Digital Object Identifier:
10.1016/j.physrep.2022.06.005

Source:
https://www.sciencedirect.com/science/article/abs/pii/S0370157322002502?via%3Dihub

Low-Frequency Noise in Downscaled Silicon Transistors: Trends, Theory and Practice

O. Marinov[1], M. Jamal Deen[1] and Juan A. Jiménez-Tejada[2]

[1]Electrical & Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (jamal@mcmaster.ca)

[2]Departamento de Electrónica y Tecnología de Computadores, Universidad de Granada, 18071 Granada, Spain

Abstract

By the continuing downscaling of sub-micron transistors in the range of few to one deca-nanometers, we focus on the increasing relative level of the low-frequency noise in these devices. Large amount of published data and models are reviewed and summarized, in order to capture the state-of-the-art, and to observe that the 1/area scaling of low-frequency noise holds even for carbon nanotube devices, but the noise becomes too large in order to have fully deterministic devices with area less than 10nm×10nm. The low-frequency noise models are discussed from the point of view that the noise can be both intrinsic and coupled to the charge transport in the devices, which provided a coherent picture, and more interestingly, showed that the models converge each to other, despite the many issues that one can find for the physical origin of each model. Several derivations are made to explain crossovers in noise spectra, variable random telegraph amplitudes, duality between energy and distance of charge traps, behaviors and trends for figures of merit by device downscaling, practical constraints for micropower amplifiers and dependence of phase noise on the harmonics in the oscillation signal, uncertainty and techniques of averaging by noise characterization. We have also shown how the unavoidable statistical variations by fabrication is embedded in the devices as a spatial "frozen noise", which also follows 1/area scaling law and limits the production yield, from one side, and from other side, the "frozen noise" contributes generically to temporal 1/f noise by randomly probing the embedded variations during device operation, owing to the purely statistical accumulation of variance that follows from cause-consequence principle, and irrespectively of the actual physical process. The accumulation of variance is known as statistics of "innovation variance", which explains the nearly log-normal distributions in the values for low-frequency noise parameters gathered from different devices, bias and other conditions, thus, the origin of geometric averaging in low-frequency noise characterizations. At present, the many models generally coincide each with other, and what makes the difference, are the values, which, however, scatter prominently in nanodevices. Perhaps, one should make some changes in the approach to the low-frequency noise in electronic devices, to emphasize the "statistics behind the numbers", because the general physical assumptions in each model always fail at some point by the device downscaling, but irrespectively of that, the statistics works, since the low-frequency noise scales consistently with the 1/area law.

Contents

I. Introduction

The above aphorism is the famous summary of the ancient Greek philosopher Plato, said approximately 2500 years ago, regarding the thoughts of Heraclitus (the Ephesus) [1]. Being not loaded with the many interpretations of this aphorism, we can simply rephrase that everything varies, including the random variation, which one usually calls noise; it just needs time this to happen. Later, in section VIII, we will show that this can be origin of 1/f noise – the most difficult for physical interpretation noise in the nature, which always "snakes out" when attempting to describe it absolutely in finite values, but at infinite limits both in time and frequency.

The low-frequency noise (LFN) is always present in electronic devices. However, obviously, the LFN is not "appreciated" due to the fact that it is assumed as undesirable effect, and small enough not to bother much, as compared to other more important and definitely physically better sound effects and useful for the practice properties of the electronic devices, such as gain, high frequency of operation, versatility in making of functions, etc. We can cite again the "present-day assessment" from [2] made in 1981, by re-quoting the Mac Donald's text in "Noise and Fluctuations" from 1962 that

"It is probably fair comment to say that to many physicists the subject of fluctuations (or "noise" to put it bluntly) appears rather esoteric and, perhaps, pointless; spontaneous fluctuations seem nothing, but an unwanted evil, which only an unwise experimenter would encounter!"

However, the low-frequency noise became prominently large, especially in small devices, and it "snakes in" in many cases as a limiting factor for applications, such as high resolution sensors, precise and stable oscillators and other; and considerable interest is given to the "slow" (as compared to the operating frequency of the devices) noise, with either 1/f power spectrum density (PSD) in frequency domain, or with bistable random telegraph signal (RTS) behavior in time. Therefore, in this work we focus on the achievements related to low-frequency noise in electronic devices, mainly in transistors, in order to identify issues related to low-frequency noise in the aggressive device downscaling nowadays, and also to attempt giving an outlook for the evolution of the issues in the near future.

Before we approach to the discussions in this work, we first briefly introduce the types of noise in respect to their spectrum. The main types of electronic noise are summarized in Table 1 in terms of current noise PSD. The thermal noise is due to random motion of charge carriers driven thermodynamically in the devices by uncorrelated scattering, and, therefore, it has uniform, or "white" spectrum, in analogy with the spectrum of the white light. The shot noise originates from the fact that the minimum charge of the carriers is the charge of the electron, $q \approx 1.6 \times 10^{-19}$C, and when many of these discrete charges overcome randomly an emission barrier and traverse the device quickly, then there are current "shots", each of short time, nearly Dirac pulses, and accordingly, each of which having uniform spectrum. Since the "shots" are randomly occurring and uncorrelated, then the spectrum of the shot noise is also uniform "white" spectrum.

The white noise, either thermal or shot noise, or both, is broadband, and it occurs from low frequencies to the maximum frequency at which the device can operate, by the assumption that the device is ideally uniform, and there is nothing else, except for charge carrier motion. In the real devices, however, there can be many other random processes, such as generation-recombination of carriers, charge trapping, phonon scattering, etc, normally with lower "speed" or "repetition rate", which, therefore, cause increase in noise spectra at the low-

frequency end. If the random process has a characteristic time constant τ, or equivalently, a rate $1/\tau$, then the process is random at time scale $t>\tau$, or equivalently, at frequency $f<1/(2\pi\tau)$, and the noise spectrum is uniform for these low frequencies. In contrary, at short time scales $t<\tau$, or equivalently, at frequency $f>1/(2\pi\tau)$, the variability of the noise signal is less, e.g. an RTS "spends" some time in "on" or "off" state before doing a transition to the other state. Consequently, the power spectrum density decays as $1/f^2$ at high frequencies $f>1/(2\pi\tau)$, and the overall spectrum of random process with characteristic time constant τ is the Lorentzian spectrum, as depicted in Table 1. The Lorentzian spectrum is found to originate usually to bistable processes, such as generation-recombination, but more precisely, the requirement for this spectrum is exponential decay in the autocorrelation function of the random process x(t), that is, $\int x(t)x(t\pm\Delta t)dt\propto\exp(-\Delta t/\tau)$, thus, the characteristic time constant τ is the correlation time in x(t).

In the last row of Table 1, the so-called "flicker" noise is given. The power spectrum density of the flicker noise is inversely proportional to the frequency in the entire frequency range, with a slope normally close 1/f, and again by analogy to the light spectrum, it is also called "pink" noise. Several concepts for the origin of the flicker noise are suggested in the literature. One of the most popular concept is superposition of Lorentzian processes with time constants distributed as $1/\tau$, by assumption of particular distributions of traps in the device structure. Another approach is to "stretch" the above exponential function for the autocorrelation, say $\tau=\tau_o+\Delta\tau$, where $\Delta\tau$ is distributed in some way, e.g. normally. Other approaches are to define ½ differential operation, random rate perturbations that decay as square root of time, etc. Overall, a unique explanation for the 1/f noise is not available at present (and most probably, it will be never available after the many suggestions made so far), although the 1/f spectrum occurs in virtually any system, from electronics, through the level or river Nile, to biology, music and finances, and there are many reasonable models and consistent explanations for 1/f noise in particular cases.

From above, the low-frequency noise in small devices is relatively increased, and explanations and models that predict this noise are available. Therefore, we review the models and their predictions extensively in this work, both numerically and by keeping the link to the physical assumptions behind the models. In order to derive a common point, we have also intentionally suppressed the controversy, which has accompanied the subject of low-frequency noise for the origin of 1/f noise, because we have observed that the predictions of the different models coincide, when the word is for numerical values and behaviors in respect to bias, temperature and device sizes; and all of the popular low-frequency noise models are somewhat mesoscopic, or better to say compact models, with the microscopic effects averaged after one or several mathematical integrations, which does not really allow to fully and precisely inspect the assumed microscopic physical origin of the fluctuations from the scattered data obtained after measurement of the noise. Indeed, after the analyses in section VIII, one may not always need to mandatory assume microscopic origin for the low-frequency noise.

To approach to the review, we first provide in section II details and the generic models related to intrinsic and coupled noise in the forms, in which the low-frequency noise is lumped in compact models, that are widely used at present. Then, in section III, we review the state of the art low-frequency noise in bipolar junction transistors (BJTs), analyzing the different factors related to the issues with low-frequency noise, such as crossover between bulk, surface and barrier noise, fabrication, superposition of Lorentzian noise, and decomposition to individual noise components in small-area BJT, problems with averaging techniques, and also, we have provided some derivations that explain details in the bias behavior of the low-frequency noise in BJT, using the generic models

from the previous section II.

Having the observations and results for BJTs, we have pursued in section IV a detailed review for the level and models of the low-frequency noise in MOS transistors, since these transistors are at the frontier of the device downscaling nowadays. From these, we found that the different models converge each to other, including for ranges (e.g. of gate oxide thickness and trap densities), at which the assumptions for the models are actually violated. Interestingly, we have observed, that the models for the noise in BJT also converge to the models for MOS transistors at some instances, such as by noise from interfacial oxides. Also, we showed that models based on superposition of distributed traps cannot discriminate clearly whether the noise is due to distribution of energy or depth of the trap in the gate oxide, since both are always together in the models. We have also provided some extensions in gaps found in the literature, such as for unequal amplitudes in RTS noise in nominally identical MOS transistors, relations between figures of merit for low and high frequency noise, and for frequency performance of the MOS transistors.

Since the downscaling of MOS transistors required modifications in the classical MOS structure (one gate MOS with uniformly doped silicon body), as well as, germanium is used widely at present to improve the mobility in npn BJTs, the low-frequency noise in the modified SiGe heterostructural bipolar transistors (HBTs) and MOS transistors with multiple gates is addressed in section V. Overall, the observation is that the low-frequency noise increases, when mixing materials or adding new layers in the transistor structures, which is different from claims made last decade, and thus, it is a potential issue for the future. Nevertheless, forward body biasing and multiple gates, seem, improve the low-frequency noise performance of MOS transistors, but cannot compensate for the increase of the scattering in the noise levels in advanced silicon based transistors, as discussed in section VI, in which it is shown that the ultimate downscaling, e.g. in carbon nanotube devices, leads to stochastic behavior of the device, although, surprisingly, the data for these device still match with the general trend (1/area).

The consequences of the increased low-frequency noise for the practice are discussed in section VII for two cases – the tradeoff between low noise and low power in low-power amplifiers and for phase noise in oscillators, providing some derivations that help to quickly estimate the noise performance of the amplifier from its consumption, and the phase noise of the oscillator from the harmonics and symmetry of the generated RF signal.

Finally, in section VIII, we provide extrapolation of the results gathered during almost one century, to identify that the low-frequency noise in nanodevices can impact the reliability even of digital circuits. Analyses in this section showed that the variations during fabrication, called as “frozen noise”, also contribute to the temporal noise in the devices; that the values for noise parameters are corresponding to the range of the Heisenberg uncertainty for the free electrons in semiconductors; and that the purely statistical accumulation of variance with time can be a mechanism generically creating 1/f noise in the nature. This accumulation mechanism is known as “innovation variance”, and it seems is overlooked for electronic devices as potential source of background 1/f noise, while it provides a reason, which explains why the geometric averaging of scattered noise data should be preferable.

Summarizing our review and analyses in section IX, we conclude that the low-frequency noise in electronic devices follows consistently the phenomenological law (1/area) in relative units for the noise power in ratio to DC power, which basically sets a barrier for downscaling of deterministically behaving devices at sizes below about 10×10nm², which is nothing, but the size range of the viruses – the smallest structures, in which the nature was able to embed reproducibility and functionality over a period in the range of about 10^9 years. The different

models for low-frequency noise appear to coincide in their predictions for noise magnitudes, which actually implies that the statistics is more important for the noise than the physical phenomena through which it causes noise in the devices. Thus, the numbers might be more important and a change in the coordinate system for noise perhaps will be helpful to describe the low-frequency noise even better than the descriptions are at present, without arguing for and against that strongly, as it happened in the near past whether the 1/f noise is due to mobility or number fluctuation. It is due to both, and in addition, perhaps due to other fluctuations, which stay hidden from us at present. It might be reasonable to state that we can observe only a "microscopic" window, if not smaller, by probing the variations as 1/f noise in electronic devices, which accumulate variations from the continuously changing matter. Therefore, we have the aphorism above, behind which is the unity of continuity and variation, and with the information available to us, let our discussion begin "flowing" with the first topic on intrinsic and coupled noise in device currents, for which, it seems, we have agreement in principle at present.

II. Intrinsic (uncorrelated, uncoupled) and coupled (correlated) behavior in a fluctuating system (behind and beyond $\Delta\mu - \Delta n$ controversy)

In this work, we use a conceptual approach of "intrinsic and coupled noise" that helps to identify the noise in electronic devices. The approach is based on the general assumption that the fluctuations in one quantity can originate intrinsically from its nature, or the fluctuations can be induced, thus, coupled from fluctuations of another quantity.

II.1. Coupled behavior

For the noise S_I in a quantity I, when S_I is coupled to another fluctuation S_V, one has

$$\frac{S_I}{{I_{DC}}^2} = K\left(\frac{g}{I_{DC}}\right)^2 S_V\,, \tag{1}$$

where K is a parameter (usually taken K=1, if no suppression or enhancement in S_I due to correlation to other process exists), I_{DC} is the average value of I, and $g=\partial I/\partial V$ is the coupling coefficient between I and V, by assuming also that I and V are immediately, instantly and fully correlated each to other. Obviously for an electronic device (e.g. resistor, diode or transistor), I, V and g are electrical current, voltage and conductance, respectively. So, the units for the power spectrum densities (PSDs) become A^2/Hz for S_I and V^2/Hz for S_V.

We illustrate eq. (1) for the voltage noise in bipolar junction transistors (BJTs) and MOS transistors, using the data from a past report of the International Technology Roadmap for Semiconductors (ITRS 2006) [3], which extended its predictions until 2020. ITRS provides values for the input referred voltage $S_{V@1Hz}$ at 1Hz normalized (multiplied) with device active area A of $1\mu m^2$ for npn BJTs and nMOS transistors, as shown in Figure 1a. Let us express $S_I/{I_{DC}}^2$ with the simplest SPICE model for 1/f noise, and use the transconductance of the transistors g_m in eq. (1), also taking K=1 and multiplying the equation with the device area A. That is

$$A\frac{S_I}{{I_{DC}}^2} = \frac{(A\times K_F)}{f} = \left(\frac{g_m}{I_{DC}}\right)^2 \frac{(A\times S_{V@1Hz})}{f}\,. \tag{2}$$

Here, the parameter K_F is a measure for the ratio noise/DC in the output current of the transistors. The ratio $g_m/I_C \approx 1/\varphi_t \approx 38.5\ V^{-1}$ in BJT, where $\varphi_t = kT/q \approx 0.026V$ is the thermal voltage at room temperature T=300K. The

ratio g_m/I_D varies with the bias of the MOS transistor, but at low gate overdrive of $(V_G-V_T)=0.1V$, g_m/I_D~13 V^{-1}. So, we write

$$(A\times K_F)\approx\begin{cases} f\left(A\times S_{V@f}\right)\Big/\varphi_t{}^2\approx 1Hz\left(A\times S_{V@1Hz}\right)38.5^2, \text{for BJT} \\ 1Hz\left(A\times S_{V@1Hz}\right)13^2, \text{for MOS at } (V_G-V_T)=0.1V \end{cases}. \quad (3)$$

The results for $(A\times K_F)$ calculated by eq. (3) are shown in Figure 1b, when the values for $(A\times S_V)$ from Figure 1a are used. Interestingly, the input referred voltage noise in BJT is about 2 orders of magnitude less than that in MOS transistors, but the difference is smaller in the output current noise; and the difference varies further at higher gate overdrive of MOS transistor, because $g_m/I_D\propto 1/(V_G-V_T)$.

Now, we discuss eq. (1) in more details. Assume that the noise is caused by a fluctuation S_Q of trapping charges, and these charges change the voltage V on a capacitance C. Then, the coupling Q=CV is given by the capacitance C, and

$$\frac{S_I}{I_{DC}{}^2}=K\left(\frac{g}{I_{DC}}\right)^2 S_V=K\left(\frac{g}{I_{DC}}\right)^2\frac{S_Q}{C^2} \quad (4)$$

Consequently, the charge Q=qN is coupled to the number of charges N by the elementary charge of the electron q (1.6×10^{-19} C), and one writes the general form of the equation for the so-called "number fluctuation" (Δn)

$$\frac{S_I}{I_{DC}{}^2}=K\left(\frac{g}{I_{DC}}\right)^2 S_V=K\left(\frac{g}{I_{DC}}\right)^2\frac{S_Q}{C^2}=K\left(\frac{g}{I_{DC}}\right)^2\left(\frac{q}{C}\right)^2 S_N, \quad (5)$$

which is widely used to study the effect of charge trapping in interfaces on the current in a device, e.g. charge trapping in the gate insulator of a MOS transistor and its drain current. This equation is for number fluctuation, since it assumes that the fluctuation S_N in the number of trapped charges causes the noise in the current I. Obviously, the unit is Hz^{-1} for the power spectrum density S_N, and S_I is also PSD in units A^2/Hz.

