Fundamental limitations of THz and Niobiumnitride SIS mixers
Dieleman, Pieter

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1998

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Chapter 2

Materials aspects and SIS characteristics

A tunnel junction mixer needs to display a very small current without applied radiation, and large current when radiation is applied. This must be combined with a sharp nonlinear current-voltage characteristic to enable efficient heterodyne mixing of the signal. In this chapter it is described how these requirements are assimilated in the fabrication process. First the most common techniques to deposit and pattern layers of metals, superconductors and insulators are discussed by reviewing the fabrication process of a superconducting tunnel junction embedded in Al circuitry. Second, the choice and fabrication of a suitable tunnel barrier on a Nb or NbN layer are explained. Third, the electrical characteristics of both junctions are examined and discussed. Fourth, the influence of barrier defects is illustrated. Fifth, materials suitable for low-loss coupling structures of radiation to the junction are examined.

2.1 Fabrication of SIS mixer structures

Substrate requirements

To obtain a chronological description of the fabrication process, we will start with the substrate. The substrate needs to be robust and it needs to exhibit a low radiation absorption loss and a small dielectric constant $\varepsilon_r$. The presence of the
substrate in the mixer block disturbs the electromagnetic wave pattern and therefore the coupling to the probes if the thickness of the substrate is smaller than the wavelength. Since the wavelength in a material is determined by the dielectric constant, a low $\varepsilon_r$ material is favored to keep the thickness and therewith the vulnerability of the substrate within reasonable handling limits. Fused quartz, with its low dielectric constant of 4, is sufficiently robust to survive grinding down to the required 40 $\mu$m.

**Sputter deposition**

The junction and circuit layers are deposited by magnetron sputtering. Sputter deposition is performed by bleeding in a noble gas like argon in a vacuum chamber which contains a target of the material to be deposited, and a substrate on which the material is to be put. The gas is ionized by a voltage applied between the substrate and the target. The gas ions are accelerated by this voltage to the target, which upon impact remove atoms from the target, which partly settle on the substrate mounted nearby. This process is schematically depicted in Fig. 2.1a. The term magnetron refers to a magnet placed in the vicinity of the target, forcing electrons to move in circles thereby increasing the path length and the ionization probability and hence the deposition rate, which is typically a few nm/s. If a niobium junction with an aluminum coupling circuit is made, the first layer deposited is an aluminum layer, which is sometimes deposited on top of a Nb layer to prevent any series resistance. The advantage of this superconducting Nb layer is that it losslessly transports the IF radiation to the IF amplifier. However if a single Al layer is used, the series resistance provides information on the DC conductivity of the material, which is of direct influence on the RF properties. In this manner RF properties deduced from the RF measurements can be compared to the RF properties calculated from the measured DC series resistance.

**Optical lithography**

The Al layer must be patterned to form the RF choke and probe structure, as shown in Fig. 2.1g. This is done by optical lithography. A polymer is used which becomes soluble in a developer liquid upon exposure to ultra-violet (UV) radiation. The polymer substance is called photoresist. A few droplets of the photoresist with solvent are dripped onto the substrate, and by high speed spinning the fluid spreads out and becomes thinner until the solvent is sufficiently evaporated.
and the photoresist layer is hardened enough to inhibit any further reduction of thickness. Subsequently the sample is baked in an oven in ambient atmosphere, to evaporate the remaining solvent. The photoresist layer is then ready to be exposed to UV radiation. A mask is designed which is placed between the UV lamp and the substrate, shielding part of the photoresist layer against the radiation. Afterwards the sample is placed in the developer which dissolves the exposed parts of the photoresist. When now a metal layer is deposited on a substrate with this patterned photoresist structure and the sample is placed in acetone, which dissolves the unexposed photoresist, the part of the metal layer on the photoresist layer is removed. This process is called lift-off, see Fig. 2.1b. When the photoresist pattern is created on an already deposited metal layer, the uncovered pieces of the metal layer can be etched. Note that with etching the positive image of the mask is obtained, whereas the mask image is obtained via lift-off.

### 2.1.1 Nb tunnel junction fabrication

Since 1983 all Nb superconductor - insulator - superconductor trilayers are produced by means of the Gurvitch fabrication process[1]. On the patterned Al ground wiring layer 100 nm of Nb is sputtered, followed by a sputter deposition of 6 nm of aluminum. Al grows on Nb in small grains, effectively smoothing out any irregularities of the Nb surface. This effect is called *wetting*, and results in a smooth surface layer of Al, ready to be oxidized. The Al oxidation takes place in a separate vacuum chamber directly coupled to the main chamber where the layers are sputtered. The Al layer is oxidized in a pure oxygen atmosphere, after which it is transported back into the main chamber for the top electrode deposition, another 100 nm of Nb. The Al layer, although a normal metal at 4.2 K, acquires superconducting properties as a result of its direct contact with the superconductor. This *proximity effect* can be thought of as Cooper pairs, responsible for the superconducting properties, leaking into the normal metal over a certain distance, the normal metal coherence length $\xi_N$. The $\xi_N$ is rather large in Al, allowing for a layer sufficiently thick to completely cover the Nb surface without a gap reduction.
2.1.2 NbN tunnel junction fabrication

NbN is a superconductor with a maximum critical temperature $T_c$ of 18 K, corresponding to a gap energy $\Delta$ of 3 meV. A NbN layer is commonly produced by reactive sputtering in a nitrogen / argon gas mixture. The Nb target is "nitridized"
2.1 Fabrication of SIS mixer structures

and subsequently sputtered onto the substrate. NbN is capable of forming several stochiometric compounds, but only one, the Nb$_1$N$_1$ B1 phase has the required superconducting properties[2]. Therefore the nitrogen concentration in the gas mixture has to be carefully optimized and controlled[3].

The normal state conductance as well as the $T_c$ and surface roughness heavily depend on the deposition conditions used: substrate material (epitaxial / amorphous), substrate temperature (ambient/heated), bombardment intensity (high/low power and pressure, short/long substrate-target distance) and gas mixture (N$_2$/Ar ratio, CH$_4$ addition). All layers deposited on non-epitaxial substrates are polycrystalline. Varying process conditions changes merely the grain size. The larger the grains, the fewer grain boundaries are present per square, lowering the resistance. However, the roughness of a layer is proportional to the grain size and this property is of utmost importance when an oxide layer with a thickness of less than a nanometer has to cover the surface homogeneously.

