0Bessel**( ) and **I1Bessel**( ) , it works for inputs up to 9.1×10^{305}.

```
Function Anti_I0(byval y as double) as double
'Calculates the inverse of the modified Bessel function, first kind, zero order
'Calls functions I0Bessel( ) and I1Bessel( )
Anti_I0 = 0
if y <= 1 then exit function
Dim x as double, y1 as double, deriv as double, deltax as double
if y < 2 then
  x = 1
else
  x = log(y) + log(20*pi)/2
end if
do
  y1 = I0Bessel(x)
  deriv = I1Bessel(x)
  deltax = (y-y1)/deriv
  x = x + deltax
loop until abs(deltax) < 1E-12
Anti_I0 = x
end function
```

13.7 Modified Bessel function, first kind, first order

The first derivative of the modified Bessel function $I_0(x)$, as used in the inverse procedure described above, is the same as the modified Bessel function $I_1(x)$. This first order function is also used in calculating diode power and detector input impedance. The series form is given by Dwight[97], but it can also be obtained by differentiating the series for $I_0(x)$ (see section **13.1**) term by term. The result is:

$$dI_0(x)/dx = I_1(x) = \frac{(x/2)}{1} + \frac{(x/2)^3}{1^2\,2} + \frac{(x/2)^5}{1^2\,2^2\,3} + \frac{(x/2)^7}{1^2\,2^2\,3^2\,4} + \ldots.. = \sum_{n=1}^{\infty} \frac{n\,(x/2)^{2n-1}}{(n!)^2}$$

From which we can see that:

Term(1) = (x/2) / 1

Term(2) = (x/2) $(x/2)^2$ / 1 . 1 . 2

Term(3) = (x/2) $(x/2)^2$ $(x/2)^2$ / 1 . 1 . 2 . 2 . 3

etc.

Term (n) = Term(n-1) × $(x/2)^2$ / n (n+1)

An OO Basic calculation procedure, obtained by changing two lines in the zero order modified Bessel function program given earlier, is shown below:

```
Function I1Bessel(byval x as double)as double
'Calculates modified Bessel function, 1st kind, first order, I1(x) = I0'(x)
I1Bessel = 0
If x <= 0 then exit function
Dim xx4 as double, term as double, sum as double, rcf as double
Dim n as integer
term = x/2
sum = term
xx4 = x*x/4
n = 0
do
  n = n+1
  rcf = xx4/( n*(n+1) )
  term = term*rcf
  sum = sum + term
loop until term/sum < 1E-12
I1Bessel = sum
end function
```

97 **Dwight** (already cited), p195, 813.2.

13.8 Polynomial used in the asymptotic form , first kind, first order

For arguments greater than about 10, the asymptotic form of the Bessel function $I_1(x)$ can be used. This was given earlier as equation (**12.20**):

$I_1(x) \approx \exp(x)\, P_{aI1}(x) / \sqrt{2\pi\, x}$

Using this form will not extend the argument range for the Bessel function itself, because it is the returned value that falls outside the floating-point range when the argument is large. The detector input impedance determination however uses a ratio of Bessel functions, $I_1(x)/I_0(x)$. The two asymptotic forms have a common factor $\exp(x)/\sqrt{2\pi\, x}$; and so the exponentials cancel when their ratio is taken, thereby greatly increasing the argument range. The Polynomial $P_{aI1}(x)$ is given by:

$$P_{aI1}(x) = 1 - \frac{4-1^2}{1!\; 8x} - \frac{(4-1^2)(4-3^2)}{2!\; (8x)^2} - \frac{(4-1^2)(4-3^2)(4-5^2)}{3!\; (8x)^3} - \ldots\ldots$$

Thus:

Term(0) = 1

Term(1) = $(4-1^2)$ / 1 . (-8x)

Term(2) = $(4-1^2)$ $(4-3^2)$ / 1 . 2 . (-8x) . (-8x)

Term(3) = $(4-1^2)$ $(4-3^2)$ $(4-5^2)$ / 1 . 2 . 3 . (-8x) . (-8x) . (-8x)

Term(n) = Term(n-1) . [4 - $(2n-1)^2$] / n (-8x)

An OO Basic function is given below:

```
Function PaI1(byval x as double)as double
'Calculates the polynomial used in the asymptotic form of I1(x).
PaI1 = 1
If x < 9.6 then exit function
Dim x8 as double, term as double, sum as double, rcf as double
Dim n as integer
term = 1
sum = term
x8 = -1/(8*x)
n = 0
do
  n = n+1
  rcf = x8*(4-(2*n-1)*(2*n-1))/n
  term = term*rcf
  sum = sum + term
loop until abs(term/sum) < 1E-9
PaI1 = sum
end function
```

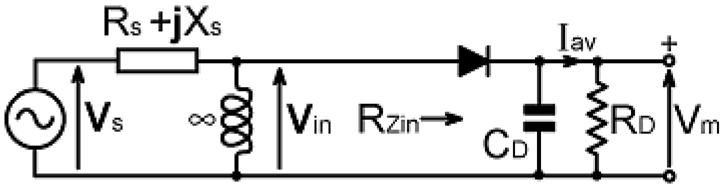
13.9 Ratio of modified Bessel functions, first order / zero order

The ratio of modified Bessel functions $I_1(x)/I_0(x)$ is used for calculating the input impedance of a diode detector. The routine below calculates both Bessel functions in the same program loop for arguments < 15, and calculates both asymptotic form polynomials in the same program loop for larger arguments.

```
Function RatioI1_0(byval x as double) as double
'Calculates ratio of Bessel funcs; I1(x)/I0(x).
RatioI1_0 = 0
If x <= 0 then exit function
Dim term1 as double, sum1 as double, rcf1 as double,
Dim term0 as double, sum0 as double, rcf0 as double
Dim n as integer
n = 0
If x < 15 then
Dim xx4 as double
  term1 = x/2
  term0 = 1
  sum1 = term1
  sum0 = term0
  xx4 = x*x/4
  do
    n = n+1
    rcf1 = xx4/( n*(n+1) )
    rcf0 = xx4/(n*n)
    term1 = term1*rcf1
    term0 = term0*rcf0
    sum1 = sum1 + term1
    sum0 = sum0 + term0
  loop until term1/sum1 < 1E-9 and term0/sum0 < 1E-9
else
Dim x8 as double
  term1 = 1
  term0 = 1
  sum1 = term1
  sum0 = term0
  x8 = 1/(8*x)
  do
    n = n+1
    rcf1 = -x8*(4-(2*n-1)*(2*n-1))/n
    rcf0 = x8*(2*n-1)*(2*n-1)/n
    term1 = term1*rcf1
    term0 = term0*rcf0
    sum1 = sum1 + term1
    sum0 = sum0 + term0
  loop until abs(term1/sum1) < 1E-9 and term0/sum0 < 1E-9
end if
RatioI1_0 = sum1/sum0
end function
```

13.10 Determining output voltage from source off-load voltage

In later sections, we will show how practical diode voltmeter problems can be transformed so that they can be solved using techniques developed for the simple detector model. In this way, the diode detector is turned into an absolute AC voltage measuring instrument; at least in the sense that an accurate measurement of the DC output can be corrected to produce an accurate measurement of the peak amplitude of a sinusoidal input. In order to do that however, it is necessary to take account of the source loading effect (perhaps not so importantly when 50 Ω signal generators are connected to detectors driving CMOS op-amps; but certainly when feeble sources such as non-invasive sampling networks are used to drive moving-coil meters). Thus, in order to complete the minimum set of mathematical tools, we need two more procedures; one to get from source off-load voltage to DC output (for modelling purposes); and one to get from DC output to source off-load voltage.



The problem of calculating the output voltage from the source voltage can be understood by considering the circuit on the right. Here $\mathbf{V}_S$ is the RMS off-load source voltage, and V_{in} is the RMS voltage produced by the source when it it loaded by the detector. $R_S+\mathbf{j}X_S$ is the source output impedance, and R_{Zin} is the detector input impedance. Thus V_{in} is the output of a potential divider formed by the two impedances, i.e.:

$\mathbf{V}_{in} = \mathbf{V}_S\ R_{Zin} / (R_{Zin} + R_S + \mathbf{j}X_S)$

The detector however does not preserve phase information, and so we can eliminate complex quantities from the analysis without affecting its generality. Thus, taking the magnitude of the denominator, and adopting the convention that an RMS voltage not written in bold is a magnitude, we get:

$$V_{in} = \frac{V_S\, R_{Zin}}{\sqrt{(R_{Zin}+R_S)^2 + X_S^2}} \qquad (13.2)$$

We still cannot calculate V_{in} however, because R_{Zin} depends on it, and so it is necessary to use an iterative procedure. That involves making an initial guess of V_{in}, calculating the average current I_{av}, and hence the input impedance , and using the results to refine the original guess. The relevant formulae and functions are:

$I_{av} = V_m / R_D$

where, using the function described in **section 13.3**:

$V_m = V_p - V_f = V_p -$ **DVfp2m**$(R_D ; V_p ; mV_T ; I_s)$

$V_p = V_{in} \sqrt{2}$

and, using equation (**12.23**) and the Bessel function ratio routine from **section 13.9**:

$R_{Zin} = V_p^2 / 2\ P_{det} = V_p / 2\ (I_{av} + I_s)$ **RatioI1_0**(V_p/mV_T)

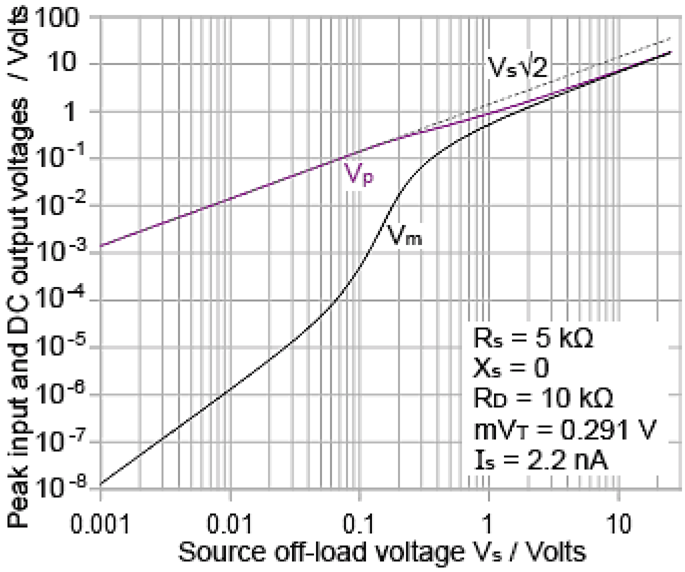
The approach was tested in the spreadsheet **det_models.ods**, sheet 6, and then coded into the Basic macro function shown below. The first method tried was that of taking the value of V_p from a round of calculation and using it directly as the starting value for the next round. This resulted in a convergent procedure provided that $|R_S+\mathbf{j}X_S|$ was less than R_D, but for very large magnitudes of source impedance and source voltages in the region of 0.1 V to 1 V (when $mV_T = 29.1$ mV), the routine would terminate at the maximum allowed number of iterations without giving the correct answer. Subsequent investigation showed that the internal value of V_p in this case would oscillate above and below the solution value, with the error neither growing nor shrinking. This behaviour was greatly curtailed by taking the seed for the next round of iteration to be the simple average of the previous and the current value. For some extremely improbable inputs however, such as a source impedance magnitude of $1000 \times R_D$, the routine could still fail. The complete solution was to use the 2:1 weighted average in favour of the previous value, thus eliminating any possibility of overshoot.

Note that the termination criterion for the program loop limits the maximum number of iterations to 128. Unmodified, it will not fail to find a solution for realistic input parameters, but the restriction gives insurance against 'program not responding' errors in the event that some awkward combination of inputs is found.