The different Δn-models for different devices (at particular set of physically based assumptions) derive different expressions for S_N and coupling parameters (C, g, K), which can be physics-, bias-, process- and design-dependent. Overall, all complex derivations based on the assumption for charge trapping end with a form similar to eq. (5). These will be presented along with the discussions on the specific devices. However, one important note should be made. The fluctuation in the number N of trapped charges is assumed in the origin of the noise in Δn-models, and this fluctuation is indirectly transferred to a fluctuation in the number n of charge carriers that provide the current I in the device, by electrostatic (Coulomb) balancing the charge using the Gauss law. So, one can mistakenly assume from the expression for drift current with a density $J=qn\mu E$ that the Δn-model is for the number n of charge carriers, by neglecting the variations in charge carrier mobility μ and in the electric field E, caused by the trapped (and thus immobile) charge. For example, the trapped charge can cause change in the mobility, so that $dg\propto\mu\partial n/\partial N+n\partial\mu/\partial N\neq\mu\partial n/\partial N$ and the estimate for g obtained from electrostatic balance $\partial n/\partial N=1$ and μ=constant will be inaccurate. To resolve these problems, a correlated (to the number of charges) mobility model (Δn-$\Delta\mu$) is used widely in nMOS transistors, and a scattering parameter is introduced.

Furthermore, both (Δn-$\Delta\mu$) and (Δn) models assume that the current flow is continuous and free of inherent fluctuations. Certainly, this is a good approximation when analyzing the average I_{DC}, but it is not exactly true,

since the scattering of the moving charge carriers is random, so the value for the mobility is a constant only on average, while the charge carrier velocity in the direction of the current flow varies (as well as, the current flow is discrete, because each charge carrier has a magnitude of q).

II.2. Intrinsic behavior

To describe these fluctuations in different physical magnitudes for the 1/f noise, one can use the Hooge equation

$$\frac{S_I}{I_{DC}^2} = \frac{S_{Qn}}{Q_n^2} = \frac{S_\mu}{\mu^2} = \ldots = S_{norm} = \frac{S_Z}{Z_{DC}^2} = \frac{\alpha_H}{nf} \tag{6}$$

where the normalized PSD of the noise $S_{norm}=S_Z/Z_{DC}^2$ for any quantity Z, i.e. current I, total carrier charge Q_n, or mobility μ, is the ratio of the power spectrum density S_Z to the square of the average Z_{DC} of this quantity. Here, n is the total number of carriers and α_H is a material dependent parameter, called Hooge parameter, which ranges usually between 10^{-5} and 10^{-3} for semiconductors, and it is approximately 2×10^{-3} in metals.

Note that Hooge eq. (6) does not discriminate between quantities when their averages are in product or ratio, i.e. $Z=z_1\cdot z_2/z_3$. Instead, it provides an estimate α_H for the variance in Z that can be attributed to one mobile particle. At an assumption that the variance (noise) originates to mobility fluctuation in the drift current $J=qn\mu E$, then the Hooge equation is for mobility noise, and if the assumption is not correct (i.e. μ=constant, but n or E varies), then the Hooge equation is not for mobility noise, but the equation is still valid, because it is based on the general principle of statistics that the variance of the average is inversely proportional to the size (average number n) of the population. Thus, care should be exercised when claiming a specific physical process as the origin of the noise, and the claim should not be based solely on the application of the Hooge equation. This point has been explained many times in the literature, for example in [4].

The Hooge equation is widely used to investigate the low-frequency noise (LFN) at assumption that the noise originates from and it is intrinsic for the conducting and semiconducting material in the transistors. For example, polysilicon resistive films are widely used in integrated circuits and the low-frequency noise of the polysilicon resistive films is characterized by the Hooge parameter α_H, e.g., as in [5], assuming that the carrier concentration is equal to the doping of the polysilicon. A second example is the LFN in organic thin-film transistors (OTFTs), which was analyzed in terms of the Hooge noise in [6,7], where the drain DC current $I_D=I_{DC}$ is assumed as a drift current proportional to the number of carriers in the channel, and

$$I_D \cdot L = q \cdot n \cdot \mu \cdot E = q \cdot n \cdot \mu \frac{V_{DS}}{L} \Rightarrow n = \frac{I_D \cdot L^2}{q \cdot \mu \cdot V_{DS}} \tag{7}$$

where V_{DS} is the DC voltage between drain and source of the OTFT, and $E=V_{DS}/L$ is the electric field in the channel of length L due to V_{DS}, assuming also an operation of the OTFT in the linear regime at low V_{DS}. Substituting in the Hooge equation (6), one gets that the normalized intrinsic noise S_{norm} due to the organic semiconductor in the channel of the OTFT is proportional to the drain-source resistance R_{DS} of the OTFT channel, that is

$$S_{norm} = \frac{S_{ID}}{I_D^2} = \frac{\alpha_H}{nf} = \frac{1}{f} \times \frac{\alpha_H \cdot q \cdot \mu \cdot V_{DS}}{I_D \cdot L^2} \propto \frac{V_{DS}}{I_D} = R_{DS} \ , \tag{8}$$

while the absolute magnitude S_{ID} of the PSD of the LFN in the drain current is proportional to DC power

dissipated on the OTFT channel, that is

$$S_{ID} = S_{norm} \cdot I_D{}^2 = \frac{1}{f} \times \frac{\alpha_H \cdot q \cdot \mu}{L^2} \times I_D \cdot V_{DS} \propto I_D \cdot V_{DS} = P_{DC} . \quad (9)$$

Eqs. (8) and (9) indicate the significance of the device resistance and applied DC power for the levels of the LFN, which are handy and were used in [6, 7, 8] to determine dependence between the mobility μ and Hooge parameter α_H in OTFT and CdSe TFT, as well as to inspect a suitable for OTFT figure of merit for the noise, given by [7]

$$\frac{S_{ID}}{(I_D \cdot V_{DS}) / (W \cdot L)} = \frac{\text{noise PSD}}{\text{DC power} / \text{device area}} . \quad (10)$$

This figure of merit includes (and compensates for) the most significant scaling factors of the LFN in electronic devices, in particular, the proportionality to applied DC power and inverse proportionality to device area. The latter is widely discussed, whereas the former is rarely addressed in the literature.

II.3. Combination of coupled and intrinsic fluctuations

By assumption that the coupled and intrinsic fluctuations in the quantity I are uncorrelated, the Hooge equation can be combined with the expressions (1) and (5) for coupled fluctuation, resulting in

$$S_{norm} = \frac{S_I}{I_{DC}{}^2} = K \left(\frac{g}{I_{DC}} \right)^2 S_V + \frac{\alpha_H}{nf} , \quad (11)$$

for the general case of coupled fluctuation, and in

$$S_{norm} = \frac{S_I}{I_{DC}{}^2} = K \left(\frac{g}{I_{DC}} \right)^2 \left(\frac{q}{C} \right)^2 S_N + \frac{\alpha_H}{nf} \quad (12)$$

for the case when the coupled fluctuation is due to fluctuation (Δn) in the number N of trapped charge.

Note again that the two terms for coupled (Δn, let's say) and intrinsic (Hooge) fluctuations in equations (11) and (12) present different sides and origins for the low-frequency noise in the current transport, and they do not contradict each other. Also, the normalized noise S_{norm}, as defined with these equations, is a power spectrum density with a unit Hz^{-1}, and is a convenient figure of merit, which estimates the noise "power" per a bandwidth of 1 Hz in ratio to the DC "power", that can be obtained by multiplying the nominator and denominator of the equations with the resistance of the device.

The above equations have constant parameters only for uniform samples at constant biasing. If the biasing changes, then the parameters also change and S_{norm} does not have a constant value along the sample. For example, in the case of current crowding, the intrinsic 1/f noise of the current cannot be given in its simple form of eq.(6). Rather, S_{norm} will be required to be expressed in an integral form [9], given by

$$S_{norm} = \iiint_{WLT} \frac{\alpha_H}{n(v) \cdot f} J^4 dv \Big/ \left(\iiint_{WLT} J^2 dv \right)^2 \quad (13)$$

where WLT is the volume of the conductive channel, v is integration variable in this volume, and n(v) and J are

the carrier concentration and the current density in the conductive channel, respectively. To the best the authors' knowledge, however, for the LFN, there has been no recent publication on the impact of the current crowding around a single trapped charge in nanometer-scale devices on the low-frequency noise, which should be prominent for single carbon nanotube transistors.

One useful observation can be made in eq.(12) for the 1/f noise in planar devices, such as bipolar junction transistors (BJTs), MOS transistors and thin film transistors (TFTs). In these transistors, one of the dimensions is almost fixed by the vertical layout, e.g. thickness of emitter-base junction and base in BJT, channel depth and oxide capacitance in MOS transistor. Therefore, the active area A plays an important role for the 1/f noise, because the numbers of carriers n, traps N and capacitance C become proportional to A. Assuming that the noise is due to uncorrelated random fluctuations, the "variance" S_N is proportional to N, and one can write from eq.(12) that

$$S_{norm} = \frac{S_I}{{I_{DC}}^2} \propto \left\{ \begin{array}{l} \frac{S_N}{C^2} \propto \frac{A}{A^2 C_A^2}, \text{if coupled noise is dominant} \\ \frac{\alpha_H}{nf} \propto \frac{1}{A\, n_A}, \text{if intrinsic noise is dominant} \end{array} \right\} \propto \frac{1}{A} \qquad (14)$$

where C_A is capacitance per unit area, e.g. oxide capacitance C_{ox} in MOS transistors, and n_A is carrier density per unit area, e.g. $n_A \propto I_E/(qA_E)$ with I_E and A_E being the emitter current and area, respectively, in BJT.

Eq. (14) implies that the low-frequency noise is inversely proportional to the device area, if one noise mechanism prevails in the device, which is usually the practical case, by keeping other conditions, e.g. biasing level, nominally the same. The biasing level in MOS transistors is a function of gate overdrive voltage (V_G-V_T), and it is the current density $J_E=I_E/A_E$ in BJTs. Thus, owing to the 1/area dependence, it is observed that S_{norm} increases in sub-micron area devices.

III. Noise in BJT

The areal dependence with increasing normalized noise in npn BJTs with smaller emitter area A_E is shown in Figure 2, using the SPICE parameter K_F defined from

$$S_{norm} = \frac{S_{I_B}}{{I_B}^2} \approx \frac{S_{I_C}}{{I_C}^2} = \frac{K_F}{f} \propto \frac{1}{A_E} \qquad (15)$$

The data for more than 150 devices are collected from [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27], and shown in Figure 2 with some points omitted for clarity. Since the numbers spread over several decades, we use geometric averaging technique [28] to extract the trend in the data from the product $A_E \times K_F$; and the trend is $A_E \times K_F \approx 5.6 \times 10^{-9}$ μm^2 with logarithmic standard deviation of σ_{dB}=3.4dB. These values are practically the same as $A_E \times K_F \approx 5.5 \times 10^{-9}$, σ_{dB}=3.7dB from a previous evaluation of the trend [29], when about a half of the data for 70 devices measured before 2005 was used, as well as in the range predicted by ITRS – see again Figure 1b. Also, the distribution of the data is very close to log-normal distribution, as shown in the insert of Figure 2.

Although the trend appears to be stable, 122 of 164 data points are within $\pm 2\sigma_{dB}$ around the average, the large deviations are apparent in Figure 2. We have inspected the details in the publications to get an insight on what causes the scattering in the data. Obviously, we found four factors: differences in measurement setups,

differences in device fabrication, crossover between different noise sources in BJT, and measurement and characterization uncertainty. The analysis of the scattered data in Figure 2 provides about the theory of the noise in BJTs developed in the past. This is discussed below.

III.1. Differences in measurement setups

It is well established that the input voltage noise S_{V_B} in BJT is low, while the input current noise S_{I_B} is high, when compared to other transistors, such as MOS and junction field effect transistors. Therefore, the values for the low-frequency noise depend strongly on the impedance Z_B of the biasing circuit in the base terminal of the BJT, even if the base DC current is unchanged. This is illustrated in Figure 3 from [12]. Consider the noise equivalent circuit for the noise measurement in Figure 3a. The DC voltage source and the impedance Z_B in base biasing circuit are changed simultaneously, so that the base current I_B is kept at a constant value of 1μA or 6μA. Then, the output noise S_{I_C} in the collector current I_C is measured, as shown with symbols in Figure 3b, and accordingly modeled by

$$S_{I_C} = \left[\frac{Z_B + r_b + r_e}{Z_B + z_B}\right]^2 \beta^2 S_{I_B} + {g_m}^2 S_{V_B}, \tag{16}$$

where r_b and r_e are the resistances of the base and emitter passive regions, $z_B=r_b+\varphi_t/I_B+(\beta+1)r_e$ is the input impedance of the BJT, $\beta\approx I_C/I_B$ is the current gain of BJT, $\varphi_t=kT/q\approx 0.026V$ is the thermal voltage at room temperature T=300K, and $g_m\approx I_C/\varphi_t$ is the transconductance of the BJT. Although the DC currents, and thus β and g_m, do not change in the experiment, the output noise S_{I_C} changes with Z_B. In particular, when $Z_B>>z_B$, then the ratio in the brackets of eq.(16) is of value close to one, the term $\beta^2 S_{I_B}$ dominates, and the normalized values for the 1/f noise in the collector and base currents become equal, because

$$S_{norm,C} = \frac{S_{I_C}}{{I_C}^2} \approx \frac{\beta^2 S_{I_B}}{(\beta I_B)^2} = \frac{S_{I_B}}{{I_B}^2} = S_{norm,B} = \frac{K_F}{f}, \text{ if } Z_B>>z_B. \tag{17}$$

Thus, a high impedance, $Z_B>>z_B$, is required in the base biasing circuit in order to measure the 1/f noise coefficient K_F correctly. Earlier publications, e.g., [30] explicitly state the condition $Z_B>>z_B$ for noise measurement, but later, some experimental setups violate the condition, especially when combining DC, low frequency and RF measurements together. Therefore, we discuss also for the cases, when $Z_B\leq z_B$.

If $Z_B<<z_B$, then the ratio in the brackets of eq.(16) is approximately

$$\frac{Z_B + r_b + r_e}{Z_B + z_B} \approx \frac{Z_B + r_b + r_e}{\varphi_t/I_B + (\beta+1)r_e} \approx \frac{Z_B + r_b + r_e}{\varphi_t/I_B + r_e I_C/I_B} \leq \frac{g_m}{\beta}(Z_B + r_b + r_e), \tag{18}$$

the output noise S_{I_C} is given by

$$S_{I_C} \leq {g_m}^2 (Z_B + r_b + r_e)^2 S_{I_B} + {g_m}^2 S_{V_B}, \text{ (if } Z_B<<z_B), \tag{19}$$

and compared to the case $Z_B>>z_B$, S_{I_C} becomes much lower, as one can see from Figure 3b. In fact, since the equivalent circuit in Figure 3a is taken at hoc, one can speculate that the physical origin of current and voltage noise in BJT is the same, assuming that $S_{V_B}=(r_b+r_e)^2 S_{I_B}$, and eq. (16) can be rewritten with the term for S_{V_B} omitted, as

$$S_{I_C} = \left(\frac{Z_B + r_b + r_e}{Z_B + z_B} \right)^2 \beta^2 S_{I_B} = \begin{cases} \beta^2 S_{I_B}, \text{if } Z_B >> z_b \approx \varphi_t / I_B + (\beta+1) r_e \\ g_m{}^2 (Z_B + r_b + r_e)^2 S_{I_B} \approx \left[\frac{I_B (Z_B + r_b + r_e)}{\varphi_t} \right]^2 \beta^2 S_{I_B}, \text{if } Z_B << z_b \end{cases} . \quad (20)$$

The speculation is quite reasonable for the modern polysilicon emitter BJT, because g_m is desired high, while r_b and r_e scale inversely with the emitter area, and r_e is affected by the interfacial oxide (IFO) between polysilicon and monosilicon layers in the emitter region. The impact of IFO on the low-frequency noise will be discussed shortly in sub-section III.2 "**Differences in BJT fabrication, IFO**".

Let us estimate the reduction of S_{I_C} by low impedance biasing of the base terminal, using eq. (20). The base current is usually less than or in the range of 10μA. The values for r_b and r_e are usually maintained less than 20-50 Ω and let Z_B=50Ω. So, the product $I_B(Z_B+r_b+r_e)$~1mV, then $I_B(Z_B+r_b+r_e)/\varphi_t$~5% and the number in the square brackets of eq. (20) becomes in the range of 1/300 (when squared). In other words, S_{I_C} is reduced about 2½ decades by low impedance biasing of the base terminal, as compared to the case of high impedance biasing. This is clearly shown in Figure 3b.

Now, let us look again at the solid triangles denoted as "Low Z_B" in Figure 2. The values for K_F are about 1-2 decades below the trend. The values were estimated from the normalized output noise $S_{I_C}/I_C{}^2$, assuming $S_{I_B}/I_B{}^2 \approx S_{I_C}/I_C{}^2$ as at a high impedance biasing of the base of the BJTs, since data for S_{I_B} are not reported in [22], and not much details are given for the measurement setup. The objective of this publication is RF circuits; and we have assumed that the noise performance was measured at low impedance conditions for biasing of the BJT base, which is consistent with the objective, but from the discussion above, the low impedance bias of the base provides uncertain values for the 1/f noise coefficient K_F in BJT. Therefore, we refrain from using these data in the evaluation of the trend for K_F in BJT. We have included the data points in the figure only to note that the measurement conditions have to be carefully inspected prior using published data in aggregated analyses and comparisons; otherwise, the numbers may mislead.