The current state of development of NbN tunnel junctions is comparable to the early years of Nb fabrication prior to Gurvitch’s discovery. Barrier materials are being tried, process conditions are being varied, but the resulting current-voltage characteristics for high current density junctions required for heterodyne mixing all show large subgap currents. See Fig. 6.1 for a comparison between $I$-$V$ curves of typical Nb and NbN high critical current density junctions. A number of research groups have similar state-of-the-art results: Jet Propulsion Laboratory (JPL) in the USA[4], Institut de Radio Astronomie Millimétrique (IRAM) in France[3,5], and the Electrotechnical Laboratory (ETL) in Japan[6]. All groups have a process of growing polycrystalline NbN on quartz substrates. ETL also has a process of epitaxial base layer growth on MgO substrates. The advantages of epitaxial growth are a generally higher gap energy, and a smooth surface for the barrier deposition. However this process is incompatible with the use of Al striplines, since the Al layer does not carry the MgO epitaxy over to the NbN junction base layer. Therefore the NbN layer on which the barrier is produced, is necessarily polycrystalline. It turns out that a polycrystalline NbN layer with large grains is easily made but is generally unsuitable for barrier deposition. The aforementioned groups have found a trade-off optimum between gap voltage and subgap “leakage” current.

The NbN junctions studied in Chs. 5 and 6 are fabricated by M. Schicke[5] at IRAM, with a recipe adapted by B. Plathner[3], obtained from M. Aoyagi and S. Takada at ETL[7].
Thermal oxide barriers on NbN

Unfortunately a Gurvitch method has not been developed for NbN tunnel junctions. Directly copying the Nb recipe turned out to be unsuccessful[8]. It is energetically favorable for the N from the NbN to hop over to the Al and form AlN[9]. The AlN at the interface between NbN and Al acts as a barrier for the proximity effect, blocking the leakage of Cooper pairs into the Al which lowers the induced energy gap value in the Al at the barrier significantly. This is shown in Fig. 2.2 where an I-V curve is shown of a NbN/Al/AlO\textsubscript{x}/Nb junction made by the Materials Science Group in Cambridge[8]. The expected energy gap value is 4 mV as shown in the theoretical characteristic. The measured curve is convincingly approximated by a theoretical model[10] which takes into account a barrier at the interface between the superconductor and the layer to be proximitized. The barrier strength is given by the parameter $\gamma_b$. The proximity parameter $\gamma_m$ is a measure of the ease with which the material is proximitized. The $\gamma$ parameters are given by:

$$
\gamma_m = \frac{\rho_{NbN}}{\rho_{Al}} \frac{\xi_{NbN}}{\xi_{Al}} \frac{d_{Al}}{2\xi_{Al}} \quad \text{and} \quad \gamma_b = \frac{2}{3} \frac{l_{Al}}{\xi_{Al}} \frac{1 - D}{D} \frac{d_{Al}}{2\xi_{Al}} \quad (2.1)
$$

in which $\rho_{NbN}$ and $\rho_{Al}$ are the respective normal state resistivities, $\xi_{NbN}$ and $\xi_{Al}$ are the respective coherence lengths of NbN and Al. The thickness of the Al layer is given by $d_{Al}$ and the mean free path in Al is given by $l_{Al}$. The variable $D$ stands for the transmission probability of the AlN barrier between Al and NbN. Fig. 2.2 shows the calculated curves for the material properties given in Table 2.1. The same model promised a high quality current-voltage characteristic when a Nb/Al/AlO\textsubscript{x}/Nb trilayer with thin Nb electrodes is sandwiched between thick NbN layers. In practice, it turned out that the Nb layers had to be so thin (2 – 4 nm) to be completely proximitized that part of the rough NbN surface would be left uncovered, again yielding NbN / Al contacts.

Directly deposited barriers

For these reasons, direct deposition of a barrier material is currently practiced. Two materials, MgO and AlN are favored. Both exhibit a lattice match with the NbN crystal structure. Fig. 2.3 shows the NbN face centered cubic (fcc) structure. The [111] planes contain alternatingly N and Nb. MgO also has a fcc crystal structure, with a lattice constant 3 % larger than that of NbN. AlN has a hexagonal close packed (hcp) structure. The [0001] hcp plane is the closest packed plane and
has a lattice match with the closest packed [111] fcc plane of NbN also shown in
Fig. 2.4. The mismatch with AlN is a mere 0.3 %.

The lattice match is most likely the main reason that these materials function
as barrier layers, since usually on unmatched surfaces the first few monolayers
are amorphous. The barrier is only a few monolayers thick, and therefore needs
to grow epitaxially. Epitaxy in this context means local epitaxy, on the grains
of NbN. (This is also the reason that buffer layers of polycrystalline MgO or AlN
increase the quality only of thin (30 – 40 nm) NbN layers, since only growth of the

\[ \begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\gamma_b & \rho_{NbN} & \rho_{Al} & \xi_{NbN} & \xi_{Al} & l_{Al} & D & \gamma_m & \gamma_b \\
\hline
\text{small} & 100 & 5 & 5 & 50 & 20 & 2 & 0.25 & 1 \\
\text{large} & 75 & 5 & 5 & 50 & 20 & 0.1 & 0.1 & 10 \\
\hline
\end{array} \]

FIGURE 2.2. Measured and calculated \( I - V \) curves of a NbN junction with Al/AlO\(_x\)
tunnel barrier. The theoretical curves are calculated with a proximity barrier strength \( \gamma_b \) of
1 and 10. Other material parameters given in Table 2.1.2 are approximately identical for
both theoretical curves.
first nanometers is promoted, after which the layer quality is nearly independent of the preceding layer). Because of the lattice mismatch, heteroepitaxy is not achieved, i.e. the NbN top electrode again starts growing with small grains[11], with lowered $T_c$ and energy gap compared to bulk NbN.
2.1 Fabrication of SIS mixer structures