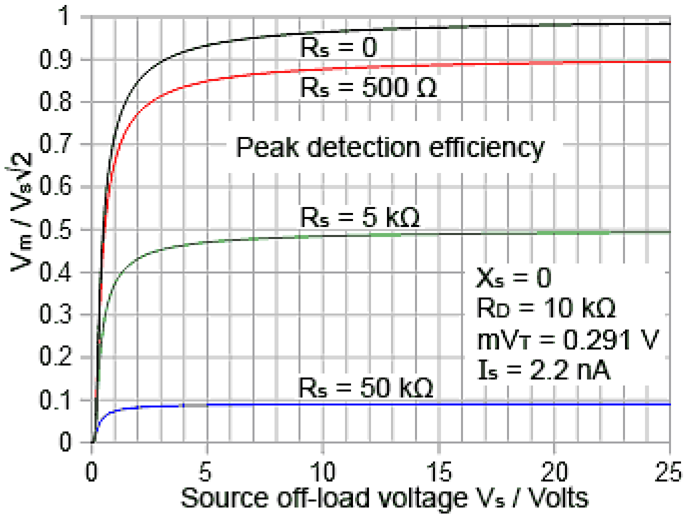
```
Function DetVs2m(byval Vs as double, Rs as double, Xs as double, RD as double, _
mVT as double, Isat as double) as double
'Calculates detector output voltage from source off-load voltage.  Version 1.00.
'Calls functions DVfp2m( ) and RatioI1_0( )
Dim Vpold as double, Vp as double, Vm as double, Iav as double, Rzin as double
Dim n as integer
DetVs2m = 0
If Vs = 0 then exit function
Vp = Vs*sqr(2)
do
  n=n+1
  Vpold = Vp
  Vm = Vp - DVfp2m(RD, Vp, mVT, Isat)
  Iav = Vm / RD
  Rzin= Vp / ( 2*(Iav+Isat)*RatioI1_0(Vp/mVT) )
  Vp = (2*Vpold + sqr(2)*Rzin*Vs/sqr((Rzin+Rs)^2 + Xs^2))/3
loop until abs(Vp-Vpold) < 1E-9 or n > 128
DetVs2m = Vm
end function
```

The results of some simulations using this algorithm are shown in the following graphs (see spreadsheet **det_models.ods**, sheet 6).

This graph shows the overall transfer function when the source impedance is a pure resistance equal to the limiting large-signal input impedance of the detector. Notice that the peak input voltage, V_p, does not begin to droop until the detector comes into its working range at an input of around 0.1 V RMS.



Here, the overall peak detection efficiency is given as the DC output voltage divided by the peak source voltage. The detector load resistance is 10 kΩ, and so the large-signal input resistance is just a little greater than 5 kΩ. The curves show the output for resistive source impedances of 500 Ω, 5 kΩ and 50 kΩ, with zero source impedance for comparison. Note that with a very large magnitude of source impedance , the linearity of the detector is improved, because the detector is then driven by something approaching an AC current source. The efficiency however is very low.



13.11 Determining source off-load voltage from output voltage

For the purpose of making absolute voltage measurements, the single most important routine is that which allows the off-load source voltage to be determined from the detector DC output. Given the work already done however, obtaining the source voltage is straightforward because it is quasi-analytical (i.e., it is analytical except for the fact that it calls on numerical routines).

Given V_m we can calculate V_p using the function described in **section 13.4**:

$V_p = V_m +$ **DVfm2p**(R_D ; V_m ; mV_T ; I_s)

also:

$I_{av} = V_m / R_D$

Using these quantities in equation (**12.23**) (with the function from **section 13.9**):

$R_{Zin} = V_p / 2\,(I_{av} + I_s)$ **RatioI1_0**(V_p / mV_T)

The on-load input voltage is the peak voltage divided by $\sqrt{2}$:

$V_{in} = V_p / \sqrt{2}$

and the source off-load voltage is given by a rearrangement of equation (**13.2**):

$$V_S = \frac{V_{in}\sqrt{(R_{Zin}+R_S)^2 + X_S^2}}{R_{Zin}}$$

A Basic macro function that performs the calculation is shown below:

```
Function DetVm2s(byval Vm as double, Rs as double, Xs as double, RD as double, _
mVT as double, Isat as double) as double
'Calculates source off-load voltage from detector output voltage.  Version 1.00,
'Calls functions DVfm2p( ) and RatioI1_0( )
Dim Vp as double, Vs as double, Iav as double, Rzin as double, Vin as double
Vp = Vm + DVfm2p(RD, Vm, mVT, Isat)
Iav = Vm / RD
Rzin = Vp /( 2*(Iav+Isat)*RatioI1_0(Vp/mVT) )
Vin = Vp / sqr(2)
DetVm2s = Vin*sqr((Rzin+Rs)^2 +Xs^2)/Rzin
end function
```

While the source off-load voltage itself is now easily obtained however, it is still necessary to calculate its standard deviation. We will return to that problem in a later section.

14. Generalised half-wave detector model

The simple diode detector model analysed in **section 12** is often assumed to provide a complete description of detector behaviour. Unfortunately however, while the various formulae and computational methods that develop around it can be used for the solution of practical problems, there is some art in the matter of applying them. Strictly, the simple model will give a good account of detectors driven by a source having extremely low DC resistance (such as an IF transformer), and using a diode with low forward ohmic resistance and low non-saturable reverse leakage (i.e., a high value of parallel resistance). It is therefore, as it stands, unsuitable for describing chokeless half-wave detectors (either series or shunt diode); and it will not account for the behaviour of realistic diodes with sufficient accuracy for absolute voltage measurement.

14.1 Series diode rectifier with port resistance and parasitics

A somewhat more general detector model is shown below. This uses the diode equivalent circuit given in **section 8**, and so offers an improvement in accuracy if we can determine the overall transfer function. Also included is port resistance (R_{port}), i.e., the DC resistance looking back into the source. Notice however that the source network does not correspond to an actual circuit, but to a generic representation of a network that has been transformed to separate the output impedance at the excitation frequency ($R_S + jX_S$) from the DC resistance (R_{port}). That transformation was introduced in **section 1.5**. The use of a perfect choke and a perfect coupling capacitor (∞ H and ∞ F), is a way of representing the mathematical separation using circuit symbols. A steady-state analysis is assumed, i.e., it is implied that infinite capacitances come pre-charged, and infinite inductances come pre-magnetised, according to the DC operating conditions. As before, a voltage with as subscript ending with ~ should be read as "the AC component of ".

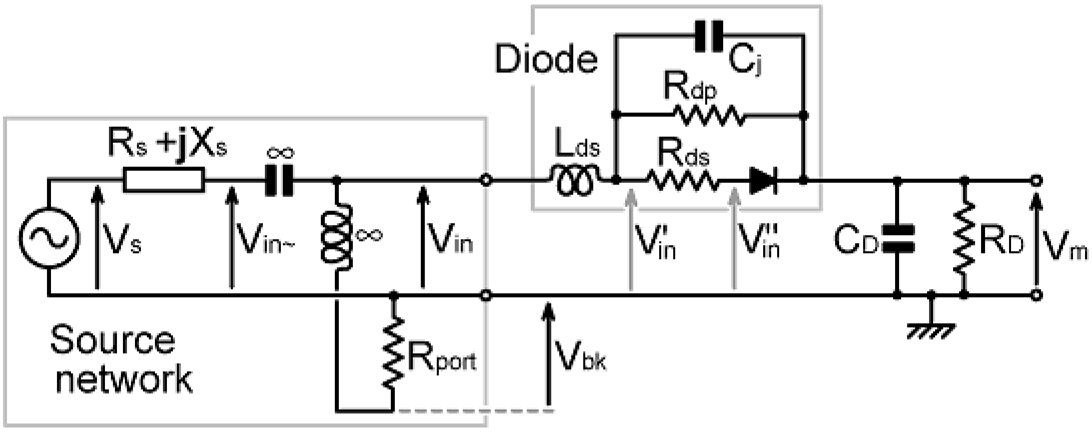


Fig. 14.1

This quasi-general diode voltmeter model is very different from the simple model of **section 12**; but if we keep sight of the practical objective (which is to determine the off-load source voltage V_S from the DC output voltage V_m), then it should be apparent that there are various rearrangements that will simplify the problem. We can start by considering the parasitic reactances.

L_{ds} is the series partial inductance of the diode and its wires or circuit tracks. Its value is about the same as that of a piece of wire of the same total length and average diameter as the diode leads from the source terminal to the smoothing capacitor. There will also be a contribution from the ground-return, but the use of a ground-plane will minimise this. In some circumstances, such as when working with surface-mount devices at VHF and below, or when using wire-ended diodes with short wires at HF, its effect will be swamped by the overall measurement uncertainty; in which case it can be neglected. The best thing to do with it is to lump it with the source impedance. In that case, when calculating the detector input voltage from the output voltage, we will get the quantity V_{in}' (or its AC component), which is not strictly observable, but we can still calculate $V_{in\sim}$ (if there is any point in doing so explicitly) and thence V_S, or we can calculate V_S directly using the modified source impedance.

C_j is the diode capacitance. This is a function of the junction depletion-layer thickness and so varies according to the instantaneous voltage (the varactor effect), but for the purpose of rectifier modelling, its value is taken to be the average at the excitation frequency. The fact that we are ignoring harmonic generation due to capacitance variation is a potential source of systematic error, but the effect is small for low capacitance signal diodes and will be lost in the overall measurement uncertainty. Thus having declared the intention of treating C_j as a simple capacitance; we can note that it cannot alter the average DC voltage across the smoothing capacitance C_D, it consumes no power, and since C_D is intentionally vastly greater in value then C_j, the only effect that C_j has is to place a weak reactive shunt across the AC input. Thus we might as well lump it with the source impedance.

The equivalent circuit that results from moving the parasitic reactances is shown below. It is now possible to combine them with the source impedance, but we will not do so yet because there is at least one more element to be transferred. For the purpose of the transformations to follow, the various voltages and currents associated with the diode network have also been marked.

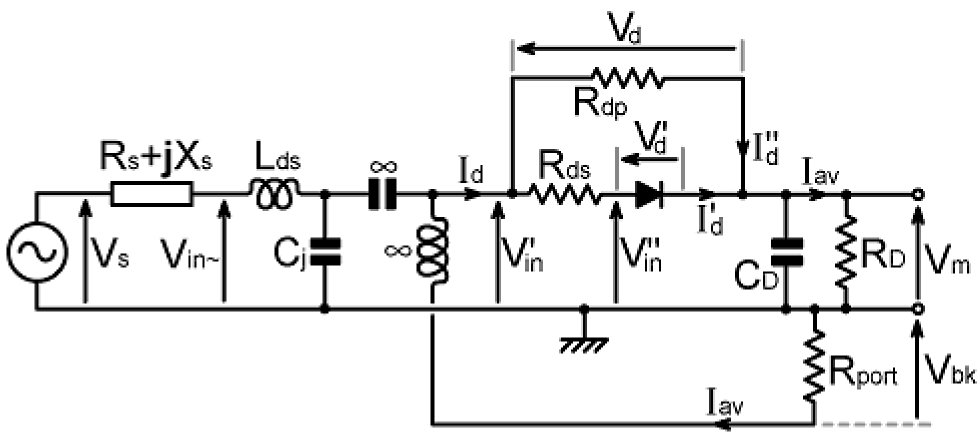


Fig 14.2

As was demonstrated in **section 12**, determining the detector transfer function depends on being able to calculate the average output current, I_{av}. That information is obtained by integrating the instantaneous diode current, I_d, over a complete cycle of the input waveform, but in this case the process is complicated by the presence of the resistances R_{ds} and R_{dp}. We can however, easily solve that part of the integral involving the non-saturable leakage current I''_d, and thereby obtain another transformation.

Previously, we adopted the convention that the AC component of the input waveform is sinusoidal and is zero and positive-going at $t = 0$, i.e.:

$$V_{in\sim} = V_p \sin\varphi$$

(some commentators prefer the cosine form, but it makes no difference to the outcome). Thus we can define the instantaneous input voltage V'_{in} as:

$$V'_{in} = V'_p \sin\varphi' - V_{bk}$$

where both the peak voltage and the phase angle of the AC component are primed because there will be a shift in both magnitude and phase due to the parasitic reactances. To the right of the point at which V'_{in} is marked on the diagram however, there are no more finite reactances, and so there will be no more phase shifts on the path to the load.

The instantaneous voltage across the whole diode (including its series resistance) is:

$V_d = V'_p \sin\varphi' - V_{bk} - V_m$

Hence the instantaneous diode current, using the diode equation (**6.1**), is:

$I_d = I'_d + I''_d = I_S [\exp(V'_d / mV_T) - 1] + (V'_p \sin\varphi' - V_{bk} - V_m) / R_{dp}$

where

$V'_d = V'_p \sin\varphi' - I'_d R_{ds} - I_{av} (R_D + R_{port})$

Note here that, because the expression for V'_d involves I'_d, there is no closed-form analytical expression for the total instantaneous current, let alone its integral (although a numerical solution is possible, as we will see later). Finding the average of the non-saturable leakage current is however straightforward.