III.2. Differences in BJT fabrication, IFO

The second factor for deviation of noise level around the trend in Figure 2 can be attributed to differences in the fabrication of BJT. The impact of several technology steps on the low-frequency noise in polysilicon emitter BJTs was reviewed several times, for example in [31], and extensively analyzed in the past [18, 19, 20, 21, 32, 33, 34, 35, 36, 37, 38, 39, 40]. From these studies, it is found that among the many sources that can contribute to the low-frequency noise, such as fluctuation in diffusion, surface recombination and charge trapping, the major noise source is associated with or located in the interfacial oxide (IFO) between poly and monosilicon layers in the emitter of BJT. The noise in polysilicon emitter BJT is attributed to IFO, because the thickness t_{IFO} (~0.3-0.8nm) of IFO strongly affects the level of the low-frequency noise. Therefore, we explicitly show and label with "var.IFO" in Figure 2 the data obtained from samples with different t_{IFO}. The samples usually are of the same emitter area, and we added eye-guide dash lines in Figure 2. Observe that a small increase of t_{IFO} with about 0.5nm results in a large increase in the level of the low-frequency noise, approximately 2 decades for the values of the noise coefficient K_F. A better view of the data in npn BJTs is given in Figure 4, together with data for other types of BJT from [32, 33, 34], and the data are analyzed in a manner similar to previous discussion in [29]. The data are available in numerical form in [41].

It is observed that the noise from IFO results into quadratic dependences $S_{I_B} \propto I_B^2$ and $S_{I_C} \propto I_C^2$, and it cannot be described as an intrinsic (Hooge) noise caused by the fluctuation in the diffusion process in BJT [10, 42] – see again eqs. (6) and (11). Therefore, it is concluded that the noise in the base and collector currents, S_{I_B} and S_{I_C}, respectively, is coupled to fluctuation of charge or conductance of IFO, as given with the left-hand term in eq. (7) or (8), and $S_{I_C} \propto S_{I_B} \propto I_B^2$. Several models for the coupled noise from IFO are reviewed in [31] and summarized in [29]. These models (described below) predict different dependences of K_F on the thickness t_{IFO} of the interfacial oxide IFO.

(i) The direct tunneling model [10] is based on thermally driven (Nyquist) random modulation of the tunneling barrier height of IFO, which in turn modulates the transparency of IFO and the current flow in the stack of polysilicon, IFO and monosilicon layers in BJT. According to direct tunneling model, it is shown in [31] that the normalized current noise increases as $K_F \propto S_{I_B}/I_B^2 \propto t_{IFO}^3$. The trend in Figure 4 confirms this cubic dependence, and the prefactor in this dependence is

$$\frac{A \times K_F}{t_{IFO}^3} \approx 38\,\mu m^{-1}, \sigma_{dB} \approx 3.4 dB \text{, for the overall dependence of the noise on } t_{IFO}. \quad (21)$$

(ii) The two-step tunneling model, as explained in [31], is derived from low-frequency noise in tunnel MOS diodes [43]. In the first step, there is a recombination of carriers from the semiconductor bands into interface states at the SiO_2 interfacial layer next to the monocrystalline silicon by the Shockley–Read–Hall process. In the second step, there is elastic tunneling of carriers from these interface states into bound or slow states in the oxide that are located close to the interface. The resulting dependence for the normalized current noise in BJT is quadratic [31], and

$$A \times K_F \propto t_{IFO}^2 \text{, is observed in several separate data series,} \quad (22)$$

as illustrated with dash-lines in Figure 4.

(iii) An exponential function, e.g. $\exp(t_{IFO}/\lambda)$, proposed in [14] empirically, can also fit some data. This is illustrated with another dash-line which fits the triangles in Figure 4 for the noise in SiGe HBTs [33, 34]. The values for λ, however, scatter; $\lambda \approx 0.5$nm for the data from [14] (open diamonds in Figure 4), while $\lambda \approx 0.4$-0.5nm for the data from [33, 34] (solid triangles in Figure 4), reaching a value, which is difficult to explain physically, even for the direct tunneling of holes. The effective mass of holes is high in SiO_2 and therefore have shorter tunneling attenuation distance $\lambda < 0.1$nm. The conservative lowest value for λ from Wentzel-Kramer-Brillouin (WKB) approximation is

$$\lambda = \frac{h}{4\pi\sqrt{2 m_o m^* \Phi}} > 0.059 nm, \quad (23)$$

where $h = 6.63 \times 10^{-34}$ Js is the Planck constant, $m_o = 9.11 \times 10^{-34}$ kg is electron mass. From [44, 45], we get that $m^* < 0.6$ is the maximum effective mass of charge in SiO_2, e.g. holes in nitrided SiO_2, and $\Phi \leq 4.5 eV = 7.2 \times 10^{-19}$ J is the maximum offset between bands of Si and SiO_2, e.g. between valence bands. Nevertheless, the range of values for t_{IFO} is narrow and the accuracy of these values is low, with an error of 0.1-0.3 nm. Therefore the different theories cannot be distinguished reliably with existing experimental results.

It is interesting to observe in eq. (21) that the prefactor 38 μm^{-1}=1/(26nm) ±3.4dB for the trend in Figure 4 corresponds to a diameter $2\pi r_{th}$, which has been regarded as the effective diameter, in which the electrostatic band bending due to a single excess (trapped) electron charge disturbs the silicon properties [46]. At room temperature, r_{th}~5nm and corresponds to the distance in which the electrostatic band bending due to a single charge is equal kT=26meV. Also, with ε being permittivity, a common term is $q^2/(A_E\varepsilon/t_{IFO})^2$ in the expressions in all models for the noise from IFO, which implies that the charge capture at IFO causes coupled noise in the BJT current, according to eq. (5). The differences between the models come from the different assumptions for the actual physical mechanism, which is behind the charge fluctuation S_N in eq. (5). These models lead to a variety of figures of merit for parameters related to the noise from IFO, e.g. tanδ~0.3-3 for direct tunneling [10]. According to [34], $S_N \propto \lambda N_t/C_{mo}^2$ or $S_N \propto D_{it}/C_{mo}^2$ for two step tunneling and random walk models, respectively, where C_{mo} is surface capacitance of IFO (may differ significantly from $\varepsilon_{SiO_2}/t_{IFO}$), and the figures of merit are $N_t/C_{mo}^2 \sim 5\times10^{29}$ $cmF^{-2}eV^{-1}$ and $D_{it}/C_{mo}^2 \sim 10^{22}$ $cm^2F^{-2}eV^{-1}$. These values for tanδ, N_t/C_{mo}^2 and D_{it}/C_{mo}^2 are solely introduced to fit data from noise measurements, and they are not examined from other experiments, to the best of our knowledge.

(iv) Non-uniform IFO model. There is some uncertainty in the models for the noise from IFO. The models predict power-law dependence between noise level and t_{IFO}, and do not suggest exponential dependence, observed in some of the experiments, especially from the samples with thicker IFO. An approach to implement exponential dependences was taken in [20] for samples of different widths W of the emitter. The results are shown with solid squares in Figure 2 and Figure 4 as function of emitter area and average IFO thickness, respectively. It is suggested in [20] that t_{IFO}=0.55nm in the middle of the emitter for samples with wide emitter, while IFO is Δt_{IFO}=0.25nm thicker at the periphery of the emitter. Then, the IFO thickness along the emitter width is assumed to decrease from $t_{IFO}+\Delta t_{IFO}$ at the edge of the emitter toward t_{IFO} in the middle of the emitter, by an empirically assumed exponential function, given by

$$t_{IFO}(w) = t_{IFO} + \Delta t_{IFO}\exp(-w/w_o),\ \ w \le W/2, \tag{24}$$

where w is the distance from the edge (w=0) toward the middle of the emitter (w≤W/2), and w_o≈80nm is a characteristic width of the thicker peripheral IFO. Schematically, the non-uniform IFO layer in the emitter opening is illustrated in Figure 5. (For convenience during integrations, the authors of [20] have chosen hyperbolic functions instead of exponential function, but they have mentioned that the choice is not unique, and other function can be used.) Then, as explained in [20] in details, the effective recombination rate is $RR=\{a_1+a_2\times\exp[a_3\times t_{IFO}(w)]\}^{-1}$, with a_1, a_2 and a_3 being constants; and the current densities of the DC and noise currents are obtained by integration along the width of the emitter, using $J(w)\propto RR(w)$ and $j(w)\propto t_{IFO}(w)^3\times J(w)^2$, respectively.

The above model for noise from non-uniform IFO is shown in [20] to have a good agreement with the experimental data. We also observe that the data are very close to the trend in Figure 4, when re-calculating the average thickness of IFO from the information in [20]. However, despite that exponential functions of t_{IFO} are used in the above model for several quantities, the overall dependence of noise level vs. average t_{IFO} remains a power law and we note that the exponential dependence of noise on t_{IFO}, although observed in [14, 33, 34] experimentally, is not explained theoretically yet. Nevertheless, the approach in [20] provides a way to analyze noise from non-uniform IFO in BJT, and can help to identify different contributions to the total low-frequency

noise in deep sub-micron BJT, in which a crossover between several noise sources is obvious; and this will be discussed next.

III.3. Crossover between different noise sources in BJT

Another reason for the scattering in the data for the normalized noise in Figure 2 is that several noise sources in BJT can contribute in different proportions. The proportions between the contributions can vary with the biasing level, and with the layout and fabrication of BJT. Note that the relation $S_I/I_{DC}^2 \propto 1/A$ in eq. (14) is valid only if strictly one noise source dominates, which is usual case for a particular device at particular bias, but not always the case, when the bias, layout, fabrication and interconnection vary, especially for sub-micron area devices. For example, the generation-recombination (GR) currents dominate the base current at low bias, while the injection through the emitter junction of BJT is low, but the noise due to GR currents modulates the injection barrier, and the GR noise is coupled to the collector current by the transconductance, resulting in a quadratic (or nearly) dependence $S_{I_C} \propto I_C^2$. However, at higher biasing, the injection current dominates, and the intrinsic (Hooge) noise in the emitter resistance or injection current takes over the GR noise, resulting in a linear (or nearly) dependence $S_{I_C} \propto I_C$ and eventually a decrease of normalized noise $S_{norm}=S_{I_C}/I_C^2$ can be observed at high bias of BJT. Such cross-over between the noise sources in BJT is reported in [21], for example, when the base current was varied over 2-3 decades, and the cross-over causes a bias-dependent variation for K_F, as illustrated earlier in Figure 2 with the solid diamonds labeled as "$S_{I_B} \propto I_B^{1.2}$". A portion of these data is presented later in Figure 6 as function of the base DC current. One explanation for such cross-over could be that the increased carrier density screens the coupling from trapped charges, but this is not elaborated for BJT, to the best of our knowledge, although explored for MOS transistors [47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59]. We will show that the evolution of the peripheral noise, as uncoupled to the emitter area in BJT, can lead to reduction of the slope in the bias dependence of the noise, while the noise from IFO has opposite evolution.

Typical types of cross-over in the noise in BJT are between areal and peripheral noise sources in the emitter region [14, 21, 35, 36, 42, 60], noise from tunneling through IFO and series resistances (carrier number fluctuation, or coupled noise to the injection process, precisely) and diffusion noise (mobility fluctuation, or intrinsic noise of the current flow in emitter and base regions) [10, 19, 21, 32, 42, 61, 62], surface and bulk noise sources [14, 21, 24, 36, 37, 42, 60], and for very small-area BJTs ($A_E \leq 0.1\mu m^2$), the extension length of the base to the contact via becomes an issue for the low-frequency noise [36, 37], since that extension is with large area compared to A_E, and it is vulnerable to the surface noise from the oxide on the top of the structure. Also, in small-area BJTs, the crossover between generation-recombination (GR), random telegraph signal (RTS) and 1/f noise becomes apparent [11, 12, 17, 18, 30, 35, 63]. The identification of noise sources uses many techniques, such as noise partitioning (or decomposition the total noise in several components), superposition of noise components, evolution of noise components and correlation between them with bias, area and perimeter of the emitter, fitting to physical models and equivalent circuits. The variety of techniques is large and a fully systematic approach in reviewing these is not possible. Nevertheless, there are several useful relations that are accumulated during the years. These are reviewed in [31] and some of them are also discussed below.

III.3.1. Intrinsic noise for the current flow

The intrinsic noise for the current flow is assumed to be due to mobility fluctuation, which follows the Hooge eq. (6) for the carriers in the emitter and base along the direction of the current flow in BJT, while the concentration

of carriers varies, as given by the theory for operation of BJT. Then, since the BJT theory is based on diffusion, the fluctuation in the mobility is transferred to fluctuation in the diffusion coefficient D_m for the minority carriers in the emitter and base regions outside the depletion region of the base-emitter junction, using the Einstein relation $D=\mu\varphi_t$. The resulting expressions for the intrinsic noise of the current flow in the base and collector currents are in the form of

$$S_{I_B} = \frac{\alpha_B'}{f} I_B \text{ and } S_{I_C} = \frac{\alpha_C'}{f} I_C \text{ , for the base and collector current noise, respectively,} \quad (25)$$

where the quantities α' are in unit Ampere and given by [62]

$$\alpha' = \frac{qD_m}{d^2} \alpha_H \ln\left(\frac{n_m(0)}{n_m(d)}\right), \text{ with } \frac{n_m(0)}{n_m(d)} \approx 1 + \frac{v \times d}{D_m}, \quad (26)$$

Respectively for S_{I_B} and S_{I_C}, $d=d_E$ or $d=d_B$ are the thicknesses of the emitter and base regions (along the coordinate of the current flow), D_m corresponds to the diffusion coefficients of the minority carriers in the emitter and base, α_H corresponds to Hooge parameter for minority carriers in the emitter and base, $n_m(0)$ are minority carrier concentrations at the emitter-base junction in the emitter and base sides, $n_m(d)$ are minority carrier concentrations at the other sides of the emitter and base (metal contact for the emitter and collector-base junction for the base), and ν are recombination velocity (~10^5cm/s) of minority carriers at the emitter metal contact or carrier saturation velocity (~10^7 cm/s) that the minority carriers in the base usually reach at the base-collector junction. Slightly different expressions for α' can be found in [61].

Eq. (25) shows that the intrinsic noise for the current flow (due to mobility or diffusion fluctuation) in the base and collector currents is a linear function of the DC currents, and it corresponds the right-hand term of eq. (11). This linear dependence is usually used to identify the diffusion noise in BJT [19, 32, 42] after splitting the total noise into linear and quadratic functions of the DC current, because other noise sources result in non-linear dependence between noise and DC currents. For example, $S_I \propto I_{DC}^2$ for recombination noise at the base surface or at the surface of the emitter-base space-charge region [42]. The expressions for the diffusion S_{I_B} are modified in [10], when IFO is present in the emitter, but the linear dependence remains. Worth mentioning, a linear dependence between noise power spectrum density and DC current in BJT is rarely observed and we will show later that such dependence can be derived from peripheral noise uncoupled to the emitter area.

III.3.2. <u>Base and collector currents are strongly correlated</u>

It is well established that the low-frequency noise in the base and collector currents are strongly correlated. Their normalized cross-correlation spectrum, called often coherence (and corresponding to correlation coefficient for random quantities), is close to one [19, 32, 37, 60]. The coherence between two noise spectra is given by

$$\text{Coherence}(s_x, s_y) = \frac{|S_{xy}|^2}{S_x S_y}, \quad (27)$$

where S_x and S_y are the individual power spectrum densities (PSDs) of the two noise spectra s_x and s_y, and $|S_{xy}|=|s_x s_y^*|$ is the magnitude of the cross-correlation PSD of these noise spectra s_x and s_y. If the coherence is close to 1, then s_x and s_y are strongly correlated and most probably originate from the same noise source, since the different noise sources are independent each from other, either by assumption, or because they contribute in

different proportion at different terminals in BJT. Simply, a coherence=1 for base and collector noise currents means that S_{I_B} and S_{I_C} are coupled in BJT; and we can use eq. (11) to investigate the coupling. The reason to address the coupling is that in the literature the correlation between base and collector currents is usually analyzed in terms of feedback on the emitter resistance r_e (see Figure 3a) and current gain β, while both the feedback and β are secondary effects of the primary diffusion process that explains the operation of BJT. The diffusion process depends on the diffusion properties of the base-emitter junction and the voltage applied on this junction. For example, generation-recombination (GR) currents in base current do not affect the DC collector current, if r_e=0, as seen from Gummel plots for DC currents in BJT, whereas the GR noise in the base current occurs in the collector current well correlated. So, in a simplified form for diffusion current in the base of BJT, we can write

$$I_{B,DIF} = I_{B0} \exp\left(\frac{V_{BE}}{\varphi_t}\right), \tag{28}$$

where $I_{B0} \propto D_m$ and $\exp(V_{BE}/\varphi_t) \propto n_m(0)$, as defined after eq. (26). Assume that $D_m \propto \mu$ fluctuates as given by eq. (25), V_{BE} is coupled to this fluctuation, but other parameters, e.g. φ_t, ν, d in eq. (26), do not fluctuate. By taking the logarithm of eq. (28), one writes

$$\frac{\partial(I_{B,DIF})}{I_{B,DIF}} = \frac{\partial(I_{B0})}{I_{B0}} + \frac{\partial(V_{BE})}{\varphi_t} = \frac{\partial(D_m)}{D_m} + \frac{\partial(V_{BE})}{\varphi_t}. \tag{29}$$

Provided that the square of the differentials corresponds to power spectrum densities and the variation of V_{BE} is coupled to the variation of the base current (correlation coefficient = ±1), using eq. (25) we get

$$\frac{S_{V_{BE}}\big|_{I_{B,DIF}}}{\varphi_t^2} = \frac{S_{I_{B,DIFF}}}{(I_{B,DIF})^2} = \frac{\alpha_B'}{f}\frac{1}{I_{B,DIF}}, \tag{30}$$

where $S_{V_{BE}}|_{I_{B,DIFF}}$ is the noise in V_{BE} coupled from the diffusion noise in I_B. Since the collector current is also coupled to V_{BE} by a relation similar to eq. (28), then $S_{V_{BE}}|_{I_{B,DIFF}}$ will add a noise component in the collector current I_C, which is coupled to the diffusion noise in the base current, and the total noise in I_C is given by

$$\frac{S_{I_C}}{I_C^2} = \frac{S_{V_{BE}}\big|_{I_{B,DIF}}}{\varphi_t^2} + \frac{S_{I_{C,DIFF}}}{(I_C)^2} = \frac{S_{V_{BE}}}{\varphi_t^2} + \frac{\alpha_C'}{f}\frac{1}{I_C} = \text{coupled} + \text{intrinsic}, \tag{31}$$

where $S_{V_{BE}} = S_{V_{BE}}|_{I_{B,DIFF}} + S_{V_{BE}}|_{IFO} + \ldots$ is the total noise in V_{BE} coupled from different noise sources, i.e. $I_{D,DIF}$, IFO and other, and $S_{I_{C,DIF}}$ is the intrinsic noise in I_C due to mobility fluctuation in the base of the transistor, as given by eq. (25). Since the concentration of minority carriers m_n depends on V_{BE}, then $S_{V_{BE}}$ can be regarded as a coupled number fluctuation in BJT. The above discussion is for unilateral case for noise propagation from input to output, for which the noise sources associated to the base and emitter of BJT couple noise into the collector current, and the intrinsic noise in the collector current is not coupled back to the base. This a reasonable approximation, because of two reasons (at least). First, the contribution of the intrinsic noise $S_{I_{C,DIF}}$ is usually negligible, and second, the backward isolation from collector to base and emitter is high as compared to the forward transmission, conservatively in the bandwidth of the low-frequency noise. Also, the first example below

demonstrates that the diffusion noise in I_B and I_C might not be discriminated each from other in experiments and cumulative experience suggests that the noise in BJT can be successfully referred to the input as a noise in the base current.