Junction patterning

After deposition of the SIS trilayer over the whole wafer, the junction is defined by etching away the superfluous material. As shown in Fig. 2.1d a 1 μm² photoresist dot is defined by optical lithography. Then the top Nb or NbN electrode is etched by means of Reactive Ion Etching (RIE) with CF₄ gas with 3 % O₂ addition. In this process an applied voltage causes the gas to be ionized. The ions and radicals created bombard the surface of the sample where the fluor radicals form volatile NbF products which are subsequently pumped away. The C atoms left over from the CF₄, together with the carbon from the photoresist which is also attacked, precipitate on the sample surface. Since the applied voltage causes the horizontal areas to be mainly bombarded with the reactive gas atoms, the carbon at the surface is removed by mechanical bombardment and reaction with O₂ to form CO and CO₂. At the side walls however the precipitated carbon is removed much slower since the bombardment is less severe. Therefore a carbon layer stays at the side walls prohibiting any lateral (sideward) etching. For this reason, with carefully chosen C : F : O atom ratio's in the plasma, RIE etching can be fast with well defined structures having only slightly smaller dimensions than the original photoresist pattern.

Since CF₄ does not attack the AlOₓ or MgO these barrier layers act as etch stoppers. The entire trilayer has to be etched to reach the Al underlayer, therefore the barrier is to be removed to expose the bottom electrode. Several methods are used:

**Wet etching.** The MgO or Al/AlOₓ layers are etched by immersing in a liquid which forms products dissolvable in the liquid[5].

**RF sputtering.** The sample is bombarded by argon ions in a sputter chamber. The developed heat is removed efficiently by employing diffusion pump oil between the substrate and the cooled metal holder. Care has to be taken to remove the oil. Acetone dissolves the oil, but also the photoresist pattern at the top side of the sample. Therefore a cotton bud is used to clean the back side of the sample[12].

**RIE "sputtering".** After etching the top electrode, the pressure is lowered and the power is increased to obtain mechanical erosion of the barrier layer. The CF₄ gas is used as sputter ions to remove the barrier mechanically. The
advantage here is that the sample can stay inside the RIE vacuum chamber, saving time and preventing contamination. A possible problem is the developed heat which can damage the photoresist or the barrier layer[13].

After removal of the barrier the remaining bottom electrode is etched with RIE (See Fig. 2.1d).

To electrically insulate the top wiring from the bottom wiring a layer of SiO$_2$ is deposited by RF sputtering (Fig. 2.1e). (Insulating targets would be charged when bombarded with positive ions only, therefore the polarity is reversed continuously (RF) to obtain electron bombardment and discharging of the target. The same is done in the RIE when the bombardment takes place on an insulating substrate). The photoresist used as an etch mask is now used for lift-off. This poses additional requirements on the trilayer etch process, since the photoresist shape degradation is to be minimized for a successful lift-off.

For the top wiring a thick (300 nm) Al layer is used since it also functions as a stripline. This layer can be defined with lift-off as well as wet-etching, dependent on the requirements on accuracy of the dimensions. A gold layer is deposited and patterned by wet etching to avoid oxidation at the contact pads. A side view of the finished junction with wiring is shown in Fig. 2.1f.
2.2 \textit{I-V} characteristics vs. barrier structure

2.2.1 Aluminum oxide growth

An interesting question which is quite important for some of the effects addressed in this thesis is how the aluminum oxide barrier grows and what the structure of the resulting layer is. Soon after the initial discovery of the Nb/AlO$_x$/Nb system\cite{1} it was investigated which properties of the Nb, Al or AlO$_x$ were the cause of the outstanding electrical characteristics by Braginsky \textit{et al}\cite{14}. A number of clear conclusions were drawn:

- The Nb-Al bond is stronger than the Al-Al bond. This facilitates complete coverage of the Nb electrode with even a very thin Al layer.

- At room temperature sputtered Al grows on Nb in small grains because of the relatively low surface mobility and hence a high Al island nucleation density\cite{15}. This facilitates a homogeneous coverage of the Nb.

- On polycrystalline Al an amorphous AlO$_x$ film grows, which is a better tunnel barrier than an epitaxial film\cite{14}.

The initial growth of an oxide layer on a polycrystalline Al layer has been investigated in detail for growth at room temperature at an oxide partial pressure of $10^{-8}$ – $10^{-5}$ Torr. The oxide exposure is usually defined in Langmuirs, $L$ (1 $L$ = $10^{-6}$ Torr $\cdot$ s) since the oxide growth rate depends similarly on the time as on the oxide partial pressure. According to Arranz \textit{et al}\cite{16} the oxide growth can be divided in three stages:

\begin{itemize}
  \item \textbf{a} \textbf{0} – \textbf{10} $L$. Single layer oxygen islands are chemisorbed. These grow to completely cover the Al at 10 $L$.
  \item \textbf{b} \textbf{10} – \textbf{1000} $L$. Oxide island nucleation starts, followed by lateral growth after an average thickness of 3 monolayers is reached.
  \item \textbf{c} $>\textbf{1000} L$. The coverage of the AlO$_x$ is nearly complete and the average thickness increases logarithmically with time and pressure.
\end{itemize}

Clearly Braginsky\cite{14} and Arranz\cite{16} present conflicting results, since one assumes amorphous oxide layers and the other monolayers of oxide on polycrystalline aluminum layers. Further studies of aluminum oxide growth, concentrated
on the process conditions required for a homogeneous tunnel barrier, are needed to resolve this discrepancy.

### 2.2.2 Dependence of electrical properties on $J_c$

When designing an SIS mixer, a number of electrical parameters are chosen to optimize the matching at the RF and IF side of the mixer and to maximize gain and minimize noise. These parameters are:

- The normal state resistance $R_N$, commonly in the range of 15–30 $\Omega$. $R_N$ is determined by the critical current density $J_c$ and the junction area $A$ via $R_N = \frac{\pi \Delta}{2e J_c A}$. Since the area is usually chosen close to 1 $\mu m^2$, the smallest practical limit of optical photolithography, the junction resistance is solely determined by the critical current density. The required critical current density is obtained by exposing the aluminum layer to a well controlled oxygen pressure for a fixed time. The dependence of $J_c$ on the oxygen exposure is depicted in Fig. 2.5, in which the solid lines are fits to data points obtained by Mallison et al [17], and the solid triangles present data found for our junctions.