The total average current is the area under the curve for the total instantaneous current.

$$I_{av} = \frac{V_m + V_{bk}}{R_D + R_{port}} = \frac{1}{2\pi}\int_0^{2\pi} I_d \, d\phi = \frac{1}{2\pi}\int_0^{2\pi} I'_d \, d\phi + \frac{1}{2\pi}\int_0^{2\pi} I''_d \, d\phi$$

Thus, if we separate that part of the average current that flows through $\mathbf{R_{dp}}$ and call it $\mathbf{-I''_{av}}$ (the reason for the minus sign will become apparent shortly), we have:

$$-I''_{av} = \frac{1}{2\pi}\int_0^{2\pi} \frac{V'_p \sin\phi' - V_{bk} - V_m}{R_{dp}} \, d\phi$$

Now observe that the sinusoidal term averages to zero over a complete cycle. Thus the constant terms are all that remain, and after evaluating the integral we get:

$-I''_{av} = -(V_{bk} + V_m) / R_{dp}$

What this means is that the diode parallel resistance acts as a DC shunt across the total load resistance $R_D + R_{port}$. The minus sign in the expression above tells us that the current flowing through R_{dp} is subtracted from the current available for charging the smoothing capacitor, i.e., it reduces the detector output. This result, actually obvious on re-inspection of **Fig. 14.2**, allows us to redraw the model once again.

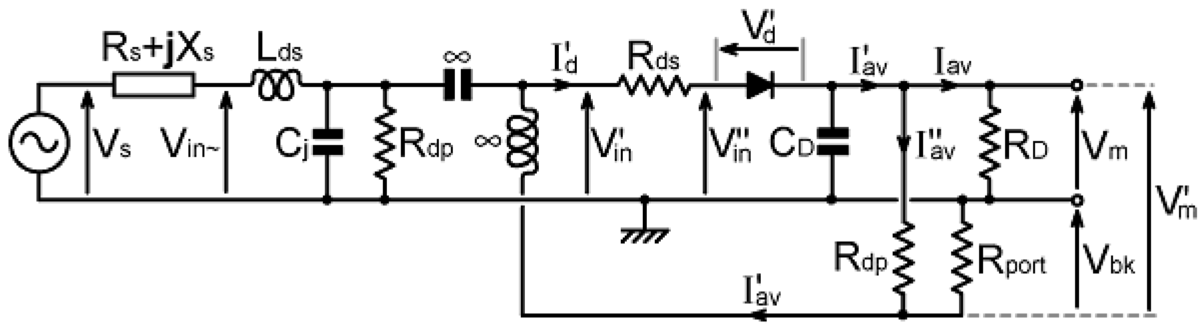


Fig. 14.3

The sign chosen above for I''_{av} allows the diagram to be labelled sensibly. Also notice that there are now two instances of R_{dp} in the equivalent circuit. The reason for that is that R_{dp}, as well as shunting the rectified output, also shunts the input signal; and since we have imposed a mathematical separation between the AC and DC parts of the problem, it is necessary to account for both effects separately.

Having worked out how to deal with R_{dp} however, it should be noted that in many instances it can be ignored. This is because modern signal diodes typically have a large R_{dp}; about 250 MΩ for a 1N5711 Schottky diode, and somewhere from a few hundred kΩ to a few MΩ for a germanium detector diode. Thus, given that most RF networks will have output impedances of less than a few hundred ohms magnitude; the AC shunting effect will usually be negligible. The DC shunting effect should however be taken into account when using a diode in conjunction with a large value of R_D or R_{port}, a sensible guideline being that it should be included if it reduces the output by more than about 0.1%.

We have now annexed most of the diode equivalent-circuit elements either into an effective source impedance, or into an effective load resistance, or both. A new, greatly simplified model results:

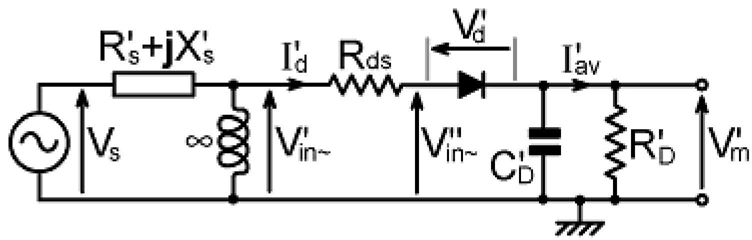


Fig. 14.4

Here the voltages $V'_{in\sim}$ and $V''_{in\sim}$ are the AC components of the voltages V'_{in} and V''_{in}. Also notice that the smoothing capacitance has been given a prime, but this is merely to warn that the transformed load resistance R'_D is not to be used for calculating the detector time constant.

For this model, V'_m is the total detected voltage, i.e.:

$V'_m = V_m + V_{bk}$

and

$R'_D = (R_D + R_{port}) \,//\, R_{dp}$

Thus V_m can be calculated as the output of a potential divider composed of R_D and R_{port}, i.e.:

$V_m = V'_m R_D / (R_D + R_{port})$

and V'_m can be calculated from V_m :

$V'_m = V_m (R_D + R_{port}) / R_D$

i.e.:

$V'_m = V_m (1 + R_{port} / R_D)$ . . . . . . (**14.1**)

The average rectified current is:

$I'_{av} = V'_m / R'_D$

and so on.

The effective source impedance is obtained using Thévenin's theorem, i.e., by writing down the impedance looking back into the source when the generator is replaced by a short circuit. Thus, by inspection of **Fig. 14.3**:

$R'_S + jX'_S = (R_S + jX_S + jX_{Lds}) // R_{dp} // jX_{Cj}$

We can use the aggregated parameter approach to solve the general diode voltmeter problem, provided that we can determine the overall transfer function. The final issue therefore, is what to do about the diode series resistance R_{ds} .

There are two ways in which we can proceed at this point. One approach is to move R_{ds} into both the source impedance and the port resistance; in which case the analysis reverts to that of the simple detector model and its input impedance as per the working in **sections 12 and 13**. The other approach is to develop numerical solutions for the transfer function of a detector with diode series resistance. Both methods should agree but, since we have no mathematically rigorous basis for the presumed equivalence, it will be instructive to try them both. In this section, we will continue reducing the problem to that of the simple detector. The detector with diode series resistance will be analysed in **section 15**.

If we go back to the circuit of **Fig. 14.3** and transfer instances of R_{ds} into both the source impedance and the load resistance we get the equivalent circuit shown below:

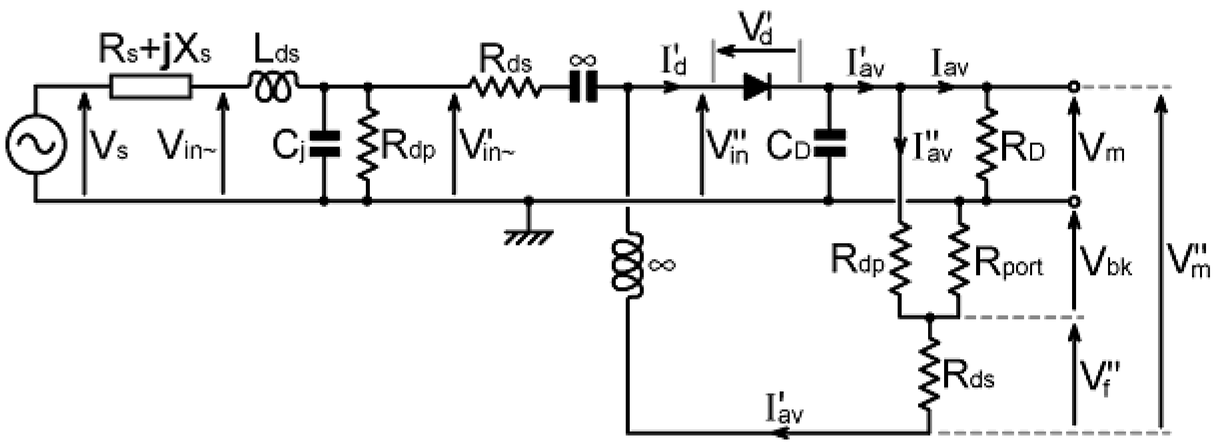


Fig. 14.5

This network can now be modelled as a simple diode detector, with new circuit parameters suitable for passing to the calculation routines described in **sections 13.10** and **13.11**. All we have to do is pass the modified parameters from the circuit below in place of the original quantities.

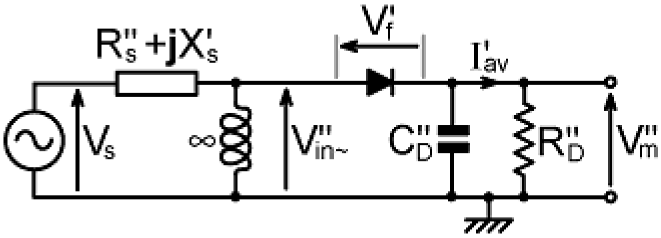


Fig. 14.6

Notice that adding R_{ds} to the effective source impedance does not change its reactance. Consequently, although the effective source resistance gets an extra prime (cf. **fig. 14.4**) the reactive element does not.

To calculate V''_m, the appropriate function call is:

V''_m = **DetVs2m**(V_S ; R''_S ; X'_S ; R''_D ; mV_T ; I_S)

and to calculate V_S, the call is:

V_S = **DetVm2s**(V''_m ; R''_S ; X'_S ; R''_D ; mV_T ; I_S)

The required parameters can be calculated using the formulae and programs given below.

Effective load resistance

The effective load resistance is given by:

$R''_D = R'_D + R_{ds}$

Where R_D' is the effective load resistance from fig. 14.4. Thus:

$R''_D = R_{ds} + R_{dp} // (R_D + R_{port})$

i.e.:

$R''_D = R_{ds} + [\ R_{dp}\ (R_D + R_{port}) / (\ R_{dp} + R_D + R_{port})\ ]$	**14.2**

A Basic macro routine that calculates R'_D is shown below:

```
Function RDeff1(RD as double, Rport as double, Rdp as double) as double
'Calculates effective load resistance RD' = Rdp // (RD + Rport)
RDeff1 = Rdp*(RD+Rport) / (Rdp+RD+Rport)
end function
```

R''_D is obtained by calling this function and adding R_{ds}:

$R''_D = R_{ds} +$ **RDeff1**(R_D ; R_{port} ; R_{dp})

Effective source impedance
To calculate R''_S and X'_S, the transformation shown in the diagram below is required:

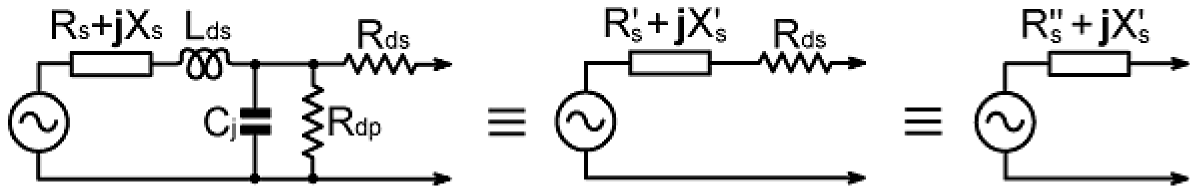


Fig. 14.7

Applying Thévenin's theorem gives us:

$R''_S + jX'_S = R_{ds} + R'_S + jX'_S = R_{ds} + (R_S + jX_S + jX_{Lds}) \,//\, R_{dp} \,//\, jX_{Cj}$ . . . . (14.7)

Extracting expressions for R_S' and X_S' from this is related to a general problem that we will refer to as the 'series-parallel to series (SP2S) transformation'. Since we will also require exactly the same derivation again later, it will be set out using a more general notation in the next subsection. The resulting formulae are coded into Basic macros, so that R_S'' and X_S' can be calculated using the following function calls:

$R''_S = R_{ds}$ + **SP2S_R**(R_S ; $X_S + X_{Lds}$; R_{dp} ; X_{Cj})

and

X'_S = **SP2S_X**(R_S ; $X_S + X_{Lds}$; R_{dp} ; X_{Cj})