Now, let us see how eq. (31) helps to deal with noise partitioning and superposition in BJT.

Example. Diffusion dominates DC and noise currents

The first example is by an assumption that the diffusion dominates both the DC and noise currents in BJT and the current gain $\beta=I_C/I_B$=constant. Using eqs. (30) and (31) for the noise in I_C we write

$$\frac{S_{I_C}}{I_C^2}=\frac{S_{V_{BE}}\Big|_{I_{B,DIF}}}{\varphi_t^2}+\frac{\alpha_C'}{f}\frac{1}{I_C}=\frac{\alpha_B'}{f}\frac{1}{I_B}+\frac{\alpha_C'}{f}\frac{1}{I_C}. \tag{32}$$

Similarly, including the coupled noise from I_C into I_B, for the noise in I_B we get

$$\frac{S_{I_B}}{I_B^2}=\frac{S_{V_{BE}}\Big|_{I_{C,DIF}}}{\varphi_t^2}+\frac{\alpha_B'}{f}\frac{1}{I_B}=\frac{\alpha_C'}{f}\frac{1}{I_C}+\frac{\alpha_B'}{f}\frac{1}{I_B}, \tag{33}$$

which is the same as eq. (32) for the noise in I_C. Since β=constant, then the ratio of the noise in I_C and I_B is

$$\frac{S_{I_C}}{S_{I_B}}=\frac{\frac{\alpha_B'}{f}\frac{I_C}{I_B}I_C+\frac{\alpha_C'}{f}I_C}{\frac{\alpha_C'}{f}\frac{I_B}{I_C}I_B+\frac{\alpha_B'}{f}I_B}=\frac{\alpha_B'\beta I_C+\alpha_C'I_C}{\alpha_C'\frac{I_B}{\beta}+\alpha_B'I_B}=\beta^2\frac{\alpha_B'I_C+\alpha_C'I_B}{\alpha_C'I_B+\alpha_B'I_C}\equiv\beta^2=\text{constant}, \tag{34}$$

and from DC and noise measurements, one cannot separate the parameters α' for the diffusion noise in I_B and I_C. One can estimate approximate values from eq. (26), but the final values for the Hooge parameter α_H will remain in an arbitrary ratio. A good guess for initial values of α_H can be obtained by attributing the noise first only to I_B and then only to I_C, and then check which number makes sense, but this is a guess – the measurement cannot reliably separate the two parameters from one device, in which the diffusion dominates both the DC and noise currents; and it is reasonable to estimate only one value of α_H for the dominant noise contribution.

The approach of finding the noise source with dominant contribution is presented in [19] in convenient form. This approach is based on choosing an appropriate equivalent circuit of the device (and measurement setup), with several, i.e. 3, possible noise sources. Then, the measured noise is attributed to each noise source, and one of them usually fits the data well, which is consistent with the assumption that one noise source practically dominates in the total noise. The approach of the dominant noise source is used several times and it identified that the dominant noise source can be presented as current noise source in the base current [19, 32, 37, 61], although the actual physical process can be in the emitter, e.g. IFO, or at the surface above base-emitter junction. Nevertheless, once the dominant noise source is known, the physical identification of the noise is much easier and focused on finding which physical model can describe the behavior of dominant noise source. This is discussed below with the second example for noise partitioning based on eq. (31) for the coupled and intrinsic noise in BJT.

Example. Noise partitioning. Non-linear dependence of current noise on diffusion noise

The second example for the use of eq. (31) is for cases when the current noise in BJT has non-linear bias

dependence with a slope different from one of the diffusion noise, and the slope changes with the bias current too, indicating cross-over between different physical origins for the noise. According to the discussions above for dominant current noise in the base of BJT, first, we rewrite eq. (31) with the term for the intrinsic noise in I_C omitted, but replaced with the diffusion noise in the base current. That is

$$S_{I_C} = \frac{S_{V_{BE}}}{\varphi_t^2} I_C^2 + \frac{\alpha_B'}{f} \frac{1}{I_{B,DIF}} I_C^2 . \tag{35}$$

Here, $S_{V_{BE}}$ denotes all noise sources coupled to V_{BE}, except for the diffusion noise in the base current, which is explicitly shown. Also, we assume that I_C is a diffusion current, while $I_B=I_{B,DIF}+I_{B,GR}$ consists of diffusion current $I_{B,DIF}$, but may have generation-recombination DC component $I_{B,GR}$ due to either bulk or surface recombination. Second, we refer the noise to the base as the current noise, using AC and DC current gain, $\beta_{AC}=\partial I_C/\partial I_B$ and $\beta_{DC}=I_C/I_B$, respectively. The equation for the base-referred current noise is

$$S_{I_B} = \frac{S_{I_C}}{\beta_{AC}^2} = \left[\frac{S_{V_{BE}}}{\varphi_t^2} \left(I_{B,DIF} + I_{B,GR} \right)^2 + \frac{\alpha_B'}{f} \left(1 + \frac{I_{B,GR}}{I_{B,DIF}} \right) \left(I_{B,DIF} + I_{B,GR} \right) \right] \left(\frac{\beta_{DC}}{\beta_{AC}} \right)^2 . \tag{36}$$

Owing to the presence of $I_{B,GR}$, both β_{AC} and β_{DC} vary with bias, even if one assumes that the ratio $\beta=I_C/I_{B,DIF}$ between the diffusion currents is constant (valid assumption until high bias is applied, that causes high level of injection of majority carriers in the base, and consequently, D_m and β decrease. The high-injection regime is usually not used in low-frequency noise measurement, but it can be reached in submicron-area BJT – see [35]). One can observe many possible cases that follow from eq. (36), since the evolution of S_{I_B} with I_B is dependent on several ratios, e.g. β_{DC}/β_{AC}, $I_{B,GR}/I_{B,DIF}$, which are bias dependent. Let us look at several of these cases.

Case 1 in second example: The diffusion is dominant in the DC currents, that is $I_{B,GR}<<I_{B,DIF}=I_B$ and $\beta_{DC}=\beta_{AC}=\beta$=constant. In this case, eq. (36) reduces to

$$S_{I_B} = \frac{S_{V_{BE}}}{\varphi_t^2} \left(I_B \right)^2 + \frac{\alpha_B'}{f} \left(I_B \right), \text{ if diffusion dominates DC} \tag{37}$$

One situation is that no noise is coupled to the diffusion process in BJT. In this situation, $S_{I_B} \propto I_B$, and this is the situation of dominant diffusion (intrinsic) noise discussed above – see eq. (33). A second situation is when the coupled noise has a constant value as a voltage noise, $S_{V_{BE}}$=constant, and it is dominant. In the second situation, $S_{I_B} \propto (I_B)^2$. Constant $S_{V_{BE}}$ is following from bias-independent fluctuations that can be regarded as resistance fluctuations, e.g. noise form IFO in the emitter [10], or charge trapping and recombination that modulate potential of the base, emitter or emitter-base junction [42]. The normalized noise of these fluctuations is nearly constant [42], that is S_I/I_{DC}^2=constant. Let us take the noise form IFO. It is shown in [10] that the noise of IFO is a fluctuation in the resistance of the emitter region (see r_e in Figure 3a).

$$\frac{S_{r_{IFO}}}{r_{IFO}^2} \approx \frac{S_{r_e}}{r_e^2} = \frac{S_{I_E}}{I_E^2} = \frac{S_{I_B}}{I_B^2} = \frac{S_{V_{BE}}\big|_{IFO}}{\varphi_t^2} = \frac{\alpha''_{IFO}}{f} = \text{constant in respect to bias}, \tag{38}$$

as far as β=constant is assumed for the case when the diffusion dominates the DC currents. Here, we have recalled the left-hand equality of eq. (30), which describes the coupling between voltage and current noise on a

pn junction. So, adding the coupled noise from IFO in eq. (37), one gets

$$S_{I_B} = \frac{\alpha''_{IFO}}{f}(I_B)^2 + \frac{\alpha'_B}{f}(I_B), \text{ if diffusion dominates DC.} \quad (39)$$

There is a critical value $I_{Bcr}=\alpha'_B/\alpha''_{IFO}$ in eq. (39) for crossover between diffusion noise and noise from IFO. If $I_B<I_{Bcr}$, then $S_{I_B}\propto I_B$ because the diffusion noise dominates. In contrary, if $I_B>I_{Bcr}$, then $S_{I_B}\propto(I_B)^2$, because the noise from IFO dominates. The crossover is observed for the large-area samples in [10], as shown in Figure 6 with squares, and similar data are reported in [19]. Based on the different slopes of S_{I_B} as function of I_B for diffusion noise and noise form IFO, these noise sources are separated and accordingly characterized [10, 19, 32], respectively, in terms of Hooge parameter α_H for the diffusion noise, since $\alpha'_B\propto\alpha_H$ in eq. (26), and in terms of $K_F=f\times S_{I_B}/I_B^2=\alpha''_{IFO}$ for the noise from IFO, as follows from eq. (38). The data suggest cumulatively that the diffusion noise is important for BJT with larger emitter area and thin IFO, while the noise from IFO dominates BJTs with emitter area less than $10\mu m^2$ and $t_{IFO}>0.5nm$. The nowadays BJTs usually are in the latter category.

Case 2 in second example (peripheral-areal noise): The diffusion is not always dominant in the DC currents. For example, at low bias, a generation-recombination (GR) process can contribute to I_B a component, given by

$$I_{B,GR} = I_{GR0}\exp\left(\frac{V_{BE}}{\eta_{GR}\varphi_t}\right), \quad (40)$$

where η_{GR} is non-ideality factor with values usually between 1.5 and 2. Following the procedure after eq. (28), the GR process couples voltage noise in the V_{BE}, given by

$$\frac{S_{V_{BE}}\big|_{I_{B,GR}}}{\varphi_t^2} = \eta_{GR}^2\frac{S_{I_{GR}}}{I_{B,GR}^2} = \frac{\alpha''_{GR}}{f} \approx \text{constant in respect to bias,} \quad (41)$$

since $S_{I_{GR}}/(I_{GR})^2\approx$constant for charge trapping [42], as mentioned above. In the presence of $I_{B,GR}$, the ratio β_{DC}/β_{AC} is

$$\frac{\beta_{DC}}{\beta_{AC}} = \frac{\frac{I_C}{I_{B,DIF}+I_{B,GR}}}{\frac{\partial I_C}{\partial(I_{B,DIF}+I_{B,GR})}} = \frac{\beta\times I_{B,DIF}}{I_{B,DIF}+I_{B,GR}}\frac{\partial(I_{B,DIF}+I_{B,GR})}{\beta\times\partial I_{B,DIF}} = \frac{I_{B,DIF}}{I_{B,DIF}+I_{B,GR}}\left(1+\frac{I_{B,GR}}{\eta_{GR}I_{B,DIF}}\right) \quad (42)$$

$$\frac{\beta_{DC}}{\beta_{AC}} = \frac{\eta_{GR}I_{B,DIF}+I_{B,GR}}{\eta_{GR}(I_{B,DIF}+I_{B,GR})} = \begin{cases} 1, \text{if } I_{B,DIF} >> I_{B,GR} \\ \frac{1}{\eta_{GR}}, \text{if } I_{B,DIF} << I_{B,GR} \end{cases} \quad (43)$$

From eq. (36), in the case of existing GR component in I_B, the base current noise becomes

$$S_{I_B} = \left(\frac{\alpha''_{GR}}{f}+\frac{\alpha'_B}{f}\frac{1}{I_{B,DIF}}\right)\left(I_{B,DIF}+\frac{I_{B,GR}}{\eta_{GR}}\right)^2 \quad (44)$$

At high bias $I_B\approx I_{B,DIF}>>I_{B,GR}$, and one observes $S_{I_B}\propto(I_B)^2$, since also $I_{B,DIF}>I_{Bcr}=\alpha'_B/\alpha''_{GR}$. At low bias, however,

when $I_{B,DIF} << I_{B,GR} \approx I_B$, several situations are possible. If $I_{B,DIF} > I_{Bcr} = \alpha'_B/\alpha''_{GR}$, then $S_{I_B} \propto (I_B)^2$, but slightly attenuated with $(\eta_{GR})^2$, and one may observe small step in the $S_{I_B} \propto (I_B)^2$ dependence at the crossover between $I_B \approx I_{B,GR}$ and $I_B \approx I_{B,DIF}$. In a situation when the coupled noise is low and I_{Bcr} is larger than the DC current for crossover between $I_B \approx I_{B,GR}$ and $I_B \approx I_{B,DIF}$, one can find a dependence that differs from $S_{I_B} \propto (I_B)^2$ and $S_{I_B} \propto (I_B)$. For this situation, we rewrite eq. (44) with α''_{GR} omitted, as

$$S_{I_B} = \frac{\alpha'_B}{f}\left(1+\frac{I_{B,GR}}{\eta_{GR} I_{B,DIF}}\right)\left(I_{B,DIF}+\frac{I_{B,GR}}{\eta_{GR}}\right), \text{ for diffusion (intrinsic) noise.} \quad (45)$$

At the higher bias end, where $I_{B,GR} < I_{B,DIF} \approx I_B$, then $S_{I_B} \propto (I_B)$. In contrary, at the lower bias end, where $\eta_{GR} I_{B,DIF} < I_{B,GR} \approx I_B$, then

$$S_{I_B} = \frac{\alpha'_B}{f}\frac{I^2_{B,GR}}{\eta^2_{GR} I_{B,DIF}} = \frac{\alpha'_B}{f}\frac{1}{\eta^2_{GR}}\frac{I^2_{GR0}}{I_{B0}}\exp\left[\frac{V_{BE}}{\varphi_t}\left(\frac{2}{\eta_{GR}}-1\right)\right] = \begin{cases}\text{from constant @ } \eta_{GR}=2 \\ \text{to } S_{I_B} \propto I_B \text{ @ } \eta_{GR}=1\end{cases}. \quad (46)$$

This equation implies that the intrinsic diffusion noise in BJT levels off when the base DC current is dominated by generation-recombination currents. This effect is observed several times when the BJT was stressed electrically or after proton irradiation, and the generation-recombination currents are increased. An example from [21] is shown in Figure 6. Before stress, circles in Figure 6, a slope close $S_{I_B} \propto (I_B)$ is observed, indicating low level of coupled noise owing to screening of surface noise from emitter-base junction, achieved by a high doping concentration at the surface of the base, using superficial base doping (SBD). Also, the step in the $S_{I_B} \propto (I_B)$ dependence is observed in this sample at the transition region from low $I_B \approx I_{B,GR}$ to higher $I_B \approx I_{B,DIF}$. After the stress, triangles in Figure 6, the $S_{I_B} \propto (I_B)$ dependence levels off due to increased $I_{B,GR}$.