- The capacitance of the junction is the product of the specific capacitance and the junction area $A$. The specific capacitance depends on the oxide thickness and dielectric constant of the oxide material and on the surface roughness, which effectively increases the area. A small capacitance facilitates matching to the RF inputs and increases the bandwidth of the device, therefore the area is chosen as small as practically possible, close to 1 $\mu m^2$. While tuning the oxygen pressure to obtain the required aluminum oxide thickness and hence the right $J_c$, the capacitance is changed as well. The dependence of the capacitance $C$ on the critical current density as measured by Van der Zant et al [18] (round dots) and Maezawa et al [19] is shown in Fig. 2.6. Although both found different dependences, it is clear that the capacitance depends heavily on $J_c$.

- The current-voltage characteristics of the SIS junction should ideally be independent of the current density apart from the normal state resistance. It will be shown in Ch. 7 however that the magnitude of the subgap current can be much larger than expected from tunneling theory, and this discrepancy
2.2 I-V characteristics vs. barrier structure

![Graph showing the dependence of $J_c$ on O$_2$ exposure.]

**Figure 2.5.** Dependence of $J_c$ on O$_2$ exposure (pressure times time). The straight lines are fits taken from Ref. [17], the triangles are data obtained from our own niobium junctions.

increases with increasing $J_c$. Therefore with increasing $J_c$ the shot noise generated by the mixer increases and the mixer gain gradually decreases.

By changing the oxygen pressure not only the normal state resistance is tuned, but the capacitance and the subgap current is affected as well, resulting in a simultaneous change of the matching at the RF and IF side and the gain and noise of the mixer itself.

**Dependence of the capacitance on $J_c$**

To present a comprehensive view of the impact of the barrier structure on the capacitance and the critical current density, we assume for the sake of the argument that we may assume the barrier to be Al$_2$O$_3$, of crystalline nature. (The aluminum oxide barrier is henceforth denoted by AlO$_x$ when the structure is unimportant). Let us furthermore assume that the barrier has areas with two monolayers, one monolayer and no monolayer, i.e. shorts, the nature of which will be described in the next section. This concept is illustrated in Fig. 2.7. Then the
critical current (for simplicity taken as the current at $V = 2\Delta/e$) for an overall area $A = A_1 + A_2 + A_0$ is given by

$$I_c = I_0[A_1T_m + A_2T_m^2 + A_0T]$$

(2.2)

with $T_m$ the tunnel transmission coefficient of one monolayer. The transmission depends exponentially on the barrier thickness, therefore the transmission of a barrier region consisting of two monolayers is $T_m^{-2}$. The pinhole transmission $T$ is derived in the next section. $A_n$ stands for the area. The monolayer transmission $T_m$ should be on the order of $10^{-4}$.

The capacitance $C$ of the junction depends quite differently on the area. The pinhole area $A_0$ is assumed to be less than 0.1% of $A$ and therefore has negligible contribution to $C$. We assume that

$$C = C_1 + C_2 = \frac{\varepsilon_0\varepsilon_r}{d}(A_1 + \frac{A_2}{2})$$

(2.3)

with $d$ the thickness of a monolayer.
Hence controlling the average critical current by varying the oxygen pressure is predominantly varying the ratio of $A_1/A_2$. As a consequence also the effective $C$ varies with oxygen pressure or $J_c$. Within this framework we can now examine what the relation between $J_c$ and $C$ is since we know the dependence of each on the relative areas. We make the assumption that $A_2$ and $A_1$ are of comparable size, the pinhole transmission $T$ is set to one, and we maintain that $A_2$ has a negligible contribution to $J_c$, but we include its contribution to $C$:

$$j_{cm} = \frac{A_1}{A}T_m + \frac{A_0}{A}$$  \hspace{1cm} (2.4)$$

$$\frac{C}{A} = \frac{\epsilon_0 \varepsilon_r}{d} \left[ \frac{A_1}{A} + \frac{1}{2} \frac{A_2}{A} \right]$$ \hspace{1cm} (2.5)$$

We define $p = A_0/A$ and $q = A_2/A$, hence

$$j_{cm} = (1 - p - q)T_m + p$$ \hspace{1cm} (2.6)$$

$$\frac{C}{A} = \frac{\epsilon_0 \varepsilon_r}{d} \left[ (1 - p - q) + \frac{1}{2} q \right] = \frac{\epsilon_0 \varepsilon_r}{d} \left[ (1 - p - \frac{1}{2} q) \right]$$ \hspace{1cm} (2.7)$$
but we know that $p$ must be small otherwise the tunnel curve would be less tunnellike:

$$\frac{C}{A} = \frac{\epsilon_0 \epsilon_r \tau_1}{d} [1 - \frac{1}{2q} \eta]$$

and with respect to the current density we may write:

$$j_{cn} = (1 - q)T_m + p[1 - T_m] \approx [1 - q]T_m + p \rightarrow q = 1 - \frac{j_{cn} - p}{T_m}$$

substituted in $C/A$ we find:

$$\frac{C}{A} = \frac{\epsilon_0 \epsilon_r \tau_1}{d} [1 - \frac{1}{2} \frac{j_{cn} - p}{T_m}] = \frac{\epsilon_0 \epsilon_r \tau_1}{2d} [1 + \frac{j_{cn} - p}{T_m}]$$

A linear increase with $J_c$ provided $j_{cn} > p$ is obtained as seen from Eq. 2.10, in accordance with the result obtained by Van der Zant et al. [18] as shown in Fig. 2.6. The $j_{cn} > p$ requirement means that the fraction of the area which carries pinhole current (times the transmission $T$ which here is put to 1) is smaller than the monolayer area times its transmission coefficient $T_m$. This is equivalent to assuming that at $V \gg 2\Delta/e$ the current is dominated by the monolayer part. The empirical facts important for the mixer design described in the beginning of Section 2.2.2 are now qualitatively understood.