The actual expressions for R''_S (or R'_S) and X'_S are rather complicated and, in principle at least, we will need to differentiate both of them with respect to each of their parameters in order to calculate the overall standard deviation of a measured voltage. This is not inherently difficult, but it is algebraically messy. It is therefore worth noting that C_j and R_{dp} are usually of considerably less statistical importance than the other parameters. If $|X_{cj}|$ and R_d are relatively large, it is still useful to include them for the purpose of reducing systematic error, but they won't make much difference to the estimated standard deviation (ESD) of either R''_S or X'_S. Therefore we can usually neglect them in that context. Hence for the purpose of calculating ESD, we merely have to differentiate the expressions:

$R''_S \approx R_S + R_{ds}$ and $X'_S \approx X_S + X_{Lds}$

Calculating DC output from total detected voltage

The total detected voltage V''_m is determined from the measured output voltage V_m by analysing the potential divider formed by R_{ds}, R_{dp}, R_{port} and R_d:

$$V_m + V_{bk} = V''_m \frac{R_{dp} // (R_D + R_{port})}{R_{ds} + R_{dp} // (R_D + R_{port})} \qquad (14.3)$$

and

$$V_m = (V_m + V_{bk}) \frac{R_D}{R_D + R_{port}} \qquad (14.4)$$

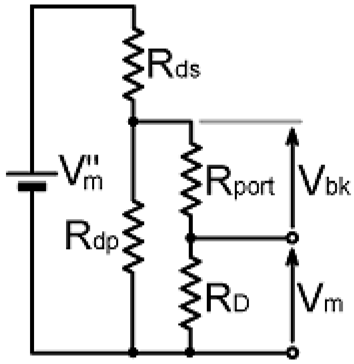


Fig. 14.8

Substituting (14.3) into (14.4) gives:

$$V_m = V''_m \frac{[R_{dp} // (R_D + R_{port})]}{[R_{ds} + R_{dp} // (R_D + R_{port})]} \frac{R_D}{(R_D + R_{port})}$$

Now if we expand the parallel resistances we get:

$$V_m = V''_m \frac{\dfrac{R_{dp} (R_D + R_{port})}{(R_{dp} + R_D + R_{port})} \dfrac{R_D}{(R_D + R_{port})}}{R_{ds} + \dfrac{R_{dp} (R_D + R_{port})}{(R_{dp} + R_D + R_{port})}}$$

Rearranging the main denominator so that all of its terms share a common denominator gives:

$$V_m = V''_m \frac{\dfrac{R_D R_{dp}}{(R_{dp} + R_D + R_{port})}}{\dfrac{R_{ds}(R_{dp} + R_D + R_{port}) + R_{dp} (R_D + R_{port})}{(R_{dp} + R_D + R_{port})}}$$

Thus, with some rearrangement of the terms in the denominator:

$V_m = V''_m \dfrac{R_D R_{dp}}{R_{ds}(R_D + R_{port}) + R_{dp} (R_D + R_{port} + R_{ds})}$	14.5

Notice here that, as $R_{dp} \to \infty$, only the terms in the denominator having R_{dp} as a factor survive, and so we get:

as $R_{dp} \to \infty$, $V_m \to V''_m R_D / (R_D + R_{port} + R_{ds})$

If we decide to ignore the diode series resistance, or set it to zero so that we can use a numerical method for the average diode current, we get:

as $R_{ds} \to 0$, $V_m \to V''_m R_D / (R_D + R_{port})$

A basic macro routine that calculates V_m from V''_m using equation (**14.5**) is given below:

```
Function DCVout( Vm2 as double, RD as double, Rport as double, Rdp as double, _
Rds as double) as double
DCVout = Vm2*RD*Rdp/(Rds*(Rd+Rport) + Rdp*(RD+Rport+Rds))
end function
```

The function call is:

V_m = **DCVout**(V''_m ; R_D ; R_{port} ; R_{dp} ; R_{ds})

or, if starting from V_m' (i.e., using a numerical diode-with-series-resistance model):

V_m = **DCVout**(V''_m ; R_D ; R_{port} ; 1 ; 0)

Notice, by inspection of equation (**14.5**), that R_{dp} makes no difference to the calculation in this case, but it must be > 0 . Inserting a value of 1 is therefore convenient.

Calculating total detected voltage from DC output

To obtain the total detected voltage V_m'' from the output voltage V_m, we can rearrange equation (**14.5**) thus:

$$V''_m = V_m \frac{R_{ds}(R_D + R_{port}) + R_{dp}(R_D + R_{port} + R_{ds})}{R_D R_{dp}}$$

This can be simplified in various ways, e.g.:

$$V''_m = V_m \left[1 + \frac{R_{port}}{R_D} + \frac{R_{ds}(R_D + R_{port} + R_{dp})}{R_D R_{dp}} \right] \qquad \textbf{14.6}$$

Notice that when $R_{ds} \rightarrow 0$, this expression reverts to equation (**14.1**).

A Basic macro that performs the calculation is given below:

```
Function DCVtot( Vm as double, RD as double, Rport as double, Rdp as double, _
Rds as double) as double
DCVtot = Vm*(1 + (Rport/RD) + Rds*(RD+Rport+Rdp)/(RD*Rdp) )
end function
```

The function call for evaluating V''_m is:

V''_m = **DCVtot**(V_m ; R_D ; R_{port} ; R_{dp} ; R_{ds})

In order to calculate V'_m (such as when using a diode-with-series-resistance model) the call is:

V'_m = **DCVtot**(V_m ; R_D ; R_{port} ; 1E9 ; 0)

R_{dp} is not required in this case, but a value of 1E12 prevents a divide-by-zero error.

>>>>> Macro needs updating:
R_{dp} is not required if $R_{ds} = 0$ is trapped automatically:

```
If Rds = 0 then
  DCVtot = Vm*(1 + (Rport/RD) )
else
  If Rdp = 0 then Rdp = 1E12
  DCVtot = Vm*(1 + (Rport/RD) + Rds*(RD+Rport+Rdp)/(RD*Rdp) )
end if
end function
```

If R_D is zero, result is undefined, so best to allow the div 0 error to occur as a warning.

>>>>>>

14.2 Series-parallel to series transformation

The diagram below refers to a frequently encountered electrical problem. A generator having finite source impedance has an impedance placed across its terminals, and we want to find the new effective source impedance.

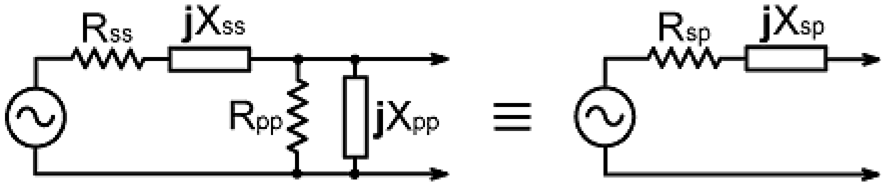


Fig. 14.9

Using Thévenin's theorem:

$$R_{SP} + \mathbf{j}X_{SP} = (R_{SS} + \mathbf{j}X_{SS}) \,//\, R_{PP} \,//\, \mathbf{j}X_{PP}$$

The parallel impedance (//) operator is commutative, so we can start the expansion by combining the series elements with the parallel resistance.

$$R_{SP}+\mathbf{j}\,X_{SP} = \frac{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}}{R_{SS}+R_{PP}+\mathbf{j}\,X_{SS}} \,//\, \mathbf{j}\,X_{PP}$$

Now including the parallel reactance we get:

$$R_{SP}+\mathbf{j}\,X_{SP} = \frac{\dfrac{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}\;\mathbf{j}\,X_{PP}}{R_{SS}+R_{PP}+\mathbf{j}\,X_{SS}}}{\dfrac{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}}{R_{SS}+R_{PP}+\mathbf{j}\,X_{SS}}+\mathbf{j}\,X_{PP}}$$

Placing the fraction in the denominator on to a common denominator gives:

$$R_{SP}+\mathbf{j}\,X_{SP} = \frac{\dfrac{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}\;\mathbf{j}\,X_{PP}}{R_{SS}+R_{PP}+\mathbf{j}\,X_{SS}}}{\dfrac{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}+\mathbf{j}\,X_{PP}(R_{SS}+R_{PP}+\mathbf{j}\,X_{SS})}{R_{SS}+R_{PP}+\mathbf{j}\,X_{SS}}}$$

i.e., after cancellation:

$$R_{SP}+\mathbf{j}\,X_{SP} = \frac{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}\;\mathbf{j}\,X_{PP}}{(R_{SS}+\mathbf{j}\,X_{SS})\,R_{PP}+\mathbf{j}\,X_{PP}(R_{SS}+R_{PP}+\mathbf{j}\,X_{SS})}$$

Multiplying out the brackets and grouping reals and imaginaries produces:

$$R_{SP}+\mathbf{j}\,X_{SP} = \frac{-R_{PP}\,X_{SS}X_{PP}+\mathbf{j}\,R_{SS}\,R_{PP}\,X_{PP}}{(R_{SS}\,R_{PP}-X_{SS}\,X_{PP})+\mathbf{j}(R_{PP}\,X_{SS}+R_{SS}\,X_{PP}+R_{PP}\,X_{PP})}$$

and multiplying numerator and denominator by the complex conjugate of the denominator:

$$R_{SP}+\mathbf{j}X_{SP}=\frac{(-R_{PP}X_{SS}X_{PP}+\mathbf{j}R_{SS}R_{PP}X_{PP})[(R_{SS}R_{PP}-X_{SS}X_{PP})-\mathbf{j}(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})]}{(R_{SS}R_{PP}-X_{SS}X_{PP})^2+(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})^2}$$

Now we can multiply-out the numerator and separate the real and imaginary parts.

$$R_{SP}+\mathbf{j}X_{SP}=\frac{-R_{PP}X_{SS}X_{PP}(R_{SS}R_{PP}-X_{SS}X_{PP})+R_{SS}R_{PP}X_{PP}(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})}{(R_{SS}R_{PP}-X_{SS}X_{PP})^2+(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})^2}$$

$$+\frac{\mathbf{j}[R_{SS}R_{PP}X_{PP}(R_{SS}R_{PP}-X_{SS}X_{PP})+R_{PP}X_{SS}X_{PP}(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})]}{(R_{SS}R_{PP}-X_{SS}X_{PP})^2+(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})^2}$$

Thus, with a reversal of the terms in the real part to get rid of the leading minus sign:

$$R_{SP}=\frac{R_{SS}R_{PP}X_{PP}(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})-R_{PP}X_{SS}X_{PP}(R_{SS}R_{PP}-X_{SS}X_{PP})}{(R_{SS}R_{PP}-X_{SS}X_{PP})^2+(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})^2} \qquad \textbf{14.8}$$

and

$$X_{SP}=\frac{R_{SS}R_{PP}X_{PP}(R_{SS}R_{PP}-X_{SS}X_{PP})+R_{PP}X_{SS}X_{PP}(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})}{(R_{SS}R_{PP}-X_{SS}X_{PP})^2+(R_{PP}X_{SS}+R_{SS}X_{PP}+R_{PP}X_{PP})^2} \qquad \textbf{14.9}$$

The OO Basic macro functions shown below perform the calculations.

```
Function SP2S_R(Rss as double, Xss as double, Rpp as double , Xpp as double ) as double
'series-parallel to series transformation.  Calculates Rsp
Dim t1 as double, t2 as double
t1 = Rpp*Xss + Rss*Xpp + Rpp*Xpp
t2 = Rss*Rpp - Xss*Xpp
SP2S_R = (Rss*Rpp*Xpp*t1 - Rpp*Xss*Xpp*t2) / (t1*t1 + t2*t2)
end function
```

```
Function SP2S_X(Rss as double, Xss as double, Rpp as double , Xpp as double ) as double
'series-parallel to series transformation.  Calculates Xsp
Dim t1 as double, t2 as double
t1 = Rpp*Xss + Rss*Xpp + Rpp*Xpp
t2 = Rss*Rpp - Xss*Xpp
SP2S_X = (Rss*Rpp*Xpp*t2 + Rpp*Xss*Xpp*t1) / (t1*t1 + t2*t2)
end function
```

14.3 Shunt diode rectifier with port impedance and parasitics

The problem of transforming the shunt-diode rectifier circuit so that it can be analysed using the simple detector model has much in common that of the series-diode case. There are however, some differences that must be taken into account.

A fairly realistic equivalent circuit for the shunt-diode detector is shown below. As is the convention in this document, a voltage placed in square brackets refers to the DC component of a waveform. The diode model used is as discussed in **section 8**.