The departure from $S_{I_B} \propto (I_B)$ and $S_{I_B} \propto (I_B)^2$ is attributed to generation-recombination currents at the surface above the base-emitter junction [14, 36, 37, 42, 60], because it is found that apart from $1/A_E$ dependence for the normalized noise ($K_F = f \times S_{I_B}/I_B^2 \propto 1/A_E$) in the base current, both the DC and noise currents in the base scale with the perimeter P_E of the emitter [11, 14, 36, 37, 60], rather than only with the emitter area A_E. So, partitioning between areal and peripheral noise in BJT takes place, by using equations in the form of

$$S_{I_B} = \frac{K_{F,A_E}}{f} I^2_{B,A_E} + \frac{K_{F,P_E}}{f} I^2_{B,P_E} \approx \frac{K_{F,A}}{f} I^2_{B,DIF} + \frac{K_{F,P}}{f} I^2_{B,GR}, \quad (47)$$

where $K_{F,A}$ ($\equiv K_{F,A_E}$) is the normalized areal noise, with $K_{F,A} \propto 1/A_E$, and $K_{F,P}$ ($\equiv K_{F,P_E}$) is the normalized peripheral noise, with $K_{F,P} \propto 1/P_E$. The peripheral noise is usually analyzed as a surface noise attributed to an area $A_P = P_E \times W_P$ at the spacer oxide [36, 37, 42], where the characteristic width W_P is taken as the width of the depletion region of base-emitter junction [37]. The fluctuation of charge trapping, or tunneling associated with the oxide at this surface A_P, is regarded as the origin for GR current and corresponding surface noise. This illustrated in Figure 7. The fluctuation of charge at/in the oxide modulates the potential in the base-emitter junction in vicinity of the trapped charge causing fluctuation in the carrier concentrations in this vicinity, and thus, current noise in BJT due to number fluctuation of carriers. The assumption that the traps affect the depletion region of width W_P of base-emitter junction is reasonable, but the value of W_P~15-20nm in [37] is

small as compared to values 50-100nm for peripheral effects. Our estimate [46] is that the potential bending around a single trapped charge results in a strong and almost constant effect in a distance of about 15nm, and the effect gradually decreases in a distance up to 50nm. This is shown with dashed lines in Figure 7. So, a "remote" coupling of the fluctuation of trapped charges is expected, and the width, as well as the depth, of coupling of surface noise is normally in the range of 50-80nm [20, 37]. Nevertheless, this range of distances is smaller than the width of the emitter, and it implies a peripheral noise that scales with the perimeter of the emitter, rather than with the area of the emitter. The surface noise can couple noise in the resistance of the base too, since the length of base extension to the metal contact can be large as compared to the emitter width, but this effect is not dominant noise source in BJT, because the doping of the base and the extension is high, and they are not depleted from carriers in principle. Therefore, outside the vicinity of depletion region of the base-emitter junction, the large number of carriers screens the fluctuation of trapped charges, although some additional noise is observed in minimum-sized BJTs [37].

To separate the peripheral noise from areal noise in BJT during experiments, one writes eq. (47) in two equivalent forms, given by

$$f\frac{A_E\times S_{I_B}}{I_{B,DIF}^2}=\left(A_E\times K_{F,A}\right)+\left(P_E\times K_{F,P}\right)\frac{A_E}{P_E}\frac{I_{B,GR}^2}{I_{B,DIF}^2},\ \approx \text{areal noise } (A_E\times K_{F,A}) \text{ at high } I_B, \qquad (48)$$

$$f\frac{P_E\times S_{I_B}}{I_{B,GR}^2}=\left(A_E\times K_{F,A}\right)\frac{P_E}{A_E}\frac{I_{B,DIF}^2}{I_{B,GR}^2}+\left(P_E\times K_{F,P}\right),\ \approx \text{peripheral noise } (P_E\times K_{F,P}) \text{ at low } I_B. \qquad (49)$$

Then, the DC currents and low-frequency noise of samples with known A_E and P_E and different ratio A_E/P_E are measured in several decades for I_B. A check for quadratic dependence $S_{I_B}\propto(I_B)^2$ is helpful to verify that the noise is coupled (from IFO and surface) and not intrinsic (from diffusion) in order to guarantee the validity of eq. (47). Next, the diffusion and surface DC currents are separated from Gummel plots for each sample in a manner so that $I_{B,DIF}\propto A_E$ and $I_{B,GR}\propto P_E$ in order to verify that the assumptions for scaling with emitter area and perimeter are valid in the experiment, e.g. $I_{B,GR}$ might not be a peripheral current in BJT. Finally, the quantities in the left-hand sides of eqs. (48) and (49) are calculated and plotted against the quantity $[A_E\times(I_{B,GR})^2]/[P_E\times(I_{B,DIF})^2]$ and its reciprocal, respectively. If the assumptions for areal and peripheral noise apply for the samples, the intersection of the linear fit in the first plot with the vertical axis should match with the slope of the linear fit in the other plot (and vice versa). If so, then one can use $(A_E\times K_{F,A})$ as the measure for the areal noise in BJT and $(P_E\times K_{F,P})$ as the measure for the peripheral noise; and proceed to physical identification of noise sources. Certainly, the procedure for separation of areal and peripheral noise is quite demanding, and we observe only portions of this procedure reported in the literature [11, 36, 37, 60].

Also, some peripheral effects in BJT, such as thicker IFO at emitter edges [20, 36] and current crowding at emitter periphery [35], may result in higher noise, but not necessary cause higher base DC current $I_{B,GR}$. In these cases, the method above may result in unrealistic values from physical point of view [37] and it has to be modified to converge the approaches in [36] to that in [31], the latter discussed earlier by the help of eq. (24) for non-uniform IFO. The modifications are the following. From DC characteristics of small and large area BJT with different ratio A_E/P_E, one has to determine the characteristic width W_P of the peripheral regions and the areal and peripheral DC current density pre-factors J_{0A} and J_{0P}, respectively, using the equation

$$I_B = (A_E - P_E W_P) J_{0A} \exp\left(\frac{V_{BE}}{\eta_A \varphi_t}\right) + P_E W_P J_{0P} \exp\left(\frac{V_{BE}}{\eta_P \varphi_t}\right) + P_E W_P J_{0GR} \exp\left(\frac{V_{BE}}{\eta_{GR} \varphi_t}\right), \quad (50)$$

so that one set of values for W_P, J_{0A}, J_{0P}, J_{0GR}, $\eta_A \approx \eta_P \approx 1$, and $\eta_{GR} \approx 2$ fits the measured curves. An initial value for W_P can be between 50nm [36] and 80nm [31]. The last term in the equation is for surface currents, and at high bias can be neglected, where also an approximation $\eta = \eta_A = \eta_P \approx 1$ holds, and the equation can be reduced to

$$\frac{I_B}{A_E} \exp\left(-\frac{V_{BE}}{\eta \varphi_t}\right) = \left(1 - \frac{P_E}{A_E} W_P\right) J_{0A} + \frac{P_E}{A_E} W_P J_{0P}. \quad (51)$$

The right-hand side of this equation reduces to J_{OA} for large-area, nearly square-shaped BJTs, and thus, J_{OA} is easy to find. With the initial value for W_P, one can find initial value for J_{OP}, using data from smaller-area, rectangular-shaped transistors; but the next step will require optimization procedure, first, to fit the data from DC measurements at high bias, varying W_P and J_{OP} in eq. (51), and then, another optimization procedure to fit all data, both at high and low bias, using eq. (50). This is not convenient. Therefore, another approach is taken in [36]. It is estimated that $P_E \times W_P / A_E << 1$, which reduces eq. (51) to

$$\frac{I_B}{A_E} \exp\left(-\frac{V_{BE}}{\eta \varphi_t}\right) = J_{0A} + \frac{P_E}{A_E} W_P J_{0P}. \quad (52)$$

The quantity in the left-hand of the equation is the current density prefactor J_{B0} for the base current in the samples, and J_{B0} is obtained for each sample from DC measurements. The values of J_{B0} are plotted against the ratio P_E / A_E, and from the slope of this plot, the values for the product $W_P \times J_{OP}$ are obtained and directly used, since, actually, $W_P \times J_{OP}$ is needed in eq. (50), in order to analyze the noise form peripheral IFO; and the GR currents can be easily removed from the measured data too.

Having $W_P \times J_{OP}$ estimated, one can obtain the peripheral I_P and areal I_A emitter currents by

$$I_P = P_E W_P J_{0P} \exp\left(\frac{V_{BE}}{\eta_P \varphi_t}\right) \approx P_E W_P J_{0P} \exp\left(\frac{V_{BE}}{\varphi_t}\right), \text{ since } \eta = \eta_A = \eta_P \approx 1 \text{ is assumed,} \quad (53)$$

and from measured base current I_B one can get the areal emitter current I_A , using

$$I_A = I_B - I_P \quad (-I_{B,GR}, \text{ if necessary}). \quad (54)$$

Care should be taken so that I_A is greater than zero.

By making substitutions $I_P \leftrightarrow I_{B,GR}$ and $I_A \leftrightarrow I_{B,DIF}$, eqs. (48) and (49) can be rewritten as

$$f \frac{A_E \times S_{I_B}}{I_A^2} = (A_E \times K_{F,A}) + (P_E \times K_{F,P}) \frac{A_E}{P_E} \frac{I_P^2}{I_A^2}, \quad (55)$$

$$f \frac{P_E \times S_{I_B}}{I_P^2} = (A_E \times K_{F,A}) \frac{P_E}{A_E} \frac{I_A^2}{I_P^2} + (P_E \times K_{F,P}), \quad (56)$$

by also neglecting the noise from surface currents. At this point, all modifications for the noise partitioning between areal and peripheral noise from IFO are made, and the procedure continues as described above for surface noise. One note should be made. We have assumed that $P_E \times W_P / A_E << 1$, which might be not precise, if

the width of the emitter is less than 0.5μm [20, 64, 65]. For such devices, either the method from [20] has to be used, or an increase in the average IFO thickness can be assumed in proportion to the decrease of the emitter width.

One interesting feature for the peripheral noise is that it causes variable slope in the bias dependence of S_{I_B}. This is continuously addressed in the literature, for example in [11, 14, 18, 21, 31, 42, 60, 62]. Since the peripheral noise is not coupled to the areal current, then the different non-ideality factors $\eta_A \approx 1 \neq \eta_P = \eta_{GR} \approx 2$ for areal and peripheral DC currents cause the variable slope, so that $S_{I_B} \propto (I_B)^m$, and $1<m<2$. The assumption is that the peripheral noise is dominant and eq. (47) can be written as

$$S_{I_B} \approx \frac{K_{F,P}}{f} I_{B,GR}^2 = \frac{K_{F,P}}{f} \frac{I_{B,GR}^2}{\left(I_{B,DIF} + I_{B,GR}\right)^2} I_B^2, \tag{57}$$

where $I_B = I_{B,DIF} + I_{B,GR}$ is the total base DC current (areal + peripheral) and $K_{F,P}$=constant is solely due to peripheral noise, which, however, is not coupled to the diffusion in the emitter area. Worth mentioning, such decoupling of peripheral noise from the BJT emitter area results in a noise current source between base and collector, as explained in [42].

The DC currents in eq. (57) are with different non-ideality factors $\eta_A \approx \eta_{DIF} = 1 \neq \eta_P \approx \eta_{GR} = 2$, and given by

$$I_{B,DIF} = I_{0A} \exp\left(\frac{V_{BE}}{\eta_A \varphi_t}\right) >> I_{B,GR} \text{ at high } V_{BE}, \tag{58}$$

$$I_{B,GR} = I_{0P} \exp\left(\frac{V_{BE}}{\eta_P \varphi_t}\right) >> I_{B,DIF} \text{ at low } V_{BE}. \tag{59}$$

So, at low bias, $I_B \approx I_{B,GR} >> I_{B,DIF}$, then $S_{I_B} \propto (I_B)^2$, and one should expect a high slope $m \approx 2$ in the bias dependence of the noise. In contrast, at high bias, $I_B \approx I_{B,DIF} >> I_{B,GR}$, then one can see that

$$\frac{I_{B,GR}^2}{\left(I_{B,DIF} + I_{B,GR}\right)^2} \approx \left(\frac{I_{B,GR}}{I_{B,DIF}}\right)^2 = \left[\frac{I_{0P} \exp\left(\frac{V_{BE}}{\eta_P \varphi_t}\right)}{I_{0A} \exp\left(\frac{V_{BE}}{\eta_A \varphi_t}\right)}\right]^2 = \frac{I_{0P}^2}{I_{0A}^{2\eta_A/\eta_P}} \frac{I_B^{2\eta_A/\eta_P}}{I_B^2}, \tag{60}$$

and S_{I_B} becomes sub-quadratic function of I_B, given by

$$S_{I_B} \approx \frac{K_{F,P}}{f} \frac{I_{B,GR}^2}{I_{B,DIF}^2} I_B^2 = \frac{K_{F,P}}{f} \frac{I_{0P}^2 I_B^{2\eta_A/\eta_P}}{I_{0A}^{2\eta_A/\eta_P}} = \frac{K_{F,P}}{f} \frac{I_{0P}^2}{I_{0A}^m} I_B^m, \text{ with } m = 2\eta_A/\eta_P \text{ at } I_B \approx I_{B,DIF} >> I_{B,GR}. \tag{61}$$

As the consequence, one can observe a lower slope $m \approx 1$ in the bias dependence of the noise at high bias, $I_B \approx I_{B,DIF} >> I_{B,GR}$, when the surface recombination is the dominant but peripheral noise source in BJT. The decrease of the slope in the bias dependence of the noise is the argued in [21] as an indication for peripheral noise, and we show sample of the data from this publication in Figure 6 with circles and triangles. Note that the bias evolution of the peripheral noise, since it is uncoupled from the areal current in BJT, is different from evolution of areal noise coupled from IFO shown in Figure 6 with squares.

III.3.3. Lorentzian noise superimposes 1/f noise

So far, we have assumed that the low-frequency noise in BJT is 1/f noise with a constant slope as function of the frequency. This is a good approximation for large-area BJT, but significant deviations in the frequency slope are prominent in submicron-area BJT. It is observed many times, for example in [11, 12, 17, 18, 30, 31, 35] that reducing the emitter area, low-frequency noise with a Lorentzian shape of the power spectrum density (PSD) superimposes on the 1/f noise. Therefore, the noise PSD in the base current of BJT is given with

$$S_{I_B} = 2qI_B + \frac{K_F}{f} I_B^2 + \sum_i \frac{B_i \tau_i}{1 + (2\pi f \tau_i)^2}, \tag{62}$$

as a superposition of shot noise $2qI_B$ with constant, "white", spectrum in respect to frequency, flicker noise with 1/f or "pink" PSD and a sum of several Lorentzian PSDs, in each of which τ_i is a characteristic time constant so that at low frequency $f<1/(2\pi\tau_i)$, the Lorentzian noise component has approximately constant magnitude $S_i(0)=B_i\tau_i$, while the magnitude of the Lorentzian noise decreases at higher frequencies $f>1/(2\pi\tau_i)$ with a slope $1/f^2$. The shot noise originates from discrete nature of the current, quantized by the electron charge q, when the carriers overcome the injection barrier of the emitter-base junction and then move fast through the short base in the BJT, resulting in very dense burst of current "shots". Each shot can be represented as a Dirac pulse, resulting in a white spectrum. The 1/f portion in S_{I_B} we have discussed above. The Lorentzian components originate from random bistable processes with amplitude $\Delta I \propto \sqrt{B_i}$, such as generation-recombination (GR), which is well established for semiconductor devices, and in particular for BJTs it is shown in [11, 17] that Shockley–Read–Hall statistics holds. It follows from this that $B_i \propto I_B^2$ and $\tau_i \propto 1/I_B$, because of the following general reasons.

The time constant τ_i of a Lorentzian component is smaller than the emission and capture time constants of the trapping GR center, and it is given by

$$\frac{1}{\tau_i} = \frac{1}{\tau_c} + \frac{1}{\tau_e} = \sigma v_{th} n' + \frac{\sigma v_{th} n' g}{\exp\left(\frac{E_F - E_T}{\varphi_t}\right)} = \sigma v_{th} n' \left[1 + \frac{g}{\exp\left(\frac{E_F - E_T}{\varphi_t}\right)}\right], \tag{63}$$

where $v_{th} \approx 10^7$cm/s is the thermal velocity of carriers in silicon; and other constants associated with each trap are capture cross-section σ, energy level E_T in eV and degeneracy factor g~1. It is reasonable to state that the base DC current I_B is proportional to both carrier concentration n' and $\exp[(E_F-E_T)/\varphi_t]$, since the Fermi level E_F in eV is nearly linear function of the bias voltage V_{BE}. Combining the constants, one can write

$$\frac{1}{\tau_i} = \frac{1}{\tau_c} + \frac{1}{\tau_e} = a_1 I_B \left(1 + \frac{a_2}{I_B}\right) = a_1 (I_B + a_2) \Leftrightarrow \tau_i = \frac{1/a_1}{(I_B + a_2)}, \tag{64}$$

where a_1 and a_2 are constants, and $\tau_i \approx$constant at "low" DC currents and decreases as $\tau_i \propto 1/I_B$ at "high" currents. Therefore, the increase of the bias current in BJT causes an increase of the corner frequency of the individual Lorentzian spectra in eq. (62).

The quadratic dependence $B_i \propto I_B^2$ can be easily understood in terms of coupled noise. As discussed above, the trapped charge bends the potentials around it, and it couples a change in the voltage at the emitter-base junction

with approximately constant amplitude $\Delta V_{BE} \propto q/C$, where C is the capacitance of the emitter-base junction. Taking the resistance of the base-emitter junction $r_\pi=\varphi_t/I_B$, one gets $\Delta I_B=\Delta V_{BE}\times I_B/\varphi_t$, and for the noise coefficient B_i, one writes

$$B_i \propto (\Delta I_B)^2 = \frac{(\Delta V_{BE})^2}{\varphi_t^2}(I_B)^2 \propto \left(\frac{q}{\varphi_t C}\right)^2 (I_B)^2 \Rightarrow B_i = a_3 (I_B)^2, \tag{65}$$

where a_3 is constant for a particular trapping GR center, a_3 can vary between the traps, but a_3 is a weak function of the bias. Combining eqs. (64) and (65), the bias dependence of the low-frequency plateau $S_i(0)=B_i\tau_i$ of a single Lorentzian noise component becomes

$$S_i(0) = B_i\tau_i = a_3 (I_B)^2 \frac{1/a_1}{(I_B + a_2)} \propto \begin{cases} (I_B)^2, \text{at low currents } I_B < a_2 \\ (I_B), \text{at high currents } I_B > a_2 \end{cases}, \tag{66}$$

where a_1, a_2 and a_3 are approximately constant in respect to bias. Thus, one observes a steeper bias dependence of $S_i(0)$ at low currents, by comparing to the slope in this dependence at high bias.