**Dependence of the subgap current on $J_c$**

Fig. 7.1 shows the development of the measured subgap current at 2/3 of the gap voltage divided by the theoretically expected thermal current[21] at 4.2 K. This thermal current is carried by quasiparticles thermally excited above the energy gap which allows them to tunnel into the other superconductor and is calculated by the standard model for the tunneling $I-V$ characteristics[21]:

$$I = \frac{1}{e R_N} \int_{-\infty}^{\infty} N_s(E)N_s(E - eV)[f_{FD}(E - eV) - f_{FD}(E)]dE$$

in which $N_s$ is the density of states in the superconductors, and $f_{FD}$ stands for the Fermi-Dirac distribution. The data points are taken from articles published by various groups[13,22–27]. Clearly for all junctions studied the subgap current increases approximately linearly with increasing $J_c$. Since the normalized thermal tunnel current is independent of the barrier thickness for an ideal tunnel
2.2 I-V characteristics vs. barrier structure

barrier, apparently with increasing current density the barrier becomes too thin to homogeneously cover the aluminum. Because of statistical fluctuations in the oxide layer thickness, at some locations oxide atoms will not be present at all. The resulting superconductor-constriction-superconductor (ScS) electrical contacts thus created are called pinholes, small shorts with transmissions close to unity[18,20,28,29].

Therefore clearly not only the capacitance depends on the current density, the actual I-V curve is affected as well!

The consequence of this analysis is that we have to assume 3 different areas to reconcile the Van der Zant et al results and our own data. $A_2$ and $A_1$ are needed to get sensible capacitance data. $A_1$ and $A_0$ are needed to account for the subgap current compared to the current above the gap voltage.

The reasoning presented above clarifies the concept of a tunnel barrier with patches with uniform thickness and pinhole conduction in parallel. With increasing current density, the average barrier thickness decreases, because $A_1$ increases at the expense of $A_2$, and $A_0$ increases at the expense of $A_1$. This mechanism provides an intuitively simple way to derive all junction parameters important for the mixer performance from the designed current density.

This view, although illustrative, will need some fine-tuning, since it predicts a linear dependence of the capacitance on the critical current as observed by Van der Zant[18], whereas for junctions with similar characteristics Maezawa et al[19] found a logarithmic dependence, as shown in Fig. 2.6.

2.2.3 Current transport in pinholes

In the previous section we concentrated on the effect of an increasing current density on the capacitance and the subgap leakage current, in which we simply described the pinholes as highly transmissive shorts. To understand fully the influence of pinholes on the J-V characteristics and the mixer performance, the actual electronic conduction mechanism in these pinholes is treated in this section.

Because of the high transmission probability of these pinholes, the dominant conduction process is multiple Andreev reflection (MAR)[30,31] rather than single-electron tunneling. The concept of Andreev reflection is visualized in Fig. 2.8. A pinhole consists most likely of a single atomic link between two superconductors. Although the MAR process occurs also in a constriction with finite $\Delta$ (See Fig. 2.9, obtained from Ref. [32], in which a superconducting Al
Figure 2.8. Schematic image of the Andreev reflection process. For different energies an incoming electron can follow different paths and get Andreev reflected 1, 2, ... times until it has gained enough energy to tunnel into the right superconductor.

Point contact is measured, it is easier to visualize the electron motion in the constriction assuming that the constriction is a very short piece of normal metal ($\Delta = 0$). In this way the transport through the resulting superconductor-normal metal-superconductor (SNS) structure can be described directly in terms of the Andreev reflection at the two NS interfaces[33]. In what follows, we use this framework for clarity.

When an electron coming from a normal conductor hits the interface with a superconductor, an additional electron from the conductor can be dragged into the superconductor to form a Cooper pair. The missing electron is called a hole and can travel back towards the superconductor on the other side, and undergo another Andreev reflection, in which a Cooper pair is destroyed. The remaining particle is again an electron and the sequence starts all over, unless the electron has gained sufficient energy to escape from the inner region into the superconductor at an energy level above the energy gap.

Obviously, the transmission probability of the middle region plays a role in the number of reflections a charge carrier will experience. At a transmission smaller than unity the current carried by these MAR rapidly decreases with decreasing voltages. This is shown in Fig. 2.9 where $I$-$V$ characteristics are plotted from an
aluminum break junction for several transmission values of the contact. This feature we have used to obtain the average transmission probability of our pinholes. As described by van der Post et al[34] a linear relation exists between the current steps at subharmonics of the gap voltage and the transmission probability of the constriction. The transmission $T$, defined here as $T = R_Q/R_N = \frac{\hbar^2}{2\pi^2} / R_N$, the quantum resistance over the normal resistance is measured to be 0.5 for pinholes in Nb junctions. The transmission value in the Arnold MAR theory[36] is a factor 1.86/0.55 smaller[34] and amounts to about 0.15 due to a different definition (This is the transmission value mentioned in Chs. 6 and 7). Henceforth, in this Chapter we will use the more common $R_Q/R_N$ definition of the transmission. Clearly an $I$-$V$ characteristic of a pinhole with a transmission probability between 0.05 and 0.6, may still look somewhat similar to that of an SIS junction[36]. For this reason, it was not until the application of MAR theory[36,37] to model the characteristics of high current density junctions[29,28] that the effect of MAR on the $I$-$V$ char-
acteristics of SIS junctions was fully appreciated. The MAR process provides an upper limit for the subgap current ratio, since the subgap current increases with increasing current density until all of the current is carried by pinholes and the $I$-$V$ curve is independent of further increase of current density. This is supported by the observation that the ratio of subgap and normal resistance ($R_{2mV}/R_N$) saturates to a value of about 1.5 (corresponding to $I_{2mV}/I_{thermal} = 30$) for $J_c$ above roughly 25 kA/cm$^2$. From Fig. 2.9 it can be seen that this $R_{2mV}/R_N$ value corresponds to a transmission probability of approximately 0.6, in accordance with the value derived from the subharmonic current steps.

### Subgap current summary

In the previous sections we have shown that the subgap current increase with increasing current densities is due to the occurrence of a significant amount of pinholes in the barrier, irrespective of the barrier structure. The harmful effect of pinholes on the performance as heterodyne mixers of Nb as well as NbN junctions will be explained in Chs. 5, 6, and 7.