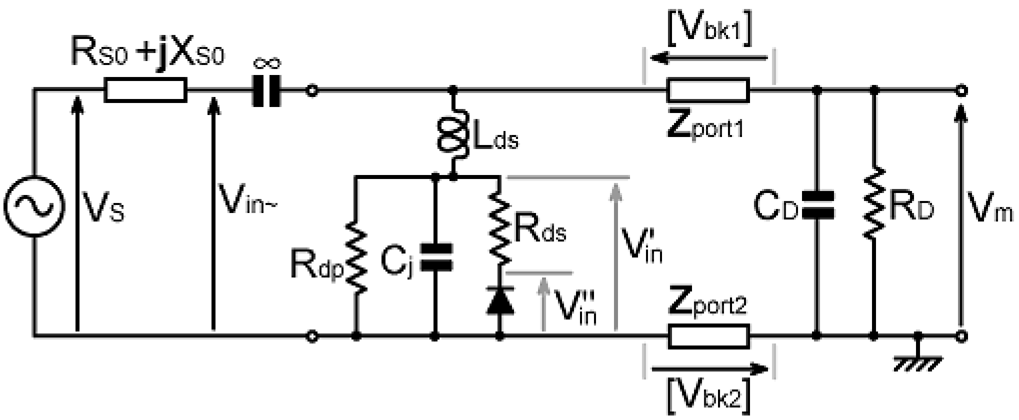


Fig 14.10

Notice that the source impedance has been written with an additional subscript 0 on each of its elements. This is done pending a transformation to make the complete network analytically equivalent to the series diode half-wave detector. Additionally, a perfect coupling capacitor has been included to make it explicit that the source is a DC open-circuit. This might seem unnecessary, given that there will be an actual coupling capacitor, but defining the series reactive element of the source impedance as negative at the input frequency is not the same as stating that there is no DC path.

The major analytical difference between the conventional detector and the shunt detector is that the latter has one or more port *impedances*, instead of just a port resistance. This is because a port resistance is defined (at least in this document) as the DC resistance in the path to the load; and in the case of the shunt detector, that path is physically separate from the source impedance. Thus we have to take both the total port resistance and the total port impedance into account.

The reason why there are two port impedances is that the shunt detector can be used for the measurement of floating voltages. Thus, assuming the grounding arrangement shown in the diagram, Z_{port2} is the ground-return impedance, whereas Z_{port1} refers to the network connecting the diode to the live output terminal. When the source AC reference is not at ground potential, it can still usually be arranged that $|Z_{port2}|$ is very small relative to $|Z_{port1}|$. In that case most analyses can proceed with $Z_{port2} = 0$, but it is important to be aware that its neglect might introduce a systematic error.

The total port impedance for analytical purposes is:

$$Z_{port} = Z_{port1} + Z_{port2}$$

If the maximum possible detector sensitivity is required, Z_{port1} can be the impedance of an RF choke; although such a choice is not recommended for precision voltmeter applications. Even if the DC path is provided by a resistor however, it is still an impedance, not a pure resistance, because high-value resistors are usually predominantly capacitive at radio frequencies when operated below the SRF.

The next step in the analysis is to separate the port resistance from the total port impedance. This is illustrated in the diagram below.

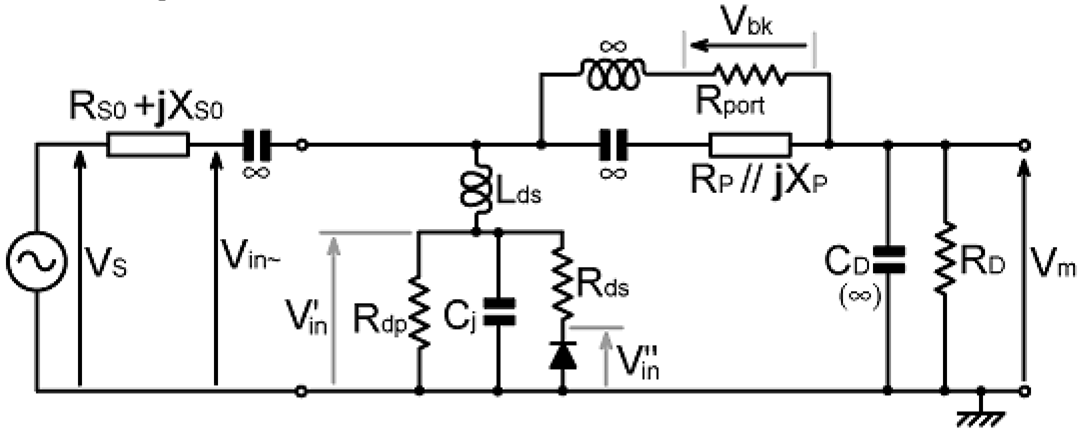


Fig 14.11

In this representation, a perfect choke channels the DC output through the port resistance, and the port impedance is made into a DC open circuit by means of a perfect coupling capacitor. Note that the port resistance is not the same as the resistive element of the port impedance. For that reason the impedance has been written as $R_P // jX_P$. The parallel form is used because the transformation has made it obvious that the port impedance is effectively in parallel with the detector input. This is so because the smoothing capacitor intentionally has a reactance that is very much smaller in magnitude than $|R_P // jX_P|$, and the AC component of the current through the smoothing capacitor makes no contribution to the DC output. Hence the next step in the transformation, as shown below, is to place the port impedance directly in parallel with the source.

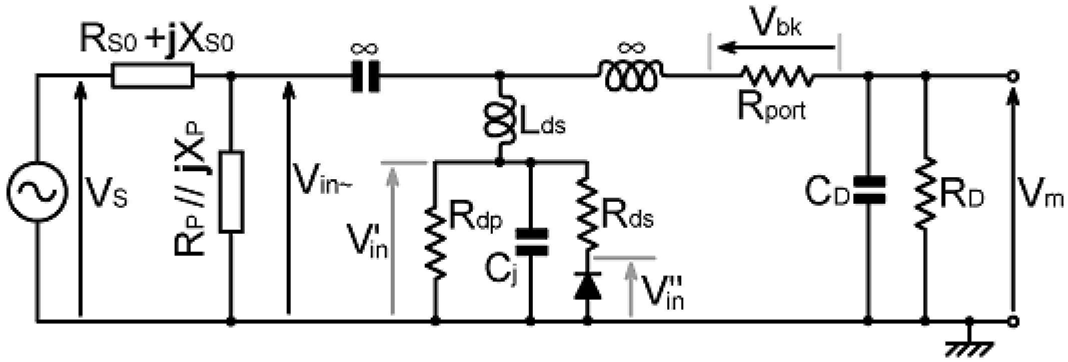


Fig 14.12

At this point it will become apparent why the source impedance has been given symbols that are different from the conventional detector case. The reason is that, while the port impedance is, notionally, already combined with the source impedance in the series diode network, in this case we need to combine the two impedances explicitly. Thus, for the *effective* source impedance (i.e., including the shunting effect of the port impedance), we can use the same definition as in the conventional case (and hence the same computer programs for enumeration) provided that we make the following transformation:

$R_S + jX_S = (R_{S0} + jX_{S0}) // R_P // jX_P$

Here, the use of the parallel form for the port impedance works to our advantage because, in the preferred case when the port impedance is that of a high-value resistor; the resistor can usually be represented as its actual resistance in parallel with a capacitance of about 0.4 pF (provided that it is operating below its SRF). Furthermore, if the source impedance magnitude is relatively low, the effect of the port resistance on the AC amplitude can be neglected, and the port impedance becomes simply the reactance of the resistor's capacitance. Note incidentally, that the port capacitance as a modelling parameter will be strongly correlated with the junction capacitance C_j. Thus, if the model is used for parameter extraction, the two capacitances will probably have to be accounted for by eliminating one and allowing the other to increase slightly.

The matter of combining the source impedance and the port impedance requires the series-parallel to series (SP2S) transformation described in **section 14.2**. Using the Basic macro routines given in that section, the required function calls are:

R_S = **SP2S_R**(R_{S0} ; X_{S0} ; R_P ; X_P) and X_S = **SP2S_X**(R_{S0} ; X_{S0} ; R_P ; X_P)

The diode parasitic reactances and parallel resistance can now be combined with the source impedance, and another instance of the parallel resistance can be combined with the DC load. This, as is shown below, gives us a configuration that can be analysed using the diode-with-series-resistance numerical method.

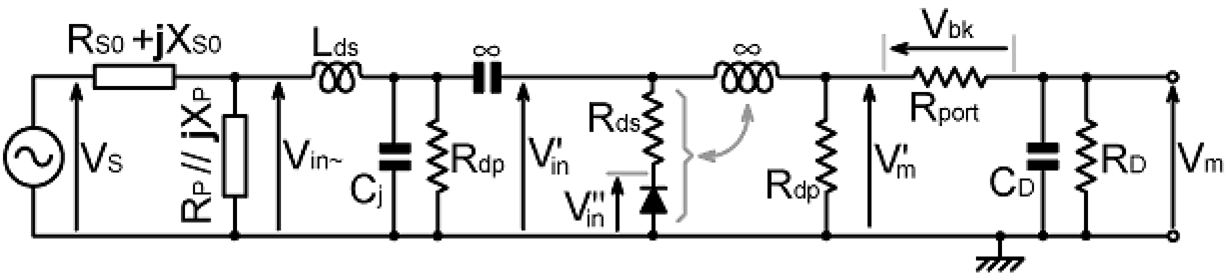


Fig 14.13

This network should be compared with **figure 14.3**, whereupon it will be noted that; if the source and port impedances are combined, and the diode with its series resistance and the ideal choke are transposed, the only difference lies in the ordering of the resistances in the DC potential divider. Since addition of resistances is commutative, this problem is analytically identical to the comparable series diode case.

Finally, instances of the diode series resistance can be combined with the source impedance and the load.

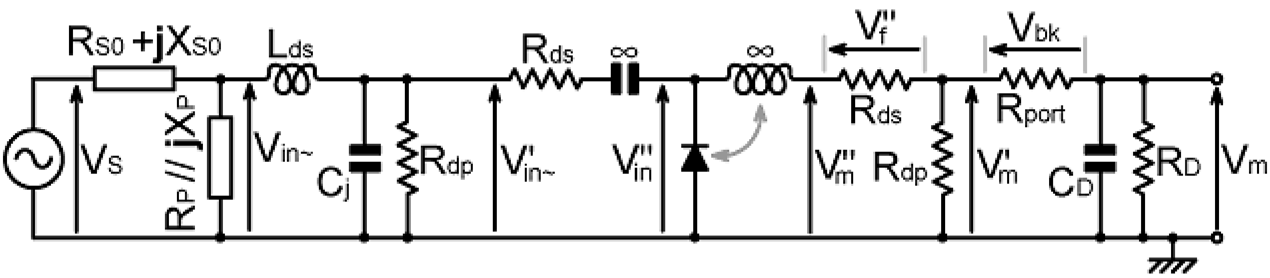


Fig 14.14

Once the port impedance is combined with the source impedance, this network is analytically identical to the one shown in **figure 14.5**. It can therefore be analysed as a simple detector using the transformations given in the previous section.

15. Detector with diode series resistance

In **section 12**, we developed the transfer function of the simple half-wave detector; the word 'simple' in that context referring to the omission of a large number of elements affecting the behaviour of practical circuits. In **section 14** however, we showed that realistic problems can be handled using the simple detector functions; the trick being to modify the source impedance and the DC load resistance to account for the various changes. There is however, one point of theoretical contention that requires further exploration.

The diode model has two resistances; the parallel resistance R_{dp}, which accounts for the non-saturable leakage current, and the junction resistance R_{ds}. There must also be a physical resistance, referred-to here as the 'port resistance', in the DC loop that completes the detector circuit. All of these are exposed to the harmonic currents created by the rectification process, and they will therefore dissipate some proportion of the harmonic energy.

If a diode voltmeter is to be calibrated using only DC standards, then it is necessary to be able to deduce the off-load source voltage from the DC output voltage. This requires that we have an accurate knowledge of the detector input impedance, which is obtained by considering the power delivered to the detector (see **section 12.3**). The detector power problem was solved however using the simple detector model, which means that the generator used had zero output impedance and is a perfect short-circuit with regard to reflected power. The derivation also made use of the assumption that the smoothing capacitor has a very low magnitude of reactance at the excitation frequency, and therefore also at higher frequencies, preventing any RF dissipation in the DC load resistor. Thus the model is such that all of the harmonic energy is dissipated in the diode itself. The effect of assigning that energy to the diode, instead of distributing it to the various resistances, will be to cause a systematic error in the determination of the input impedance. It is reasonable to expect that this discrepancy will be small in most instances; but without some way of quantifying it, it is not possible to be sure.