The bias dependence of Lorentzian noise coupled from single generation-recombination center in BJT is illustrated in Figure 8 with an example from [12]. The evolution of the power spectrum density in Figure 8Figure 8a includes an increase of the corner frequency and an increase of plateau of the Lorentzian spectrum, as suggested by eqs. (64) and (66), respectively. The bias dependence of different parameters in Lorentzian spectrum is illustrated in Figure 8b. The noise coefficient B_i increases strongly with the bias according to eq. (65), as shown with squares in Figure 8b. The corner frequency $1/(2\pi\tau_i)$ also increases with bias, but with lower slope according to eq. (64), as shown with circles in Figure 8b. The knee in the bias dependence of $S_i(0)$ is also clearly visible from the triangles, and it is in agreement with the prediction of eq. (66). Similar bias dependence of the Lorentzian noise is observed in [35], where it is observed that the ratio $S_i(0)/\tau_i$ is a function I_B, and this function is with a constant slope. This is expected from eqs. (64) and (66), because

$$\frac{S_i(0)}{\tau_i} = \frac{B_i\tau_i}{\tau_i} = B_i \propto (I_B)^m, \text{ with } m=2\eta_A/\eta_P, \tag{67}$$

depending on whether areal or peripheral noise is coupled in BJT, as discussed above. If the Lorentzian noise originates from the emitter area, then $m\approx 2$, as shown by eq. (65). If the Lorentzian noise originates from the emitter periphery, the slope is expected to be lower, $m\sim 1$, as discussed after eq. (61). In [35], the slope is m=1.3 and indicates peripheral Lorentzian noise, which agrees with the investigation on the current crowding in BJT addressed in this publication.

Crossover between 1/f and GR noise

The relative contribution between 1/f noise and Lorentzian noise is found to be dependent on the emitter area [17, 18, 30, 31, 33]. In large-area BJTs, the portion of 1/f noise is dominant in the low-frequency noise, as shown in the right-hand plot of Figure 9. By decreasing the emitter area, however, the low-frequency noise becomes dominated by Lorentzian noise as shown in the left-hand plot of Figure 9. Nevertheless, by averaging the noise from several small-area BJT one gets nearly 1/f spectrum, and what is even more interesting, the values for the product $(A_E\times K_F)$ obtained from the average noise of small devices match the values obtained from large-area BJT, in which the Lorentzian noise is low. Therefore, it is deduced in [17, 18, 30, 31, 33, 60] that the 1/f noise

can be due to superposition of many Lorentzian components originating from a large number of trapping GR centers in large-area BJT at a given trap density, while in small-area BJTs, having approximately the same trap density, the number of traps becomes small and the individual Lorentzian components become distinguishable. So, the origin 1/f noise can be superposition of Lorentzian noise, and the term for the 1/f noise in eq. (62) can be modeled in terms of this superposition. The model is developed in [17, 66, 67] and reviewed in [18, 31]; and similar model is used in [33, 60] to get insight for the noise in SiGe HBTs.

The model [17, 66, 67] uses the assumption that the trap density is constant in the emitter area, and complies with the observations that the low-frequency noise, on average, is with 1/f power spectrum density (1/f), having magnitude inversely proportional to the area (A_E) and quadratic function of the base current (I_B^2). The variation of noise around the average is expected to be inversely proportional to the square root of the area. Also, when the emitter area is small, bistable Random Telegraph Signal (RTS) noise is expected from individual traps. Let us see how one can get 1/f noise from the superposition of Lorentzian components at the above assumptions.

Consider the right-hand term in eq. (62)

$$S_{I_B} = \sum_{i=1}^{A_E \times N_{tA}} \frac{B_i \tau_i}{1+(2\pi f \tau_i)^2}, \text{ with } \quad B_i = 4(\Delta I_i)^2 \tag{68}$$

where N_{tA} is the constant trap density per unit emitter area, $A_E \times N_{tA}$ is the total number of traps in the emitter area A_E which couple RTS noise, ΔI_i is the amplitude of the RTS from the i^{th} trap. The multiplier 4 depends on the ratio between capture τ_c and emission τ_e times in RTS [4, 68, 69], it can decrease, but usually is taken constant for simplicity [70].

From the discussion for eq. (65), ΔI_i is coupled from single trapped charge q, and it is

$$\Delta I_{B,i} = a_i \left(\frac{q}{\varphi_t C} \right) I_B , \tag{69}$$

where $C=C_A \times A_E$ is the capacitance of the emitter-base junction, $C_A=\varepsilon_{Si}/W_J$ is the capacitance per unit area of the depleted region of the emitter-base junction, $\varepsilon_{Si}=1.04\times10^{-12}$ F/cm is permittivity of silicon and W_J is the width of the depletion region, which is bias dependent. The parameter a_i represents the strength of amplitude coupling, it corresponds to $\sqrt{K}$ in the general equations (1), (4), (5), and it is expected that $a_i<1$ in the following analysis, since the coupling is attenuated by the diffusion impedance $r_\pi \| (2\pi f C_{DIF})^{-1}$ of the forward biased junction. The expressions for W_J are given in [71], for example, and for an abrupt junction, they are

$$W_J = \sqrt{\frac{2\varepsilon_{Si}}{q} \frac{N_d + N_a}{N_d N_a}} \sqrt{V_{bi} - V_{BE}} , \text{ where } \quad V_{bi} = \varphi_t \ln\left(\frac{N_d N_a}{n_i^2} \right) , \tag{70}$$

where N_d and N_a are the volume doping concentrations at the emitter-base junction, and $n_i=1.5\times10^{10}\approx10^{10}$cm^{-3} is the carrier concentration in intrinsic silicon at room temperature.

Substituting in eq. (68), one gets

$$S_{I_B} = 4\left(\frac{q W_J}{\varphi_t \varepsilon_{Si}} \right)^2 \frac{I_B^2}{A_E^2} \sum_{i=1}^{A_E \times N_{tA}} \frac{a_i^2 \tau_i}{1+(2\pi f \tau_i)^2} . \tag{71}$$

The number of acting traps $A_E \times N_{tA}$ around the Fermi level and their coupling a_i vary between samples and with bias. For large emitter area BJT, however, one can use average value a_{avg}=average(a_i) and assume a continuous distribution g(τ) for the traps in respect to their time constants τ_i, within a range from τ_{min} to τ_{max}, by writing

$$N_{tA} = N_{tA} \int_{\tau_{min}}^{\tau_{max}} g(\tau) d\tau \text{ , so that } \int_{\tau_{min}}^{\tau_{max}} g(\tau) d\tau = 1 . \tag{72}$$

So, eq. (71) becomes

$$S_{I_B} = 4 a_{avg}^2 \left(\frac{q W_J}{\varphi_t \varepsilon_{Si}} \right)^2 \frac{I_B^2}{A_E^2} A_E N_{tA} \int_{\tau_{min}}^{\tau_{max}} \frac{g(\tau) \tau d\tau}{1 + (2\pi f \tau)^2} . \tag{73}$$

Furthermore, it needs g(τ)×τ=b=constant in order to obtain 1/f spectrum from the integral. This is because

$$\int_{\tau_{min}}^{\tau_{max}} \frac{d\tau}{1 + (2\pi f \tau)^2} = \frac{1}{2\pi f} \left[\operatorname{arctg}(2\pi f \tau_{max}) - \operatorname{arctg}(2\pi f \tau_{min}) \right] \approx \frac{1}{4f} , \tag{74}$$

at conditions $(2\pi f_{min} \tau_{max}) >> 1$ and $(2\pi f_{max} \tau_{min}) << 1$. Assume that the noise is measured in 4 decades, e.g. decimal $\log(f_{max}/f_{min})=4$, then the time constants of the traps should be distributed as g(τ)=b/τ at least 2 decades above and below to satisfy the conditions for obtaining 1/f noise from superposition of Lorentzian spectra. This means, that $\log(\tau_{max}/\tau_{min}) \geq 8$. The reasons for the g(τ)=b/τ distribution in a wide range are summarized in [70] and discussed in length in [2, 68]. The parameter b is then estimated from eq. (72) as

$$\int_{\tau_{min}}^{\tau_{max}} \frac{b}{\tau} d\tau = b \ln \left(\frac{\tau_{max}}{\tau_{min}} \right) = 1 \Rightarrow b = \frac{1}{\ln(\tau_{max}/\tau_{min})} , \tag{75}$$

and the substitutions in eq. (73) yield that the power spectrum density of the noise in the base current due to superposition of Lorentzian components in large-area BJT is

$$S_{I_B} = \left(\frac{q W_J}{\varphi_t \varepsilon_{Si}} \right)^2 \frac{a_{avg}^2}{\ln(10)} \frac{N_{tA}}{\log(\tau_{max}/\tau_{min})} \frac{1}{f} \frac{I_B^2}{A_E} = \left(\frac{q W_J}{\varphi_t \varepsilon_{Si}} \right)^2 \frac{a_{avg}^2}{\ln(10)} \frac{N_{tAdec}}{f} \frac{I_B^2}{A_E} \propto \begin{cases} 1/f \\ 1/A_E \\ I_B^2 \end{cases} , \tag{76}$$

while in small-area BJT, the Lorentzian components are accompanied with RTS noise in time domain with an amplitude of

$$\Delta I_B = a_{avg} \left(\frac{q W_J}{\varphi_t \varepsilon_{Si}} \right) \frac{I_B}{A_E} . \tag{77}$$

From eq. (76), it is clear that the superposition of Lorentzian components from individual trapping or generation-recombination centers reassembles that the low-frequency noise, on average, is with 1/f power spectrum density, having magnitude inversely proportional to the area and quadratic function of the base current, as it has been stated above. The model in [17, 67] defines the quantity N_{tAdec} as areal density of traps per frequency decade, which can be experimentally obtained from large-area BJT, using

$$N_{tAdec} = \frac{N_{tA}}{\log(\tau_{max}/\tau_{min})} = A_E \frac{S_{I_B}}{I_B^2} f\left(\frac{\varphi_t \varepsilon_{Si}}{qW_J}\right)^2 \frac{\ln(10)}{a_{avg}^2} = \frac{(A_E \times K_F)}{0.43 \times a_{avg}^2}\left(\frac{\varphi_t \varepsilon_{Si}}{qW_J}\right)^2 , \quad (78)$$

and varying the coupling coefficient a_{avg} so that the number of Lorentzian components in smaller-area BJT matches the number $A_E \times N_{tAdec} \times \log(f_{max}/f_{min})$, and the measured RTS amplitudes ΔI_B agree with eq. (77) in sub-micron area BJT. Our experience implies that $a_{avg} \sim 0.07$ is a reasonable initial value, when using the barrier capacitance of the junction in eqs. (77) and (78). If one includes the diffusion capacitance, then the values for capacitance and coupling coefficient should be scaled up about one decade, but their ratio will be approximately the same, and therefore, ΔI_B and N_{tAdec} will remain unchanged. The areal density of traps per frequency decade N_{tAdec} is used also in statistical model for MOS transistors in [72].

Here we give some numbers relevant to the data in Figure 9 in order to provide impression for the order of magnitudes. $N_d = 10^{20}$ cm^{-2}, $N_a = 1.5 \times 10^{18}$ cm^{-2}, $V_{bi} = 1.07$ V, $V_{BE} = 0.81$ V, $W_J = 15$ nm, $A_E \times K_F = 4 \times 10^{-9}$ μm^2, coupling $a_{avg} = 0.07$, $N_{tAdec} = 2.4 \times 10^8$ cm^{-2}dec^{-1}, $\log(\tau_{max}/\tau_{min}) = 10$ dec, $N_{tA} = 2.4 \times 10^9$ cm^{-2}, $N_{tAeV} = N_{tA}/(0.026\text{ eV}) = 9.2 \times 10^{10}$ cm^{-2}eV^{-1}, since the charge only in the traps within 26meV around the Fermi level fluctuates [4, 69]. For the measurement frequency range of $\log(f_{max}/f_{min}) = 4$ decades, one can have 1.5, 6 and 23 Lorentzian components for devices of emitter area $A_E = 0.16$, 0.64 and 2.4 μm^2, respectively, but these components are distinguishable only in the smallest device in Figure 9, since one needs about one frequency decade separation between individual components in order to see them clearly in the spectrum. The values above are obtained from new analysis of the data. These values are within the order of magnitudes reported in [17], but they differ up to a half decade (perhaps because the aforementioned multiplier 4 in eq. (68) has to be reduced to 1 according to Fermi statistics [4, 68, 69]), while the measured data are the same. This indicates that there is an inherent characterization uncertainty for the low-frequency noise, which is normally in the range of several dB, and this is discussed next.

III.4. Measurement and characterization uncertainty – experimental accuracy, fitting and averaging

The last reason for the scattering in the data for the normalized noise in Figure 2, which we want to address, is that the measurement and characterization of low-frequency noise have inherent uncertainty. The sources of this uncertainty are several and include instrument inaccuracy, deviations from the ideal 1/f slope in the spectra, use of different characterization methods; and all these interfere with the experience and preferences of the individual researcher. We shall discuss now several difficulties that one has to overcome in experimental works.

III.4.1. Accuracy and the smoothness of the measured spectra

The first problem is the accuracy and the smoothness of the measured spectra. The measurement setup usually uses DC bias sources, low-noise preamplifier and spectrum analyzer. The bias sources are normally of good accuracy, with error not more than 1%, which causes less than 0.1dB uncertainty for normalized noise. However, battery bias (or filter) is required to ensure low level of noise floor in some experiments; and wire-wound potentiometers and resistors are used to provide the desired bias current or voltage. The DC accuracy of such setup usually is in the range 5%-10%, which causes up to about 1dB uncertainty for normalized noise. The next source of inaccuracy is the gain in the preamplifier and its finite impedance. At medium frequency of 100Hz, the gain in a typical low-noise preamplifier has inaccuracy of about 1%-3%, but at low frequency (1Hz) the blocking capacitor at the input of voltage amplifiers causes roll-off and the impedance of the amplifier together with its input cable is a problem in measurements of devices at low currents (nA) even in the range of 1kHz. The right choice in such case is a current preamplifier, but the DC current has to pass through the amplifier, no blocking of

this current is possible, resulting in decreased sensitivity of the setup in some cases, especially at high DC currents (>1mA). Overall, even after careful preparation of the measurement setup, the error associated with the preamplifier and its front-end is in the range 3%-5%, which adds another 0.4dB uncertainty for the normalized noise. The error from the spectrum analyzers is usually low, within 1%, but it requires that the measurement range is chosen accordingly. However, in some cases such as RTS noise with small or high duty cycle and occasional spikes, the range has to be increased to avoid saturation. Practically, the error from spectrum analyzer is about 3%, which is another 0.25dB uncertainty for the normalized noise. Adding all the errors, the instruments usually cause 2dB uncertainty in low-frequency noise experiments, in which 3-6 decades for the DC biasing currents are considered.

The smoothness of the measured spectra is also a problem in low-frequency noise experiments. At the low-frequency end of the spectrum, one usually desires a resolution of 1Hz or better. This, in turn, results in time window for signal capturing of 1 s or more. Then, since the noise is a random signal, the Fourier transformation of single captured record is with large scatter around the average, e.g. 10 dB or more. Therefore, many records, e.g. 100, need to be captured and averaged, which increases the measurement time to 10 and more minutes per spectrum, in order to obtain spectrum with scattering about 2dB. So, the measurement time becomes an issue for low-frequency noise, and it causes about 30% uncertainty for the normalized noise, if no additional method for averaging is used. One example for insufficient averaging during measurement is shown in Figure 10 at the tradeoff with measurement time. The scattering in this figure is about ½ decade and one will meet with difficulties to examine the normalized noise from this figure with accuracy better than 1/4 decade.

III.4.2. Deviations from the ideal 1/f slope in the spectra

The second problem related to the experimental accuracy for the normalized noise is the variability of the slope in the 1/f-like spectra. The variations are typically two: the steepness of the slope is different from 1/f; and the slope of the spectrum varies with the frequency, having humps owing to large Lorentzian components in the spectrum. There are physically based models with variable slope in noise spectrum, e.g. analytical in [2, 73], numerical in [68] and the aforementioned for superposition of Lorentzian components [17, 66, 67]. These models are based on deviation from uniform trap density or gaps in the uniform distribution, but a mature characterization technique based on any of those models is not available. There is no standard technique that resolves the problem of the variability of the slope in the 1/f-like spectra. However, some approaches are popular. These are choice of frequency of interest, fitting of formal, semi-empirical or Monte Carlo model, and post-measurement averaging.