### 2.3 Influence of MAR on THz radiation mixing

#### 2.3.1 Effect of microwave radiation on MAR

The interaction of microwave radiation with the multiple Andreev reflection process has received little attention over the past 25 years. Around 1973, two experimental papers appeared\[38,39\] which deal with this subject. In both it was observed that higher order subgap structures decrease more rapidly with increasing power than lower order structures. An explanation of this effect may be that the more often a quasiparticle travels back and forth, the more often it encounters a modulated density of states in one of the superconducting electrodes. This effect is modeled by an \textit{ad hoc} modification of Tien-Gordon theory\[40\] taking into account the order of the MAR process responsible for the subgap structures studied\[39\]. This modification is implemented as follows. The magnitude of single-particle tunneling features like the current step at the gap voltage $2\Delta/e$ is decreased by $J_0(\alpha)$ in which $J_0(\alpha)$ is a Bessel function of the first kind and $\alpha$ is the normalized radiation power. With increasing power the original features at $V_0$ are imaged as "satellite structures" at voltages $V_0 \pm n\hbar\omega$ ($n = 1, 2, 3, \ldots$). The magnitude of
2.3 Influence of MAR on THz radiation mixing

these satellite structures according to the Tien-Gordon theory is given by $J_n(\alpha)$. However, the MAR-induced subgap structures behave quite differently upon irradiation. The decrease in magnitude of a structure at voltage $eV = 2\Delta/m$, involving $m - 1$ Andreev reflections is quite accurately given by $J_0(m\alpha)$, which hints in the direction of multiply encounters of a modified density of states upon traversals of a quasiparticle.

A recent theoretical paper by Gunsenheimer and Zaikin[41] concludes roughly the same: Higher order processes are more severely affected by applied microwave radiation. However their theory assumes a non-adiabatic process: The energy of the quasiparticle in the constriction itself is modified by the radiation, rather than the density of states of the superconducting electrodes as in the adiabatic Tien-Gordon theory. Since emphasis is put on the physical processes in the constriction rather than on a quantitative comparison of the experimental observations, the validity of their non-adiabatical approach is left unconfirmed.

2.3.2 Radiation modifies MAR modifies mixing properties

In Ch. 6 the influence of MAR on the shot noise generated by the junction without applied radiation is well described and is considered to be well understood. From Section 2.3.1 it is clear that the MAR process is changed under the influence of radiation in a way which is not predicted by Tien-Gordon. Unfortunately, the Tucker theory, like the Tien-Gordon theory is based on the assumption that current is carried by a first-order tunneling process. Therefore the conclusion is clear: The performance of an SIS mixer in which MAR occurs is not accurately described by the Tucker theory! Our engineering approach to this situation is the following. The voltage at which the junction is commonly biased upon operation is close to the gap voltage, where only one Andreev reflection takes place. This is a fortunate property, since this means that the electrical characteristics are hardly influenced (as explicitly stated in Ref. [41] and experimentally observed in Ref. [39]). Therefore we can still we trust the Tucker theory to be valid in the voltage range of interest.
Chapter 2  Materials aspects and SIS characteristics

2.4  Integrated tuning structures

Soon after the outstanding performance of the SIS detector was appreciated, the devices were used at higher frequencies than the original 33 to 115 GHz. This however posed severe demands on the fabrication of ever smaller mixer blocks with backshort- and E-plane tuners to compensate for the intrinsic capacitance of the SIS junction. The invention of on-chip tuning by means of integrated striplines has appeared the key to further enhancement of the bandwidth, and operating frequency. Because the integrated circuit is made in the same process as the SIS junction, optical lithography is used to define the dimensions. A backshort is still used, although this takes skilled and dedicated manufacturing. A major advantage of this approach is that the stripline is made of superconducting niobium, which losslessly bridges the impedance gap between the probes and the junction. With integrated tuning structures near - quantum limited performance has been obtained at frequencies up to 700 GHz, the gap frequency of Nb of which material both the stripline and the junction are made.

The junction itself performs well at frequencies beyond 700 GHz, whereas the stripline suffers from severe losses since the Nb strongly absorbs the incoming radiation since Cooper pairs are broken up and the RF losses are comparable to those of Nb in normal state. Unfortunately low-\( T_c \) superconductors with a \( T_c \) above 6 K have a normal state DC conductance which is an order of magnitude worse than that of normal metals like gold, copper of even aluminum, which is a normal metal at the SIS operating temperature of 4.2 K, since the \( T_c \) of Al is 1.2 K. Therefore at higher frequencies a metal is used as low-loss stripline material. Although good conductors at low-frequencies, well designed normal metal stripline circuits have shown to absorb 70 % of the incoming radiation.

Current research on the utilization of SIS devices at THz frequencies is concentrated on the design of low - loss tuning structures of aluminum and the use of alternative superconductors as lossless stripline material to be used up to 1200 GHz, the upper frequency limit of the junction itself.

2.4.1  Transmission line operation

The tuning circuit used, a stripline, belongs to the class of transmission lines, which are intended to transport RF signals from an output port of device 1 to
the input port of device 2, possibly with dissimilar input - or output impedances. The mode of operation is similar to an anti-reflection coating or a laser cavity: an incoming wave is partially reflected at the input, the transmitted part travels toward the end of the line where the junction is located where again a part is reflected, which, when arriving at the input, interferes with the previously reflected wave. Depending on the impedance and length of the stripline, the reflected waves can destructively or constructively interfere. When the interference is destructive, the reflection is zero and all radiation is transmitted to the junction, apart from the portion which is dissipated in the stripline during traversal of the wave. A stripline consists of a narrow strip, separated from a wide ground plane by a dielectric material. In our case the strip and ground plane are Al layers, SiO$_2$ functions as a dielectric. (See Fig. 1.8).