To allow the dissipation of harmonic energy in the resistances associated with the diode, it is necessary to include those resistances in the process of integrating the instantaneous current to obtain the average current. Then to find the detector input impedance, it is necessary to include those resistances in the integration of the instantaneous VI product to find the power. Before doing that however, a certain pragmatism is indicated in the matter of what to include.

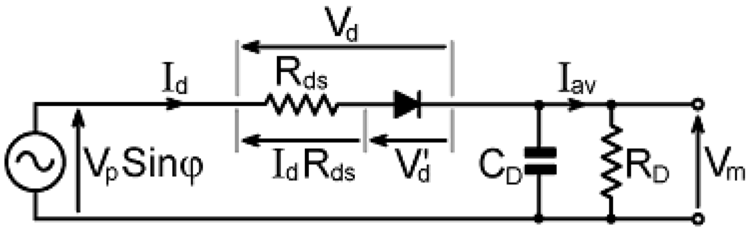
The first issue is that, for practically any diode that we might want to use, R_{dp} is large; and as was discussed in **section 14.1**, there is some point in including it in the effective DC load, but there is usually little point in including it in the effective source impedance. Moving it into the source impedance is a way of accounting for its dissipation at the excitation frequency, and if there is some doubt about the need to do that, then there is even more doubt about the need to consider its absorption of energy from higher harmonics. Thus, while it will certainly complicate matters to include R_{dp} in the power integration, it is unlikely to produce any effect of statistical significance. For that reason, will not bother to do it.

The second issue is a matter of complexity. The resistance looking back into the source will be different at different frequencies. We have already separated its DC component from its component at the excitation frequency, but to make a full account of all of the harmonic energy, we will need a component at each if the harmonic frequencies. This is workable if all of those components are the same, but they will only be so if the source impedance is a pure resistance. Moreover, if the source can be modelled as a generator with a pure resistance in series with it, then that resistance can be combined with the diode series resistance.

It follows, that while we will not be able to take account of harmonic energy to a high degree of accuracy in all situations, we will certainly be able to discover the magnitude of its effect by including R_{ds} in the integrations.

15.1 Instantaneous diode current

In the diagram below, the simple detector model of **section 12** has been modified by the inclusion of diode series resistance. In order to find the average diode current I_{av}, from which the DC output voltage V_m is calculated, it is necessary to integrate the instantaneous current I_d over a complete cycle of the input waveform. The first task therefore is to determine I_d. This is given by the near ideal diode equation (see **section 6**):



$I_d = I_S [\exp\{ V'_d / mV_T \} - 1]$

Where:

$V'_d = V_d - I_d R_{ds}$

i.e.:

$I_d = I_S [\exp\{ (V_d - I_d R_{ds}) / mV_T \} - 1]$

The problem with this expression is that it has instances of I_d both inside and outside the exponential bracket. This means that there will be no closed-form analytical solution. We will therefore use an iterative method, based on the procedure first introduced in **section 13.3**. We start by giving different symbols to the two instances of I_d:

$y = I_S [\exp\{ (V_d - x R_{ds}) / mV_T \} - 1]$

An initial value for x is easily obtained by calculating I_d on the assumption that $R_{ds}= 0$, i.e.;

$x_0 = I_S [\exp(V_d / mV_T) - 1]$

x then has to be adjusted until y agrees with it to some sensible degree of accuracy, i.e:

$|x - y| \leq$ maximum acceptable error.

The change in y due to a small change in x is defined thus:

When $x \rightarrow x + \delta x$, $y \rightarrow y + \delta y$

and a solution occurs when:

$x + \delta x = y + \delta y$

where, for small increments:

$\delta y \approx \delta x\ \partial y/\partial x$

Thus:

$x + \delta x \approx y + \delta x\ \partial y/\partial x$

i.e.:

$\delta x \approx (x - y) / (\partial y/\partial x - 1)$

The derivative is obtained using the chain rule:

$\partial y/\partial x = I_S (-R_{ds} / mV_T) \exp\{ (V_d - x R_{ds}) / mV_T \}$

Notice that if $R_{ds} = 0$, the derivative is always zero and the initial estimate for x is the solution.

A Basic program that performs the iteration is shown below. Algorithm development was carried out in the spreadsheet **det_models.ods**, sheet 2. The convergence criterion |x-y| has been set so that I_d is calculated to the nearest pico amp.

```
Function DinstI(byval Vd as double, Rds as double, mVT as double, Isat as double) as double
'Calculates diode instantaneous current when diode has finite series resistance
Dim x as double
x = Isat*(exp(Vd/mVT)-1)
DinstI = x
If Rds <= 0 then exit function
Dim arg as double, y as double, deriv as double, diff as double, deltax as double
do
  arg = (Vd - x*Rds)/mVT
  y = Isat*(exp(arg)-1)
  deriv = -Isat*Rds*exp(arg)/mVT
  diff = x-y
  deltax = diff/(deriv-1)
  x = x+deltax
Loop until abs(diff) < 1E-12
DinstI = x
end function
```

15.2 Average diode current by numerical integration

The DC output of a detector is given by the product of the average diode current and the effective load resistance; i.e., assuming zero port resistance and zero non-saturable leakage current:

$V_m = I_{av} R_D$

The average current is the area under the curve for the instantaneous current. This can be obtained by integrating the instantaneous current (I_d) over one complete input cycle and dividing by the length of a cycle, i.e.;

$$I_{av} = \frac{V_m}{R_D} = \frac{1}{2\pi}\int_0^{2\pi} I_d \, d\phi$$

This was carried out in **section 12.1** for the case with $R_{ds} = 0$ and led to a solution involving the zero-order modified Bessel function of the first kind. We know immediately however, from the working in the previous section, that when $R_{ds} > 0$, this integral is not analytical because I_d is not analytical. The most obvious solution to this problem is to use a numerical integration method. Such procedures are generally less efficient than analytical methods, and so before applying them, it is sensible to look for ways in which computational redundancy can be eliminated.

Throughout this document, we have made the assumption that the reactance of the smoothing capacitor is very small (i.e., analytically zero) at the excitation frequency. This is actually not an approximation in the determination of the DC output for a given detector input, because RF currents in the smoothing capacitor cannot affect the average (recall that sinusoids average to zero over an integer number of cycles), but it will affect the detector input impedance slightly. Still, this effect, essentially an adjustment of the distribution of harmonic currents, is very small in comparison to the issue we are trying to resolve; which is whether or not it is safe to neglect the dissipation of harmonic energy in the diode series resistance. Therefore we will retain the large-smoothing-capacitance approximation, which has the effect of eliminating output-voltage droop in the interval between the diode-current pulses.

If the DC output voltage does not droop, then the current-spike that occurs at the peak of the input voltage waveform will be symmetric about the maximum (as in the diagram on the right). If that is the case then, assuming that the input voltage is defined as $V_p \sin\varphi$, the area under the curve between $\pi/2$ and $3\pi/2$ radians will be the same as the area under the curve between $3\pi/2$ and $5\pi/2$ radians (where $5\pi/2$ is the same as $\pi/2$ in the next cycle of the waveform). It follows that the average current can be found by integrating the instantaneous current between the limits $\pi/2$ and $3\pi/2$ and dividing the result by $3\pi/2 - \pi/2 = \pi$.

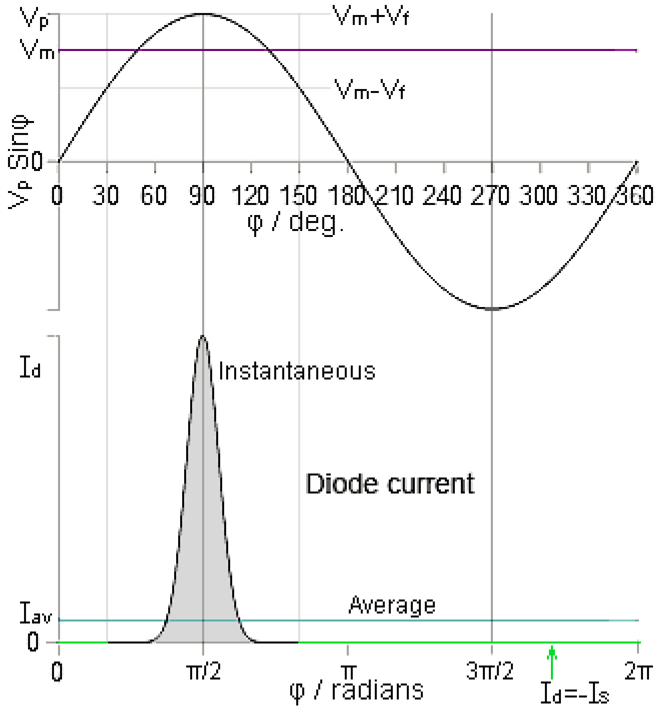


Fig. 15.2

Thus:

$$I_{av} = \frac{1}{\pi} \int_{\pi/2}^{3\pi/2} I_d \, d\phi \qquad (15.1)$$

This change of limits reduces the computational burden of numerical integration by a factor of 2.
A further reduction in the computational overhead can be had by noting that, when a diode with infinite parallel resistance is reverse biased, the current is limited to the reverse-saturation leakage current. Moreover, it can be shown by simulation (e.g., by playing with the numbers in the spreadsheet **det_models.ods**, sheet 2), that for a diode with $R_{ds} = 0$, the saturation current ($-I_S$) is just reached when the diode reverse bias is the negative of the forward voltage drop under dynamic conditions, i.e., $-V_f$. Since we have to input values of the peak and the DC output voltages (V_p and V_m) in order to perform the integration, we have a value of $V_f = V_p - V_m$ that can be used to find the part of the cycle in which the current is constant. Sensible values of R_{ds} make very little difference to the point at which $-I_S$ is reached; but even if the value chosen is unusually large, the effect will be to increase V_f and thereby reduce the length of the region in which we assume that $I_d = -I_S$.

To find the points at which the diode either comes out of or goes into the reverse saturation, we note that this transition occurs when:

$V_p \sin\varphi_x - V_m = -V_f$

Where φ_x has two solutions because there are two crossover points in each cycle of the driving waveform. Substituting for V_p:

$(V_m + V_f) \sin\varphi_x - V_m = -V_f$

$(V_m + V_f) \sin\varphi_x = V_m - V_f$

$\sin\varphi_x = (V_m - V_f) / (V_m + V_f)$

Thus in the region between 0 and $\pi/2$ we have:

$\varphi_x = \arcsin[(V_m - V_f) / (V_m + V_f)]$

Since we need to integrate between $\pi/2$ and $3\pi/2$ however, the solution we want is:

$\varphi_x = \pi - \arcsin[(V_m - V_f) / (V_m + V_f)]$	15.2

Now the integral becomes:

$$I_{av} = \frac{1}{\pi} \left[\int_{\pi/2}^{\phi_x} I_d \, d\phi + \int_{\phi_x}^{3\pi/2} I_d \, d\phi \right] \qquad (15.3)$$

with the possibility of giving a fairly accurate solution for the second integral by simply setting its indefinite value to $-I_S \varphi$. Thus, applying the limits:

$$I_{av} = \frac{1}{\pi}\left[\int_{\pi/2}^{\phi_x} I_d \, d\phi - \left(\frac{3\pi}{2} - \phi_x\right) I_S\right] \quad (15.4)$$

There is however, a good reason why using this form might not be a good idea. This is that the solution for the diode forward voltage, $V_f = V_p - V_m$, has to be determined by a process of iteration. Thus, when the integration routine is called at the beginning of the iteration process, the submitted values of V_p and V_m might not be very accurate. If that is so, then the crossover point will not be accurate, and the assumption that the function is flat from φ_x to $3\pi/2$ might not be valid. Thus actual evaluation of the second integral in (15.3) is advisable. There will still be a large saving in computation time however because, as we will see, the numerical integration process can be made adaptive, so that the integration of the nearly-flat region will be much faster than the integration of the serpentine region from $\pi/2$ to φ_x .

Another episode of the Simpsons

The most basic numerical integration method is that of dividing the area under a curve into contiguous strips, drawing a straight line between the two points at which the curve intersects each strip, and then adding the areas of all the resulting trapezoids between the desired limits. This, of course, is the trapezoid rule. Since it involves straight line approximations to the curve, it can give an exact value of the integral of any linear (first-order) function. If the function is linear and free from discontinuities, a single strip is sufficient. If the function is of quadratic or higher order however, it produces an approximation, which improves slowly as the number of strips is increased.