Choice of frequency of interest

The approach of choosing a frequency of interest from the whole spectrum is suitable for volume noise characterizations, e.g. industrial tests. This approach minimizes the time of measurement and compresses the data into a small volume, so that comparisons between many samples and biasing regimes become feasible. The approach is particularly suitable when the goal is to examine the evolution of the noise level with bias or to obtain a statistics for the noise levels among devices on one or from several wafers. Due to its simplicity, the approach of choosing a frequency of interest is widely used, almost in every single publication, and it allowed verifying that the variation of the noise is a reciprocal function of the device area, as shown with one example from [74] in Figure 11 for MOS transistors. However, the approach of choosing a single frequency from the spectrum is vulnerable to large uncertainty owing to deviation of noise spectrum from 1/f slope, since the

approach inherently assumes 1/f slope when estimating the normalized noise at 1Hz, as required for the noise parameter K_F. A deviation δ=10-20% from 1/f slope causes error 10dB×δ also multiplied by the number of the frequency decades of extrapolation when referring the noise to 1Hz. A frequency 10Hz is typically chosen in experiments to minimize the measurement time, and therefore, 10-20% (0.4-0.8dB) error is easily introduced for the value of K_F. In addition, the randomness of the corner frequency f_c of Lorentzian noise in sub-micrometer area devices causes variation from high values for estimated K_F, if $f_c \approx 10$Hz, to low values, if f_c is one-two frequency decades apart from 10Hz. The uncertainty for the value of K_F in this case is very large – one and more decades, as one can deduce from Figure 11, and one can estimate meaningless value for K_F, if only one frequency point from the noise spectrum is used for the purpose. Therefore, one has to do statistical analysis of many data obtained by the approach of choosing a frequency of interest in order to obtain representative value for the normalized noise and the noise parameter K_F.

Fitting of formal, semi-empirical or Monte Carlo model

The other approach of fitting of formal, semi-empirical or Monte Carlo model to the measured noise spectra is more reliable and less vulnerable to variations in the slope of the spectrum, and therefore is used often in the research on low-frequency noise. The fitting model is chosen usually in the form of eq. (62), and it is enhanced with two more parameters, resulting in

$$S_I = S_{white} + \frac{K_F}{f^{B_F}} I^{A_F} + \sum_i \frac{B_i \tau_i}{1 + (2\pi f \tau_i)^2}, \tag{79}$$

where B_F=0.8…1.2 reflects the deviation of the flicker noise spectrum from the ideal 1/f slope, and A_F=1…2 reflects the bias dependence of the flicker noise. The approach of fitting a model, however, has several drawbacks. The first drawback is that it cannot be formalized into algorithm, because the selection of values for A_F, B_F and number of Lorentzian components is not unique, and the selection is also dependent on the pre-selected optimization criterion, the latter also chosen by preferences of the individual researcher. The second drawback, which follows from the first, is that the fitting procedure requires intensive human assistance in an interactive manner, which makes the approach very slow and vulnerable to human errors and individual preferences and skills. The last, but not the least, drawback is that the value and the unit for K_F varies with values of B_F and A_F. As the consequence, the quantitative comparison for the level of the flicker noise from different measurements and samples is almost impossible. Nevertheless, the first two terms of eq. (79) are implemented in the device models for circuit simulations and the evaluation of the corresponding parameters K_F, A_F and B_F helps in the design practice. Therefore, the approach of fitting a formal noise model for BJTs is taken in [32, 34, 37]. Interestingly, persons from industry are co-authoring these publications.

III.4.3. Data processing

The third problem related to the experimental accuracy for the normalized noise is the data processing from which K_F is evaluated. As mentioned above, the scattering in the measured noise spectra is about 2dB and the variations between individual values from different measurements and devices can be larger than 1 decade – see again Figure 11. In such situation, one has to perform post-measurement averaging in order to evaluate the noise and to obtain a representative value for the normalized noise and K_F.

The simplest way to perform averaging of a noise spectrum is to plot the spectrum in a log-log scale and to draw a line where the density of measured points is the highest. Form the intersect of the extrapolated line with the

axis for noise level at 1Hz, one gets the power spectrum density S(1Hz) at frequency 1 Hz, then the normalized noise and K_F are calculated. This approach is used very often, due to its simplicity, but it is vulnerable to human errors. An example that demonstrates the range of errors by manual fitting of data from noise measurements is given in Figure 12. The squares in this figure are the data reported by the authors in a numerical form in a table. The diamonds are the same data, but reported in a graphical form. The discrepancy between numerical and graphical data is apparent, although the manual fit of the data and the fitting of the two data series using least mean square method yield equations, which result in almost overlapping lines when plotted together. Looking closer at the numbers of the fitting equations, the values of the parameters vary about 5% both for the prefactor and exponential coefficient. So, one expects at least 0.2dB uncertainty for the normalized noise when the data from noise measurements are processed manually in a graphical form. This uncertainty could be much higher, if the data scatter, as in Figure 10 or in Figure 11.

Another approach that is equivalent to averaging is to analyze the noise in a bandwidth, rather than at individual frequencies. This is a standard procedure for resistors [75, 76, 77, 78], but rarely used for electronic devices, perhaps because the noise in a bandwidth is an integral measure, the frequency slope of the noise is essentially inaccessible from the noise in a bandwidth, and also the noise in a bandwidth is not implemented in device models, since it is expected to be a result from simulations. An attempt for modeling the noise in a bandwidth is presented in [72].

Numerical methods for averaging

The numerical methods for averaging of noise spectra can be divided into two groups – arithmetic (precisely, root-mean-square RMS) and geometric averaging of power spectrum density (PSD) (or derivates of PSD, such as $f \times PSD \times Area/I_{DC}^2$).

For several power spectrum densities S_i, with i=1, 2 …i_{max}, the arithmetic (RMS) averaging uses

$$S_{avg} = \frac{1}{i_{max}} \sum_{i}^{i_{max}} S_i \text{ , for arithmetic (RMS) mean,} \tag{80}$$

and

$$\sigma = \sqrt{\frac{1}{i_{max} - 1} \sum_{i}^{i_{max}} \left(S_i - S_{avg}\right)^2} \text{ , for arithmetic (RMS) standard deviation.} \tag{81}$$

The geometric averaging [28, 29, 34, 36, 79] uses logarithm of the power spectrum densities, given by

$$S_{dB,i} = 10dB \times \log\left(S_i\right), \text{ PSD in dB.} \tag{82}$$

Then,

$$S_{dB,avg} = \frac{1}{i_{max}} \sum_{i}^{i_{max}} S_{dB,i} \text{ , for geometric mean,} \tag{83}$$

and

$$\sigma_{dB} = \sqrt{\frac{1}{i_{max} - 1} \sum_{i}^{i_{max}} \left(S_{dB,i} - S_{dB,avg}\right)^2} \text{ , for geometric standard deviation.} \tag{84}$$

When the variations between individual spectra S_i are small, e.g. in large-area BJT, then both methods yield similar results, that is

$$S_{dB,avg} \approx 10dB \times \log\left(S_{avg}\right) \text{ and } S_{dB,avg} + \sigma_{dB} \approx 10dB \times \log\left(S_{avg} + \sigma\right). \quad (85)$$

However, if the variations between individual spectra S_i are spread over one decade or more, then the arithmetic averaging results in variation larger than the average [72], $\sigma > S_{avg}$, it becomes impractical (because the noise is attributed to σ rather than to S_{avg}), and the geometric averaging is more suitable for sub-micrometer area devices, since the distribution of noise variation tends to log-normal distribution [28, 29, 79, 80]. Therefore, many authors use geometric averaging [17, 28, 29, 34, 36, 66, 67, 79, 80, 81, 82]. Worth mentioning, it is empirically observed that the noise variations are better described by log-normal distribution, and substantial work is expected to explain the origin of this empirical observation [79, 80]. A reason is given later in section VIII.6. "**Consequences from statistical nature of LFN – distributions in spectra, techniques of averaging, data volume and coordinates, instrumentation**". Interestingly, assuming Poisson distribution of threading dislocations in strained silicon wafers, the geometric model of noise variation explains the data in [82].

However, one should be careful when using the equations for geometric averaging and modeling of noise, in order to preserve the physical consistence of figures of merit (FOM). For example, defining FOM for relative variation (in respect to mean level of noise), the arithmetic (RMS) averaging suggests [72, 79]

$$FOM_{ari} = \frac{\sigma}{S_{avg}}, \text{ for relative variation from arithmetic (RMS) averaging,} \quad (86)$$

The FOM for relative variation from geometric averaging is actually σ_{dB} [28, 79, 80, 81], because it follows from eq. (85) that [17, 66, 67]

$$FOM_{geo} \equiv \sigma_{dB} \Rightarrow \frac{\sigma}{S_{avg}} + 1 = 10^{\left(\sigma_{dB}/10dB\right)}, \text{ for relative variation from geometric averaging,} \quad (87)$$

whereas the $FOM = \sigma_{dB}/|S_{dB,avg}|$ as attempted in [33, 34, 36] is just a number that will violate the rules for physical units, if one tries to convert it to ratio of noise level. In the above equations, note again that σ is standard deviation and S_{avg} is average of noise S, the latter being a variance (usually per unit frequency) in principle.

To provide impression for the noise variation in BJTs, we present in Figure 13 several views of the few data available from literature. These data have been also obtained using geometric standard deviation according eq. (84). The shaded areas in the figure represent cases when the variation of the noise is larger than the average level, that is $\sigma > K_F$. The top-left plot in the figure shows the data vs. emitter area, as originally reported. One can see that the noise variation increases in small-area BJT and in sub-micron area BJT the variation is dominating over the average. The slope in this plot is somehow low, only -2.5dB/dec, but the bottom-left plot, obtained using eq. (85) or eq. (87), clearly shows that the absolute variations are a strong function of the emitter area, $\sigma \propto (A_E)^{-1.5}$, and taking into account that $K_F \propto (A_E)^{-1}$, then one observes the aforementioned increase of $(A_E)^{-0.5}$ in noise variation as function of decreasing device area, as predicted by Monte Carlo simulations [17, 66, 67] and analytically in [72].

More interesting, and to the best of our knowledge not explored explicitly yet, are the observations that one can make in the right-hand plots in Figure 13. The main observation is that the noise variation scales with the average noise. This is clear in the bottom-right plot, where the slope 3:2 suggests a trend $\sigma \propto (K_F)^{1.5}$ for the

absolute variation σ of the noise as function of the average noise K_F, and reassembles the corresponding slope -3:2 for the dependence $\sigma \propto (A_E)^{-1.5}$. Owing to the limited number of experimental data points, it is difficult to identify properly the factors that cause the relation between noise variation and average noise. For example, there is no unique line that can be drawn in the top-right plot. The overall behavior in Figure 13 suggests that $\sigma \propto K_F / \sqrt{A_E} \propto A_E^{-1.5}$, but the limited amount of experimental data does not allow parameterizing the dependence reliably. Nevertheless, for BJTs with emitter area $A_E < 0.3\mu m^2$ and $K_F > 3\times10^{-8}$, the noise variation appears to be more important than the average noise, since the noise variation affects the repeatability in the noise performance of the devices. Consequently, large discrepancy between simulations based only on average noise and the real noise performance of the individual device is expected for nowadays BJT. More discussions on averaging techniques are given in section VIII. "**Outlook for the LFN**" after eq.(469).

Summing all factors discussed above we estimate that the uncertainty of each data point in Figure 2 is about 3dB, owing to different measurement setups and techniques for model fitting and averaging. This uncertainty, perhaps, contributes significantly to the value of σ_{dB} that is shown in the figure. Nevertheless, we use the data for npn BJTs from Figure 2 as a benchmark in comparisons to other devices. When comparing the values for K_F, we use the data which are within and close to the interval $A_E \times K_F \approx 5.6\times10^{-9}\ \mu m^2 \pm 2\sigma_{dB}$. When comparing the values for input voltage noise S_V, the same reference data are referred as the base voltage noise using the relation $f \times A_E \times S_{V_B}(f) = A_E \times K_F \times (\varphi_t)^2$ from eq.(3), which applies when the noise in BJT is coupled from the emitter area to the diffusion currents, according to the discussions on eq. (36).

Comparative study of noise in npn and pnp BJTs

Past publications , for example [83], conclude after comparative studies on devices from complementary technologies (BiCMOS) that the noise in pnp BJTs is lower than the noise in their npn counterparts. We have searched for published data for low-frequency noise in silicon pnp BJT. The data are collected from [12, 32, 83], stored in numerical form in [84], and shown in Figure 14 together with data and trend for the noise in silicon npn BJTs. The information for noise in silicon pnp BJTs is less available in the last two decades, as compared to the data for silicon npn BJTs and the more recent interest on SiGe HBTs. In contrary to the above comparative studies, we did not found data for the noise in pnp BJT, which could confirm the conclusion of the comparative studies. This is illustrated in Figure 14. The open circles and gray lines are from Figure 2 and represent data and trend for noise coefficient K_F in npn BJTs. The solid symbols and black lines in Figure 14 represent data and trend for noise coefficient K_F in pnp BJTs, and they are above the trend for npn BJTs. Also, studies [34] on noise in complementary SiGe HBTs indicate that the noise in pnp transistors is higher than the noise in npn transistors. Interestingly, the data for pnp BJTs tends also to log-normal distribution, as shown in the insert of Figure 14. Nevertheless, the difference in noise coefficients K_F between npn and pnp transistors is within the range of scattering of the data for low-frequency noise, and it does not lend much confidence to make a general conclusion whether the noise is lower in npn or in pnp BJTs. Some differences in the bias dependence of the noise and for the impact of IFO in npn and pnp transistors are reported in [32], which makes the fair comparison of the noise in these transistors difficult.

IV. Noise in MOS transistors

The low-frequency noise in MOS transistors is widely studied because of several reasons. First, a physical model has been proposed in [85] long time ago and before any model for noise in BJT was accepted. Second, the MOS

technologies are the driving force in the downscaling of electronic devices in the last decades. Third, the low-frequency noise in MOS transistors is higher than the noise in BJTs, as one can see in Figure 15. In this figure, the data are for 134 nMOS transistors from [22, 47, 48, 49, 50, 51, 72, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123], and for 53 pMOS transistors from [47, 52, 86, 88, 95, 96, 97, 100, 102, 104, 105, 109, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132]. Many data points overlap in Figure 15, the lowest levels for noise in MOS transistors is higher than the highest levels of noise in BJT, and the data scatter about 3.5 decades for MOS transistors, while the data scatter less than 2 decades for BJT. Owing to these observations in the figure, it is clear that the control of the noise in MOS transistors is more difficult than it is in BJT. The main reason is that the operation of the MOS transistor is governed by a surface current transport, which is exposed to interface phenomena at the semiconductor-gate dielectric interface, while in BJT the current transport is mostly in the bulk of the semiconductor and it is "less" sensitive to the surface of the semiconductor. The consequence is that the current research on low-frequency noise in MOS transistors is focused on the fabrication of the gate stack at the semiconductor surface and although important and studied, the variation of the noise with the channel size is not dominant in the scope of many publications. Therefore, many researchers use samples of "standard" gate area, e.g. around 2-3μm^2, or 10-20 μm^2, but different gate stacks, and unique trend in Figure 15 cannot be placed accurately. Nevertheless, the product $W \times L \times S_{V_G}$ for MOS transistors tends to log-normal distribution, as shown in the insert of Figure 15, and the mean in this distribution is about 2.5 decades above the average noise in BJT, which is in agreement with past ITRS predictions about the difference in the noise between MOS and bipolar transistors [3] – see again Figure 1a. Interestingly, when looking at the distributions for the product $W \times L \times S_{V_G}$ separately for nMOS and pMOS transistors in Figure 16, the average noise in pMOS transistors is about 3dB (2 times) lower than the noise in nMOS transistors, but at the same time, the distribution of the noise values is broader for pMOS transistors, which practically does not lend much confidence to conclude which type of MOS transistors is with lower levels of noise. Comparisons for noise in pMOS and nMOS transistors report opposite observations – noise in pMOS transistors is lower [95] and higher [83, 98] than in nMOS transistors. Having the above observations in mind, and the data for noise available, we begin the discussion on the physics and models underneath these observations and the scattered data of low-frequency noise in MOS transistors.

IV.1. Models and predictability

It is widely accepted that the low-frequency noise in the drain current of the MOS transistor is coupled by the charge and mobility fluctuations related to the interface between semiconductor and gate dielectric. The approach of intrinsic and coupled noise allows comparing the low-frequency noise of MOS transistors to other devices, e.g. bipolar junction transistors, as it is discussed in the previous section. So, with the obvious for MOS transistor notations, we rewrite eqs. (11) and (12) as

$$\frac{S_{I_D}}{{I_D}^2} = K\left(\frac{g_m}{I_D}\right)^2 S_{V_G} + \frac{\alpha_H}{n_{eff} f}, \tag{88}$$

for the purpose of comparing input referred voltages, and

$$\frac{S_{I_D}}{{I_D}^2} = K\left(\frac{g_m}{I_D}\right)^2 \left(\frac{q}{C_{ox}}\right)^2 S_{N_t} + \frac{\alpha_H}{n_{eff} f}, \tag{89}$$

for the purpose of analyses of charge trapping at interface between semiconductor and gate dielectric. Here, I_D [A] is the DC drain current, S_{I_D} [A^2/Hz] is the power spectrum density (PSD) of the output drain current noise.

For the coupled portion of noise, popular as noise from number fluctuation Δn, the relevant quantities are transconductance $g_m=\partial I_D/\partial V_G$ [A/V=S], PSD of input referred (gate) voltage noise S_{V_G} [V^2/Hz], gate capacitance per unit area C_{ox} [F/cm^2], PSD of trapped charge per unit area S_{N_t} [cm^{-4}/Hz], and K is coupling parameter which depends on many factors, including bias, according to different physical models, but practically it is taken K≈1.