The impedance of the line mainly depends on the width $w$ of the strip, the SiO$_2$ thickness $d$ and the distance over which the radiation penetrates into the stripline material. This distance is called the skindepth $\sigma$ in the case of normal metals and penetration depth $\lambda$ for a superconductor. The propagation constant is given by

$$\gamma = \sqrt{Z Y} = \alpha + j\beta$$

(2.12)

The characteristic impedance of the line is given by

$$Z_0 = \sqrt{Z/Y}$$

(2.13)

where $Z$ and $Y$ are the series impedance and shunt admittance per unit length of the line and $w$ and $d$ are the stripline width and thickness of the dielectric respectively. The wavelength in the line is $\lambda = 2\pi/\text{Im}\left(\gamma\right)$. The real part of $\gamma$ ($\alpha$) is the attenuation coefficient in Nepers per meter. The shunt admittance $Y$ per unit length is given by

$$Y = j\omega \left(\frac{w}{d}\right) \epsilon_0 \epsilon_r$$

(2.14)

where $\epsilon_r$ is the relative dielectric constant of the dielectric and $\epsilon_0$ is the permittivity of free space. The series impedance $Z$ is given by:

$$Z = j\omega \mu_0 \frac{d}{w} + \frac{Z_{s1} + Z_{s2}}{w}$$

(2.15)

where $Z_{s1}$, $Z_{s2}$ are the surface impedances of the strip and the ground plane and $\mu_0$ is the permeability of free space. The first term of $Z$ covers the geometrical
The inductance of the line, whereas the second part is determined by the surface currents of the microstrip and therefore by the material parameters. The effect of the surface impedance on the stripline behavior is discussed in Section 2.5.

The impedance seen at distance $l$ from a certain load $Z_L$ at the end of the stripline is calculated from

$$Z_l = Z_0 \left( \frac{Z_L \cosh(\gamma l) + Z_0 \sinh(\gamma l)}{Z_0 \cosh(\gamma l) + Z_L \sinh(\gamma l)} \right)$$

Eq. 2.16 describes the actual impedance transformation. The operation is illustrated in Fig. 2.10 in which the impedance seen at a distance $l$ from a junction situated at the end of the stripline is calculated for a frequency of 1 THz. The junction is assumed to have a $J_c$ of 10 kA/cm$^2$ and an area of 1 $\mu$m$^2$, corresponding to a resistance $R_N$ of 22 $\Omega$ and a capacitance $C$ of 85 fF[44]. The stripline is assumed to have a width of 10 $\mu$m and a SiO$_2$ thickness of 250 nm. Close to the junction the impedance is equal to the junction impedance, since the length $l$ is negligible compared to the wavelength in the stripline. At the end of the stripline the real as well as the imaginary part of the impedance steeply rises, until close to a distance of 46 $\mu$m, where the imaginary part suddenly drops to zero, resulting in
2.4 Integrated tuning structures

2.4.2 Low-loss radiation coupling

For frequencies below 700 GHz, this "endloaded structure" is quite convenient since with appropriately chosen line width the real as well as the imaginary part of the junction impedance is transformed into the probe impedance. However the very sharp change in the imaginary part close to the end of the stripline induces large RF currents and hence excessive losses in normal conducting materials. Fortunately, this build up of the imaginary part of the impedance can be used to benefit the stripline operation with a double junction circuit[48,49]. Fig. 2.11a is a magnified image of the impedance change with distance for a stripline with 3 μm width and a SiO$_2$ thickness of 350 nm. For clarity the imaginary part of the junction impedance is also drawn. At 4.8 μm distance of the junction, the real impedance is the complex conjugate of the junction impedance. Therefore if at this position a similar junction is placed, their impedances tune out and a real resistance $R_N/2$ remains to be matched to the real probe impedance. This can be done without sharp impedance changes, as Fig. 2.11b shows. With this double junction design the lowest noise temperatures are obtained at 1 THz[24].

**Figure 2.11.** a. The real and imaginary part of the impedance seen at distance $l$ from the junction. At a distance of 4.8 μm the imaginary part is the complex conjugate of that of the junction. b. The impedance at distance $l$ from the junction nearest to the probes. The impedance varies smoothly from 0 to 50 Ω.

a real impedance close to the probe impedance of 50 Ω real.
2.5 Surface resistance of metals and superconductors

Apart from the design, the efficiency of a stripline is determined by RF losses in the materials comprising the line. For the geometries and dimensions of the striplines used for submm radiation coupling radiation losses and dielectric losses can be ignored, leaving only the conductor losses. These losses are denoted by the surface impedance $Z_s$ which influences the stripline characteristics according to Eq. 2.15. $Z_s$ is defined as the electric field $E_x$ applied in the $y$-direction, parallel to the surface of the material divided by the total current induced:

$$Z_s = \frac{E_x(\omega)}{\int_0^t J_y(x, \omega)dx}$$  \hspace{1cm} (2.17)

in which $J_y(x, \omega)$ is the current density in the $y$ direction at distance $x$ from the surface. The thickness of the layer is denoted by $t$. Depending on the frequency of the electric field, three distinct regimes can be discerned: the local limit, the anomalous limit and the extreme anomalous limit[50].

The surface resistance is in the local limit when the mean free path $l_e$ is smaller than the skin depth $\sigma$. The skin depth is the characteristic decay length of the electric field in a normal metal. It is given by $\delta = 2/\sqrt{\omega\mu_0\sigma}$ in which $\sigma$ is the conductance of the metal. For a superconductor, the Pippard coherence length $\xi = 1/(\xi_0^{-1} + l_e^{-1})$ has to be compared to the penetration depth $\lambda$ which is frequency independent. The conductivity $\sigma$ of a superconductor on the other hand does depend on the frequency and is given by the Mattis-Bardeen equations[46]. What happens in the local limit is that the electric field varies slow enough for the electron to scatter a number of times before the electric field is reversed. This means that the local relation $J = \sigma E$ holds, hence the name local limit. The surface impedance in the local limit is given by

$$Z_s(\omega) = \sqrt{\frac{i\omega\mu_0}{\sigma}} \coth(\sqrt{i\omega\mu_0\sigma t})$$  \hspace{1cm} (2.18)

At higher frequencies, the time between elastic collisions and the period of the oscillations becomes comparable, and the local limit is no longer valid.
This occurs when \( l_e \approx \delta \) for normal metals and \( \xi \approx \lambda \). The material is then said to be in the anomalous limit and the calculation of the surface impedance becomes quite complex\[50,51\].

The extreme anomalous limit is reached when the mean free path and the coherence length are much larger than the skin depth and penetration depth respectively. The electric field now varies so fast that the electron is accelerated in a number of cycles before it scatters, which simplifies the calculation significantly. In this case the surface impedance for a normal metal is given by

\[
Z_s = \frac{1}{3^{1/2} \pi^{1/3} (1 + \sqrt{3}i) \left( \frac{9 \sigma^2 \mu_0^2 l_e}{16 \sigma} \right)^{1/3}}
\] (2.19)

The remarkable feature of this limit is that the surface resistance does no longer depend on the conductivity \( \sigma \), but rather on the Fermi velocity, since \( l/\sigma \propto v_F^{-2} \)[12]. Therefore once the normal metal fulfills \( l \gg \delta \), further improvement of the surface resistance is impossible.