The problem with the trapezoid method is that the reason for wanting to perform a numerical integration is that there is no known analytical solution. This usually means that the integrand is effectively of high order; i.e., it has highly variable curvature; which suggests that using a first-order approximation for each curve segment is going to result in a large number of strips if it is to give useful accuracy. An obvious improvement therefore, is to increase the order of the function used to approximate curve segments.

Numerical integration using a second-order (quadratic) function to approximate each line segment is known as 'Simpson's rule' [98], or more informatively, 'Simpson's 1:4:1 rule'.

$$\int_a^b f(x)dx \approx \frac{(b-a)}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right] \quad (15.5)$$

The second-order approximation to the integral turns out to be the 1:4:1 weighted average of the function evaluated at the lower limit, the mid point, and the upper limit, of the segment. If the integrand is second-order and continuous across the interval, then the solution is exact. Otherwise the value returned is an approximation, which improves as the difference between the limits a and b is reduced.

It is possible to fit the curve segments to functions of even higher order. In fact, the trapezoid rule and Simpson's 1:4:1 rule are the first two in a series known as the 'Closed Newton-Cotes formulae' [99]. The cubic method is Simpson's 1:3:3:1 rule, and the quartic method is Boole's rule, which uses the 7:32:12:32:7 weighted average; the function being evaluated at equally-spaced intervals over the integration range in each case. Notice that the function always has to be

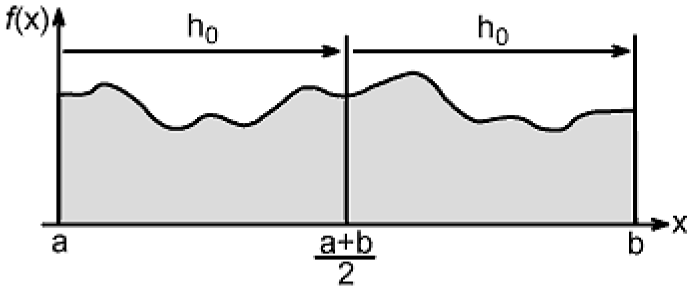
98 http://en.wikipedia.org/wiki/Simpson's_rule
99 http://en.wikipedia.org/wiki/Newton-Cotes_formulas

evaluated at p+1 points, where p is the order of the approximating function.

As the order of the fitting function is increased, the number of segments required to achieve a given degree of accuracy is reduced. Increasing the order however, requires increasing the number of points to be evaluated across a segment. Consequently there is not a linearly-increasing advantage in increasing the order in comparison to the simple expedient of using more segments. The greatest benefit is obtained by going from first-order to second-order, and for that reason, Simpson's 1:4:1 rule is the most widely-used method. Another advantage of the 1:4:1 rule is that it can be implemented in a manner that automatically determines the number of segments required without the need to use arrays to store function-evaluation data.

Most code examples for Simpson's-rule integration require the number of segments to be given as an input parameter. Presumably, the user will have performed an exhaustive evaluation of the entire working parameter-space prior to using the method to solve actual problems, or will have some other clairvoyant way of knowing that the result will be sufficiently accurate with a particular choice. The 1:4:1 weighted average however, is convenient for an algorithm that successively doubles the number of segments without either wasting the previous calculations or needing to store them explicitly[100]. The new result after each doubling is checked against the previous result, and the process continues until convergence occurs. In this way, Simpson's-rule integration becomes adaptive, and advance characterisation of the integrand is not required.

The version of the adaptive algorithm used here works as follows. Firstly, a crude estimate of the integral is obtained by applying Simpson's rule across the entire integration range. This is shown in the adjacent diagram, where the interval between the three samples of $f(x)$ is designated h_0, so that:



$h_0 = (b-a)/2$

If we call this initial estimate S_0, then we get (by comparison with equation **15.5**):

$S_0 = [\, f(a) + 4f(a+h_0) + f(a+2h_0) \,]\, h_0 / 3$

We can however, also de-construct the problem in a different way by designating the samples as odd or even according to the number of instances of the sampling interval that have been added to the lower limit. Then we can write the sums of odds and evens separately. For the initial estimate we get:

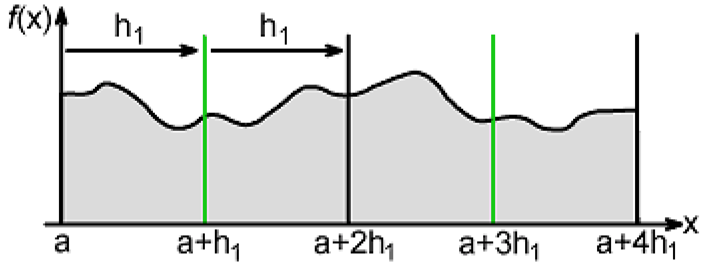
$\Sigma_0(\text{even}) = f(a) + f(a+2h_0)$

$\Sigma_0(\text{odd}) = f(a+h_0)$

so that:

$S_0 = [\, \Sigma_0(\text{even}) + 4\Sigma_0(\text{odd}) \,]\, h_0 / 3$

100 The author was introduced to this method by Bob Weaver, who sent a code example in a private e-mail communication, 26th April 2010, 06:20. Bob also describes the method and gives examples on his '**Numerical Methods for Inductance Calculation**' web page: http://electronbunker.ca/CalcMethods2c.html (accessed 25th Sept. 2014)



Now we double the number of segments by halving the sampling interval.

$h_1 = h_0 / 2$

If we call the new estimate S_1, then we get:

$S_1 = [\,f(a) + 4f(a+h_1) + f(a+2h_1)\,]\,(h_1 / 3) + [\,f(a+2h_1) + 4f(a+3h_1) + f(a+4h_1)\,]\,(h_1 / 3)$

i.e.:

$S_1 = [\,f(a) + 4f(a+h_1) + 2f(a+2h_1) + 4f(a+3h_1) + f(a+4h_1)\,]\,(h_1 / 3)$

but now let us separate-out the odds and evens:

$\Sigma_1(\text{even}) = f(a) + 2f(a+2h_1) + f(a+4h_1) = \Sigma_0(\text{even}) + 2\,\Sigma_0(\text{odd})$

$\Sigma_1(\text{odd}) = f(a+h_1) + f(a+3h_1)$

and

$S_1 = [\,\Sigma_1(\text{even}) + 4\Sigma_1(\text{odd})\,]\,h_1 / 3 = [\,\Sigma_0(\text{even}) + 2\,\Sigma_0(\text{odd}) + 4\Sigma_1(\text{odd})\,]\,h_1 / 3$

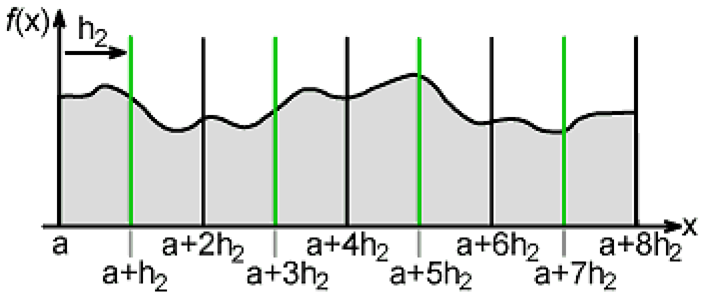
Thus it turns out that the summations from the previous round can be combined to give the evens for the current round, so that it is only necessary to calculate the odd samples. The process is also recursive, so that:

$\Sigma_k(\text{even}) = \Sigma_{(k-1)}(\text{even}) + 2\,\Sigma_{(k-1)}(\text{odd})$

and

$S_k = [\,\Sigma_k(\text{even}) + 4\Sigma_k(\text{odd})\,]\,h_k / 3$

The loop indexing required for the calculation of each new set of odd samples is however not particularly obvious, because the number of points increases exponentially. This is perhaps illustrated by the case for k=2, shown on the right. The following table unravels the problem.



cycle no. k	No. of segs.	odd evaluation points, $a + i\,h_k$	n = max. value of i	No, of segs used
0	1	i = 1	1 = 2(0+1)-1	
1	2	i = 1 , 3	3 = 2(1+1)-1	
2	4	i = 1 , 3 , 5 , 7	7 = 2(3+1)-1	
3	8	i = 1 , 3, 5 , ... , 15	15 = 2(7+1)-1	
4	16	i = 1 , 3, 5 , ... , 31	31 = 2(15+1)-1	
k	2^k	i = 1 , 3, 5 , ... , n_k	$n_k = 2(n_{(k-1)}+1)-1$	$n_{segs} = (n_k+1) /2$

The sum of odd samples can be calculated in a for-next loop, with a step size of 2. As can be seen from the table, the maximum loop count for the k^{th} iteration, n_k, can be calculated from the maximum loop count for the $(k-1)^{th}$ iteration using:

$$n_k = 2 (n_{(k-1)} + 1) - 1$$

The termination criterion can be based on $|S_n-S_{n-1}|$ falling below a certain value, or $|1-S_n/S_{n-1}|$ falling below a certain value, depending on the need for an absolute or a relative measure of convergence.

The total number of segments used for a calculation can be determined from the final maximum loop count using:

$$n_{segs} = (n_k+1) /2$$

The total number of points calculated is $2n_{segs} +1 = n_k+2$

A numerical solution for diode average current when R_{ds} is finite, using the adaptive method just described, is given below. For I_d values, it makes successive calls to the diode instantaneous current routine **DinstI**( ) described in **section 15.1**. The form of the integral is as per equation (**15.1**)

>>>>>> work in progress

Below is the first version of the algorithm, which performs the integral between π/2 and 3π/2 in one go. It works fine as a proof of concept, but it is slow. The much faster version that breaks the integral into two parts at the point φ_x is in the spreadsheet macro library, but I haven't documented it yet or produced the final fully optimised version.

>>>>>>>>>>

```
Function DavI(byval Vp as double, Vm as double, mVT as double, _
Isat as double, Rds as double) as double
'Calculates average diode current when diode has finite series resistance (Rds)
'Calls function DinstI( ). 'Adaptive Simpson's rule integration procedure
'based on a code example supplied by Bob Weaver (eml 2010:04:26 06:20).
Dim a as double, h as double, sumev as double, sumod as double
Dim previous as double, integral as double, diff as double
Dim n as integer, i as integer
a = pi/2
h = pi/2
sumev = DinstI(Vp*sin(a)-Vm, Rds, mVT, Isat) + DinstI(Vp*sin(3*a)-Vm, Rds, mVT, Isat)
sumod = DinstI(Vp*sin(2*a)-Vm, Rds, mVT, Isat)
previous = (4*sumod+sumev)*h/3
n=1
do
  sumev = sumev + 2*sumod
  sumod = 0
  n = 2*(n+1)-1
  h = h/2
  for i = 1 to n step 2
    sumod = sumod + DinstI(Vp*sin(a+i*h)-Vm, Rds, mVT, Isat)
  next
  integral = (4*sumod + sumev)*h/3
  diff = integral / previous -1
  previous = integral
loop until abs(diff) <= 1E-9
DavI = integral / pi
end function
```

Note that for the simple detector case, I_{av} can be calculated from (V_p-V_f)/ R_D , where V_f is obtained using the routine **DVfp2m**(R_D ; V_p ; mV_T ; I_S) (see **section 13.3**). If R_{ds} is set to zero in the call to **DavI**(V_p ; V_m ; mV_T ; I_S ; R_{ds}) above, and the convergence criterion of ≤ 1 part in 10^9 change since last cycle of iteration (as shown) is used, then, for practical input voltages, the numerical method uses 64 segments and agrees with the Bessel function method to better than 9 decimal places of micro amps (see spreadsheet **det_models.ods**, sheet 2).

15.3 Output voltage from peak input voltage

The problem of calculating the output voltage. V_m, from a given peak input voltage for the simple detector was addressed in **section 13.3**. An iterative approach is required because calculation of the diode average current, from which the output voltage is obtained, requires values of both the peak input voltage and the output voltage. Thus it is necessary to input trial values of V_m and adjust them until the calculation agrees with the input. The same conditions prevail when the diode has non-zero series resistance, although the practicalities are somewhat different.