The surface impedance of a superconductor in the extreme anomalous limit is given by Eq. 2.18 in which the Mattis-Bardeen\[46\] frequency dependent complex conductivity \( \sigma_s = \sigma_1 - j \sigma_2 \) is used instead of the constant DC conductivity \( \sigma_0 \) expected for a normal metal.

To examine in which limit we are experimentally, the typical materials parameters of the metals and superconductors used are listed in Table 2.2 for a frequency of 1 THz.

The \( l_e/\delta \) and \( \xi/\lambda \) ratios indicate that Al is in the anomalous limit, whereas Nb and NbN are in the local limit. The calculated surface resistances are plotted in Fig. 2.12.

### Table 2.2. Material parameters of Nb, Al and NbN at 5 K (all lengths are in nm).

<table>
<thead>
<tr>
<th>Material</th>
<th>( \sigma (\Omega^{-1} m^{-1}) )</th>
<th>( l_e )</th>
<th>( \lambda )</th>
<th>( \xi )</th>
<th>( \delta )</th>
<th>( l_e/\delta )</th>
<th>( \xi/\lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb</td>
<td>( 2.5 \cdot 10^7 )</td>
<td>10</td>
<td>75</td>
<td>100</td>
<td>0.1</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>NbN</td>
<td>( 7 \cdot 10^5 )</td>
<td>0.4</td>
<td>400</td>
<td>600</td>
<td>0.0007</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>( 2 \cdot 10^8 )</td>
<td>125</td>
<td>--</td>
<td>--</td>
<td>36</td>
<td>3.5</td>
<td>--</td>
</tr>
</tbody>
</table>
Figure 2.12. Calculated surface resistances of NbN, Nb, and Al layers with characteristics as tabulated in Table 2.2. The calculated NbN values are lower than those of Nb since the $T/T_c$ ratio is twice as large in NbN. The measured values are obtained from a polycrystalline NbN layer which most likely resembles well the NbN layers used for junction and stripline when fabricated on fused quartz substrates.

2.5.1 Nb surface impedance

To calculate the surface resistance of Nb and NbN we use Eq. 2.18 with the calculated conductivity for the extreme anomalous limit using the Mattis-Bardeen theory. Comparison of measured and calculated stripline losses show that the measured surface impedance for Nb layers follows nicely the calculated Mattis-Bardeen results, at least below the gap frequency [52–54].

2.5.2 Al surface impedance

For aluminum a complete numerical calculation of the surface resistance in the anomalous limit gives values about 40% higher than the extreme anomalous equation yields [55]. Therefore the sensitivity of the receiver can be increased in theory with 40% when the extreme anomalous limit can be reached. Since the DC conductivity is determined by the mean free path, with careful optimization of process conditions the DC conductivity of Al can be improved, until the mean free path is limited by the layer thickness rather than the impurity concen-
tration. This is done by slow evaporation of aluminum in an ultra high vacuum (UHV) environment, yielding smooth layers with low impurity content[12]. Also fast sputtering is known to decrease the impurity concentration, but this usually results in a rougher layer with smaller grains and therefore a larger influence of grain boundaries on the DC or RF conductivity. Since first trials of UHV Al layer employment resulted in degraded junction quality, results in this thesis are obtained with sputtered Al layers.

2.5.3 NbN surface impedance

Above 800 GHz the surface resistance of Al is lower than that of Nb. The calculated surface impedance of NbN is a lot smaller than that of Nb at frequencies below the gap frequency of NbN. However the graph shows that the measured values are two orders of magnitude larger than expected![56] This is the main reason why many groups have chosen to embed the Nb junctions in Al instead of NbN, since clearly the Al outperforms NbN at frequencies beyond 600 GHz (See also Ch. 3) and furthermore Al is a lot easier to include in the fabrication. The reason for the poor RF conductivity in NbN is not completely understood. The effect of metallic regions in the superconductor[57] or increased surface roughness[58] both appeared to produce minor effects only. The surface resistance increase is most likely due to the fact that NbN is a polycrystalline material when sputtered onto a substrate which is not lattice matched with NbN. The NbN can be considered as a network of weakly coupled grains with a lateral grain size $a$ and grain thickness equal to the layer thickness[57,59]. Although a supercurrent can flow through the layer[60], the surface resistance of the material is increased since also penetration of the radiation into the grain boundaries is to be taken into account. With this model it is found that for grain sizes $a < 100 \mu m$ the presence of grain boundaries increases the surface impedance for the materials studied in Ref. [59]. This assumption is supported by the observation that NbN grown on heated or lattice matched substrates exhibits larger crystals, and correspondingly lower surface resistances[61]. Moreover, the dependence of the DC resistance can be explained by the presence of grain boundaries as well[62]. Obviously, these observations can provide merely a partial answer, since Nb and Al are polycrystalline as well. Therefore it is suspected that the grain boundary resistance in NbN is higher than that in Nb, Al or even the recently studied NbTiN (See Section 1.6.1), although the origin of this resistance increase is unknown.
2.6 Conclusions

The fabrication of state-of-the-art SIS junctions with high critical current densities and nevertheless low subgap currents is an intricate process, with many conflicting requirements. The result however is a nearly perfect tunnel junction. Nearly perfect, since the subgap current, which is the main source of mixer noise, irrepressibly increases with increasing current density, for Nb as well as NbN junctions.

The large intrinsic capacitance of an SIS junction is compensated for by an impedance transforming circuit, a stripline. This stripline, when made of Nb, works losslessly up to 700 GHz, after which the RF losses are so severe that other materials are required. Aluminum layers provide lower losses, even lower than those in NbN layers, although still a significant part of the incoming radiation is absorbed. NbTiN, recently employed as stripline material for SIS impedance matching, may be a very promising candidate up to 1500 GHz. Further progress requires clever microwave design together with low-loss materials development.

References

[48] V.Y. Belitsky, S.W. Jacobsson, L.V. Filippenko, C. Holmstedt, V.P.