Using the diode average current routine developed in the previous subsections we have:

$V_m = I_{av} R_D = R_D$ **DavI**(V_p ; V_m ; mV_T ; I_S ; R_{ds})

If we call the two instances of V_m y and x, then:

$y = I_{av} R_D = R_D$ **DavI**(V_p ; x ; mV_T ; I_S ; R_{ds})

A good first estimate for x can be obtained by assuming that $R_{ds} = 0$ and using the routine **DVfp2m**(R_D ; V_p ; mV_T ; I_S). This works very well for small inputs, because the voltage across R_{ds} makes little contribution to the overall forward voltage when the diode current is small. Subsequent adjustments to x however need to be determined with some precision, because a small change in x gives rise to a large change in y, and the calculation can easily become unstable. We therefore need to use the derivative $\partial y/\partial x$ to determine the shift, with the small difficulty that this derivative cannot be expressed analytically. A solution is to use finite-difference numerical differentiation.

$\partial y/\partial x \approx R_D$ [**DavI**(V_p ; $x+\delta x/2$; mV_T ; I_S ; R_{ds}) + **DavI**(V_p ; $x-\delta x/2$; mV_T ; I_S ; R_{ds})] / δx

The finite difference, δx , is determined by trial and error, and in this case, values between 10^{-3} and 10^{-6} were found to work well. Notice incidentally, that some implementations of finite-difference differentiations use $f(x)$ and $f(x+\delta x)$ respectively for the two function evaluations. This saves computation time, because $f(x)$ will already have been calculated, but it also reduces the accuracy of the result and introduces bias into any residual error that remains when the calculation is deemed to have converged.

Once we have a derivative, calculation of the shift is as in the various similar procedures in the set of detector utilities accompanying this work:

$\delta x \approx (x - y) / (\partial y/\partial x - 1)$

A Basic routine that performs the calculations given below. Note that it returns the diode forward voltage drop V_f, which can be subtracted from the peak input voltage to determine the detector DC output.

Regarding the matter of computational efficiency; notice that the function makes 3 calls to the numerical integration routine **DavI**( ) on each round of iteration. **DavI**( ) in turn, if it uses say 64 segments, will make 127 calls on the instantaneous current routine **DinstI**( ) each time it is called. When using interpreted Basic, execution is is slow; except for small input voltages, in which case R_{ds} makes very little difference to the result and the initial estimate is close to the final value.

>>>>>>>>>>>>> work in progress

The procedure below has a numerical instability. The problem has yet to be solved.

see spreadsheet **det_models.ods**, sheet 9

>>>>>

```
Function DVfRp2m(RD as double, Vp as double, mVT as double, Isat as double, _
Rds as double) as double
'calculates forward voltage of a diode half-wave detector
'when diode has finite series resistance. 'Calls functions: DVfp2m( ) and DavI( )
Dim Vf as double
Vf = DVfp2m(RD, Vp, mVT, Isat)      'starting estimate assumes Rds=0
DVfRp2m = Vf
If Rds <= 0 OR Vp/mVT > 708 then exit function
Dim x as double, y as double, dx as double, deriv as double, diff as double, deltax as double
x = Vp-Vf
dx = 0.000001          'sets finite difference in [volts] for numerical differentiation.
do
  y = Rd*DavI(Vp, x, mVT, Isat, Rds)
  diff = x-y
  deriv = Rd*( DavI(Vp, x+dx/2, mVT, Isat, Rds)-DavI(Vp, x-dx/2, mVT, Isat, Rds))/dx
  deltax = diff/(deriv-1)
  x=x+deltax
loop until abs(diff) < 1E-7       'sets precision for termination.
DVfRp2m = Vp-x
end function
```

15.4 Comparison of numerical integration and transformation methods

15.x Peak input voltage from output voltage

Works down to 1 microvolt with dx=0.00001
reducing dx beyond that causes no. of segs used in DavI() to exceed maxint (32767).
Note that there is little to be gained from including Rds in the model when Vm is very small.
Adaptive termination criterion is needed to avoid stability issues. Forced termination for n>256 was required during debugging to prevent infinite looping. Leaving it in removes the risk of 'program not responding' errors when termination criterion is adjusted.

>>>> work in progress

Algorithm has a singularity problem because deriv-1 passes through zero in the valid argument range. Might need a completely different approach - needs more thought. Whether this is worth the effort depends on comparison with method of section **14**]

>>>>>>>

```
Function DVfRm2p(RD as double, Vm as double, mVT as double, Isat as double, _
Rds as double) as double
'calculates peak detection error (forward voltage) of a diode half-wave detector
'when diode has finite series resistance. 'Calls functions: DVfm2p() and DavI()
Dim Vf as double, x as double
Vf = DVfm2p(RD, Vm, mVT, Isat)    'starting estimate assumes Rds=0
DVfRm2p = Vf
x = Vm + Vf    'x is the initial estimate for peak input voltage Vp
If Rds <= 0 OR x/mVT > 708 then exit function
Dim y as double, dx as double, deriv as double, diff as double, deltax as double
Dim maxdiff as double
dx = 0.00001    'sets finite difference for numerical differentiation
maxdiff = 1E-6/Vm    'Adaptive termination criterion (less strict for small Vm)
Dim n as integer
do
  n=n+1  'For forced loop termination.  Not used in calculation
  y = Vf + Rd*DavI(x, vm, mVT, Isat, Rds)
  diff = x-y
  deriv = Rd*( DavI(x+dx/2, Vm, mVT, Isat, Rds)-DavI(x-dx/2, Vm, mVT, Isat, Rds))/dx
  deltax = diff/(deriv-1)
  x = x+deltax
  Vf = x-Vm
loop until abs(diff) < maxdiff or n > 256
DVfRm2p = Vf
end function
```

99. Work in progress

>>>>>>>>>>> Under construction

Remaining topics to be covered in part 1:
- Comparison of transformation and numerical integration methods.
- Error analysis (calculation of measurement standard deviation)
- Fitting diode data and extraction of diode parameters.
>>>>>

xx. Comparison of methods.

xx. Error analysis for absolute AC voltage measurement.
Need to show how uncertainties propagate into the working parameters and into the final result.
+ Routines for calculating the standard deviation.

xx. Parameter determination

Stray points:

>>> Heavy smoothing - loss of generality. Timescale ratios. Video 10:1, voltmeter 10000:1. Can be included in numerical integration.

>> Diode detector input impedance. Comment under general model:
>>
Note also that the input impedance of a detector will often be slightly capacitive at high frequencies. This is due to the junction capacitance of the diode. For the 1N5711 Schottky diode (for example), this capacitance is about 2pF. Hence placing several diodes in parallel is not a preferred method for improving the linearity of broadband or high-frequency detectors.

>>>>>>>>>> Orphan refs:

"**Errors in SWR Meters**", Albert E Weller, WD8KBW, QEX Correspondence, June 1982, p2.
Errors due to diode non-linearity.

"**Diode Voltmeters**", Albert E Weller Jr., WD8KBW, QEX, April 1983, p3, cont p6.
Non-linearity compensation scheme.

These go with:
"**The Theory of Diode Voltmeters and Some Applications**", Albert E Weller, WD8KBW, QEX Jan 1984, p7-14.

"**Calibrating Diode Detectors**", John Grebenkemper, KI6WX, QEX, Aug 1990, p3-8.
[This article confuses saturable and non-saturable leakage current and consequently makes a mess of diode parameter extraction].

>>> Early draft of parameter extraction article.

Diode correction function
When using a diode detector to make accurate RF voltage measurements, mathematical correction for diode non-linearity can be obtained by fitting the diode forward conduction characteristic to an expression of the form:

$V_f = V_1 \ln(I_f) + V_0 + I_f R_{ds}$

Where R_{ds} is the diode series resistance, $V_0 + I_f R_{ds}$ is the forward voltage when $\ln(I_f) = 0$, and V_1 is the gradient of the corresponding graph. We can see how this simple correction function comes about by starting with the diode equation and including a term to allow for the fact that the diode will have some ordinary series resistance (R_{ds}). Hence:

$V_f = I_f R_{ds} + m V_T \ln[(I_f / I_S) +1]$

V_f has to go to zero when $I_f = 0$. Hence the +1 term inside the logarithm bracket is there to set a limiting condition, which is:

$\ln[(I_f / I_S) +1] \rightarrow 0$ as $I_f \rightarrow 0$

I_S for a Schottky signal diode is usually somewhere around 10^{-9} Amps (i.e., 1nA). Hence, when the detector is used with inputs somewhat greater than the forward conduction threshold, then $I_f / I_S >> 1$, and we can neglect the +1 term without noticeable effect. This leads to the large input approximation:

$V_f = If R_{ds} + mV_T \ln(I_f / I_S)$

But we don't know I_S , and so we perform a substitution to capture this lack of knowledge in a dimensionless parameter (I_{ref} / I_S). Thus:

$V_f = I_f R_{ds} + mV_T \ln[(I_f / I_{ref})(I_{ref} / I_S)]$

Now, making use of the identity ln(pq)=ln(p)+ln(q), we get:

$V_f = I_f R_{ds} + mV_T \ln(I_{ref} / I_S) + mV_T \ln(I_f / I_{ref})$

which is in the form:

$V_f = I_f R_{ds} + V_0 + V_1 \ln(I_f / I_{ref})$

The choice of I_{ref} is arbitrary, and it is sensible to use some convenient engineering multiple of Amps (i.e., usually 1 µA or 1 mA for signal diodes). Hence, adopting $I_{ref} = 1$ µA :

$V_f = I_f R_{ds} + V_0 + V_1 \ln(I_f / [\mu A])$

i.e., the current inserted into the logarithm bracket has effectively been divided by its units to make it dimensionless, which means that all we do in practice is take the logarithm of the forward current in µA to calculate the rightmost term. Now we can rearrange the equation in the form y=a+bx and carry out a linear regression analysis[101], i.e.;

$V_f - I_f R_{ds} = V_0 + V_1 \ln(I_f)$

where $y = (V_f - I_f R_{ds})$, $a = V_0$, $b = V_1$ and $x = \ln(I_f)$

When calibrating a small-signal detector for currents of < 1mA, R_{ds} can often be assumed to be zero if a Schottky or gold-bonded diode is used. The series resistance of some diodes however, particularly the germanium point-contact variety, can be fairly large and so cannot be ignored. In order to allow for a finite R_{ds}; a simple trick in calculation (program or spreadsheet) is to make provision for the term $I_f R_{ds}$ to be subtracted from the measured value of V_f, but initially to set R_{ds} = 0. If the regression-line shows significant curvature, R_{ds} can be adjusted by hand or by iteration to get the smallest standard-deviation of fit.

Once the parameters V_0, V_1 and R_{ds} are determined, the function we started with:

$V_f = V_1 \ln(I_f) + V_0 + I_f R_{ds}$

will return a value of V_f for a given value of I_f that is good over several decades of current (provided that the temperature is close to what it was when the fitting data were collected). Note however, that the regression function does not have the correct limiting condition to return a true value for V_f when $I_f \rightarrow 0$; and so I_f of less than about 100 nA should be trapped as an illegal input. Alternatively we can solve for the diode equation parameters using:

$V_1 = mV_T$ and $V_0 = mV_T \ln(I_{ref} / I_S)$

in which case V_f can be calculated from the diode equation directly and the limit at $I_f = 0$ will be correct..

In an example given elsewhere [see Data Analysis], the regression function for a 1N5711 diode (neglecting R_{ds}) was found to be:

$V_f = 0.158342 + 0.029060 \ln(I_f / [\mu A])$ Volts

Taking V_T =25.3 mV, this gives m=1.15 and I_S = 4.3 nA. Using the diode equation (with the parameters in full precision as calculated from the fit) produces V_f values that are barely different from those given by the fitting function provided that $I_f >> I_S$ (see spreadsheet **1N5711.ods**). Note that the determined diode equation parameters are not necessarily realistic, because R_{ds} had been neglected in this case, but they are accurate for several more decimal places than are required for correcting experimental detector readings (The SPICE parameters for the Agilent 1N5711 [102] are I_S = 2.2 nA. R_{ds} =25 Ω). Attempting to include the $I_f R_{ds}$ correction in the fit resulted in a negative value for R_{ds}, indicating that there are insufficient data to determine the extra parameter in this case. If the best fit is obtained with $R_{ds} < 0$, then the parameter is fitting noise and should be set to 0.

■

101 see, for example, **Data Analysis**, DWK. http://g3ynh.info/zdocs/math/
102 IN5711 data.