IV.1.1. Intrinsic noise (mobility fluctuation)

For the intrinsic portion of noise, popular as noise from mobility fluctuation $\Delta\mu$, α_H is the Hooge parameter and n_{eff} is effective number of carriers in the channel [9]. The effective number $n_{eff} \leq n$ takes into account the non-uniform distribution of the total number n of carriers in the channel, when the drain voltage is not zero, and can be deduced by using eq. (13), but the procedure of evaluation is complicated, and one usually takes $n_{eff} \approx n$ when investigating the low-frequency noise in MOS transistors in strong inversion regime at gate biasing $V_{GS}>V_T$ above the threshold voltage V_T, especially when the transistor operates in linear (ohmic) mode at low drain bias V_{DS}~50mV<V_{GS}-V_T. Evidently, even before assuming any physics for the noise, quite a lot approximations and assumptions can be taken differently by different researchers in the equations for the low-frequency noise in MOS transistor, and this causes large spread in the values reported for the parameters related to the noise, and sometimes even to controversial conclusions. We illustrate this with one example for the intrinsic component of the noise.

Assume that $V_{GS}-V_T \geq 0.1V$ and $0 \leq V_{DS} \leq V_{GS}-V_T-0.05V$. Such biasing of MOS transistor is usually regarded as operation in ohmic mode, and one takes $n=WLC_{ox}(V_{GS}-V_T)/q$, neglecting that the charge concentration is lower at the drain side of the channel. Assume also that the intrinsic (Hooge) low-frequency noise is dominant (e.g. pMOS transistor), so that the Hooge parameter is estimated from eq. (88) as

$$\alpha_H @ n = f\frac{S_{I_D}}{{I_D}^2}n = f\frac{S_{I_D}}{{I_D}^2}\frac{WLC_{ox}\left(V_{GS}-V_T\right)}{q}, \text{ at approximation } n=WLC_{ox}(V_{GS}-V_T)/q. \quad (90)$$

Now, let us increase the "accuracy" of the calculation of the total number of carriers, assuming gradual approximation for charge concentration. The average number of carriers n_{avg} is the mean value of the number of carriers at source and drain sides, $n_{avg}=(n_{source}+n_{drain})/2$, and n_{avg} becomes

$$n_{avg} = \frac{1}{2}\left[\frac{WLC_{ox}\left(V_{GS}-V_T\right)}{q}+\frac{WLC_{ox}\left(V_{GS}-V_T-V_{DS}\right)}{q}\right] = \frac{WLC_{ox}}{q}\left(V_{GS}-V_T-\frac{V_{DS}}{2}\right). \quad (91)$$

Accordingly, the Hooge parameter is estimated as

$$\alpha_H @ n_{avg} = f\frac{S_{I_D}}{{I_D}^2}n_{avg} = f\frac{S_{I_D}}{{I_D}^2}\frac{WLC_{ox}}{q}\left(V_{GS}-V_T-\frac{V_{DS}}{2}\right), \text{ at } n_{avg}=(n_{source}+n_{drain})/2. \quad (92)$$

Let us increase further the "accuracy" of the calculation using the integral for current crowding given by eq. (13), rewritten for 1D case, by assuming constant current density J, charge sheet approximation (the thickness of inversion layer T≡Dirac function, yet not accurate for sub-100nm MOS transistors) and thus WTJ=constant.

$$S_{norm} = \frac{S_{I_D}}{{I_D}^2} = \frac{\int_0^L \frac{\alpha_H}{n'(x)\cdot f}(WTJ)^4\,dx}{\left(\int_0^L (WTJ)^2\,dx\right)^2} = \frac{\alpha_H}{fL^2}\int_0^L \frac{dx}{n'(x)}\,, \tag{93}$$

where the coordinate x=0…L is along the channel length, and

$$n'(x) = \frac{\partial n}{\partial x} = \frac{WC_{ox}}{q}\left(V_{GS} - V_T - V_{DS}\frac{x}{L}\right). \tag{94}$$

Then,

$$\int_0^L \frac{dx}{n'(x)} = \int_0^L \frac{dx}{\frac{WC_{ox}}{q}\left(V_{GS} - V_T - V_{DS}\frac{x}{L}\right)} = \frac{qL}{WC_{ox}V_{DS}}\ln\left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}}\right) \tag{95}$$

So, one gets the following expression for the Hooge parameter

$$\alpha_H @ n_{eff} = f\frac{S_{I_D}}{{I_D}^2}\frac{WLC_{ox}V_{DS}}{q\ln\left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}}\right)}, \text{ with } \quad n_{eff} = \frac{WLC_{ox}V_{DS}}{q\ln\left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}}\right)}, \tag{96}$$

where the effective number of carriers in the channel, n_{eff} [9], is lower than the above approximations $n=WLC_{ox}(V_{GS}-V_T)/q$ and $n_{avg}=WLC_{ox}(V_{GS}-V_T-V_{DS}/2)/q$. Provided that one processes the same data from one sample, but using different assumptions for number of carriers, Figure 17 illustrates possible discrepancies in the values estimated for Hooge parameter, since the value of α_H is proportional to the estimated number of carriers in eqs. (90), (92) and (96).

One can see from Figure 17 that 20% to 50% variations (1dB to 3dB) in the estimated values for α_H can be easily introduced by changing only the characterization model. Similar vulnerability is discussed in [133] for the coupled part of the low-frequency noise in MOS transistors (popular also as number fluctuation Δn) when the drain bias is not low, and brings the transistor in saturation mode, resulting in doubling the noise levels (3dB increase). Thus, ½ decade in the scattering of the data in Figure 15 we attribute to the use of different characterization models and procedures, since no standard method is currently accepted for MOS transistors.

IV.1.2. Coupled noise component (number fluctuation)

The coupled noise component in MOS transistors is widely studied in terms of trapping at semiconductor-oxide interface states [134] or charge trapping in gate oxide, as originally suggested in [85]. These approaches are known as random walk and tunneling Δn models for 1/f noise in MOS transistors, respectively. They are illustrated in Figure 18 and discussed below.

Another possibility for charge trapping and associated coupled noise component is when the traps are in the semiconductor [135], either in the inversion or/and in the depletion layer of the MOS transistor channel. The exploration of this possibility is feasible, when Lorentzian noise spectra are present, but random-telegraph signal (RTS) noise waveforms are not observed, and the noise is attributed to traps with particular energy, in contrary

to the expectation for uniformly distributed in energy and space traps in gate dielectrics. This possibility of trapping in the semiconductor of the MOS transistors is rarely addressed in the literature, since the interface effects are dominant in the MOS transistors, and not discussed below.

Random walk model

The random walk model [134] assumes that a charge carrier is captured at an interface state and after a time τ it moves to a neighboring interface state. The mean free path of the charges "walking" at the semiconductor-insulator interface is taken as atomic distance $l \approx 0.2$nm and the cross-section of charge trapping is taken $\sigma=10^{-16}$ $cm^2=(0.1nm)^2$. For charge sheet (2D) approximation of the inversion layer in MOS channel, the distribution of time constants is

$$g(\tau) = \frac{\sqrt{\sigma}}{2\pi l}\frac{1}{\tau} \approx \frac{0.1}{\tau}. \tag{97}$$

For the random walk, it is derived in [134] from superposition of Lorentzian spectra (see eqs. (73), (74) and (75) earlier) that the gate referred voltage noise in eq. (88) is

$$S_{V_G} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{1}{WL}\frac{\sqrt{\sigma}}{2\pi l} kTD_{it} \approx \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT}{WL} 0.1D_{it}, \text{ for random walk,} \tag{98}$$

where the interface states are assumed to be distributed uniformly with density $D_{it}[cm^{-2}eV^{-1}]$=constant.

Tunneling model

The tunneling Δn model for 1/f noise in MOS transistors, which was originally suggested in [85], has been followed up widely [47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 72, 74, 80, 82, 83, 87, 89, 91, 95, 101, 103, 106, 107, 110, 124, 126, 127, 128, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145]. It assumes that some charge carriers are trapped in the depth of the gate oxide. The tunneling probability is exponentially decaying function of the distance x_{ti} from semiconductor-insulator interface to the position of the trap in the oxide. Therefore, the rate $1/\tau$ of tunneling events is

$$\frac{1}{\tau} = \frac{1}{\tau_o}\exp\left(-\frac{x_{ti}}{\lambda}\right), \tag{99}$$

where $1/\tau_o$ is attempt rate (frequency) taken by different assumptions in the range of from $10^7 s^{-1}$ [134] to $10^{10} s^{-1}$ [49, 55, 140], and λ is tunneling (electron wave) attenuation distance, given by eq. (23). The values for λ range between 0.06 nm and 0.22 nm for different semiconductor-insulator pairs, electrons and holes – see Table 2, because λ is function of the product of effective mass of the carrier in the insulator and the energy offset between the bands in semiconductor and insulator, according to Wentzel-Kramer-Brillouin (WKB) approximation. For $Si\text{-}SiO_2$ and electrons, one usually takes λ=0.1nm – see [146], for example.

At an assumption for uniformly distributed traps in the oxide depth, $\partial N_t/\partial x_{ti}$=constant, as shown in many publications, for example in [70], it follows from eq. (99) that $\ln(\tau) \propto x_{ti}/\lambda \Rightarrow \partial\tau/\tau = \partial x_{ti}/\lambda$, and the distribution of the traps vs. their time constants becomes

$$g(\tau) = \frac{\partial N_t}{\partial \tau} = \frac{\partial N_t}{\partial x_{ti}}\frac{\partial x_{ti}}{\partial \tau} = \text{constant}\frac{\lambda}{\tau} = \frac{b}{\tau}. \tag{100}$$

Then, the superposition of Lorentzian spectra of the individual traps in the oxide produces 1/f noise – see again the procedure by eqs. (72), (73), (74) and (75). In this way (see [28] for example), the gate referred voltage noise in eq. (88) becomes as

$$S_{V_G} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT}{WL}\lambda N_t, \quad \text{for tunneling in gate insulator,} \tag{101}$$

where the traps in the oxide are assumed to be distributed uniformly with density $N_t[cm^{-3}eV^{-1}]$=constant.

Comparison of random walk and tunneling models

Comparing eqs. (98) and (101), one sees that the random walk and tunneling models converge, although the physical assumptions behind these models are different. The reason is that D_{it} in the random walk model and N_t in the tunneling model are assumed uniformly distributed both in space and energy, in order to obtain $1/\tau$ distribution for the time constants of the traps, and from this – 1/f noise. This is discussed in more detail in [28, 147], where also is observed that D_{it} from the low-frequency noise technique converges with charge pumping technique in the frequency range 10kHz to 100kHz that is accessible by both techniques. The deduction in [147, 148] is that the $1/\tau$ distribution arises from traps' effective cross-section σ_{eff} apparent at semiconductor-dielectric interface, and σ_{eff} is exponential function from both activation energy E_B of interface states and distance x_{ti} from semiconductor-insulator interface, given by

$$\frac{1}{\tau} \propto \sigma_{eff} \propto \exp\left(-\frac{E_B}{kT}\right)\exp\left(-\frac{x_{ti}}{\lambda}\right). \tag{102}$$

Accordingly,

$$\partial[\ln(\tau)] = \frac{\partial\tau}{\tau} = \frac{\partial E_B}{kT} + \frac{\partial x_{ti}}{\lambda}, \tag{103}$$

and following the procedure in eq. (100), one sees that the $1/\tau$ distribution can be achieved either or both assuming uniform distribution for the activation energy at the interface, $\partial D_{it}/\partial E_B$=constant, or uniform spatial distribution of the traps in the depth of the gate insulator, $\partial N_t/\partial x_{ti}$=constant. Figure 19 illustrates how the uniform distributions in trap energy E_B=0.2...0.6eV and tunneling distance x_{ti}=0.8…2.3 nm produce 1/f noise. Therefore, one can quite arbitrary attribute the 1/f noise to interface states or oxide traps, and this demonstrates the convergence between random walk and tunneling models for the gate referred 1/f noise voltage S_{V_G} in MOS transistors. In fact, both trapping mechanisms can superimpose, one can combine tunneling and random walk models, and we write for S_{V_G} in eq. (88) that

$$S_{V_G} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT}{WL}\left(\lambda N_t + 0.1D_{it}\right). \tag{104}$$

Rewritten in terms of charge fluctuation, S_{N_t} in eq. (89) is then given by

$$S_{N_t} = \frac{1}{f}\frac{kT}{WL}\left(\lambda N_t + 0.1D_{it}\right). \tag{105}$$

One can assume even a two-step process: first carrier capture in interface state and then random walk and tunneling. Perhaps, this is the reason why the range for τ_o in eq. (99) is taken with values in several decades by

different researchers. There are also other issues with the model for S_{V_G} in eq. (104), which we will discuss below, but before that, we will inspect the variation of the data in Figure 15 caused by the term $(C_{ox})^2$ in the denominator of eq. (104).

Effect of the oxide capacitance, C_{ox}

For the purpose, we plot the ITRS [3] figure of merit for S_{V_G} ($FOM_{S_{VG}}$) against the equivalent oxide thickness EOT of gate insulator, with $EOT=\varepsilon_{SiO_2}/C_{ox}$, where $\varepsilon_{SiO_2}=3.45\times10^{-13}$ F/cm is the permittivity of SiO_2 and C_{ox} is the gate insulator capacitance per unit area. From eq. (104), $FOM_{S_{VG}}$ is given by

$$FOM_{S_{VG}} = \left[1\mu m^2 S_{V_G} \middle| (V_{GS}-V_T) \approx 0.1V \right] \frac{f}{1Hz} \frac{WL}{1\mu m^2} = \frac{q^2 kT(\lambda N_t + 0.1 D_{it})}{C_{ox}^2} \propto \frac{1}{C_{ox}^2} \propto EOT^2, \quad (106)$$

and $FOM_{S_{VG}}$ represents input referred gate voltage PSD at 1Hz of MOS transistor of unit gate area $(WL)_{ref}=1\mu m^2$ at low gate overdrive voltage $(V_{GS}-V_T)\approx0.1V$. According to scaling rules, one expects that $FOM_{S_{VG}}\propto(C_{ox})^{-2}\propto EOT^2$, and the deviation from this proportionality can be attributed to material and technology variations via N_t or D_{it}.

We have collected data from [22, 47, 48, 49, 50, 51, 72, 82, 86, 87, 88, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 149] for nMOS transistors and from [47, 52, 86, 88, 94, 95, 96, 97, 100, 102, 104, 105, 109, 124, 125, 126, 128] for pMOS transistors, and plot the result for $FOM_{S_{VG}}$ in Figure 20. The data are stored in numerical form in [150]. Many data points overlap in Figure 20 and the scattering in the data is large – about 4 decades. No trend can be estimated directly form the data points in the figure. The trend line for the average and variations $\pm\sigma_{dB}$ in this figure are obtained from statistical analysis of the data according to $FOM_{S_{VG}}$ of eq. (106), following the procedure of eqs. (83), (84) and (85), and using geometric averaging of the quantity X_{dB}, where X_{dB} is given by

$$10^{X_{dB}/10dB} = \frac{FOM_{S_{VG}}}{EOT^2} = \frac{WLS_{V_G}(1Hz)}{EOT^2}, \ V^2/Hz. \quad (107)$$

The distributions obtained from the statistical analysis are shown in Figure 21. The histogram in Figure 21a illustrates that the overall distribution tends to log-normal distribution with an average 6×10^{-5} V^2/Hz and standard deviation $\sigma_{dB}\approx10$ dB (1 decade). These values are obtained when using 106+38 data points shown in Figure 20 for nMOS+pMOS silicon transistors, 13+12 data points shown later in Figure 22 for MOS transistors with strained lattice and germanium content in the channel, and 16+5 data points from ITRS predictions made for the period 2006-2020 [3] for RF and Analog MOS.

The trend line "Average" and the margins $\pm2\sigma_{dB}$ for MOS transistors in Figure 20 correspond to the above stated values. For a comparison, the data and the trend for BJT vs. IFO thickness from Figure 4 are referred as input voltage noise, using $A_E\times S_{V_B}(1Hz)=A_E\times K_F\times(\varphi_t)^2/1Hz$ according to eq. (3), and added to Figure 20. Two speculative observations can be made in Figure 20. One is that the noise in MOS transistors is, on average, what was expected in past ITRS predictions [3]. The second is that the noise levels (and LFN models) in MOS transistors and BJT converge. The crossover between the noise in these transistors corresponds to the minimum EOT and the maximum IFO, which is about 0.8nm equivalent thickness of SiO_2. Although the crossover is

apparent at the boundary between conduction and insulation of thin dielectrics, these observations are speculative, because the physical assumptions in the models are different and the scattering of the data is large. For example, the distributions of noise in nMOS and pMOS transistors have different statistical modes, two decades apart from each other, as shown in Figure 21b, while at the same time, the data for these devices scatter in the same range of 5 decades in a very similar manner. The latter implies that the type of conduction, electron or hole, in MOS transistors might not be the major concern for the noise in these devices in the same way as the channel area and dielectric thickness are unable to capture the noise levels in different MOS transistors. So, the major factor for the LFN level in MOS transistors appears to be the fabrication of the gate stack, rather than the thickness of the gate insulator; and the current research is mostly focused on this issue. However, this research is quite empirical, since no model is available to explain physically 3 and more decades of variations. For example, there are reports with oxide trap density $N_t > 10^{20}$ $cm^{-3}eV^{-1}$, which should result into degeneration of insulator band-gap, since, for example, the band gap of HfO_2 is about 6eV (Table 2), then the oxide trap concentration becomes in the range of $6eV \times N_t \sim 10^{21}$ cm^{-3}, while the atomic density is in the range 10^{22} cm^{-3}.

IV.1.3. Empirical factors that impact the scattering of data in noise measurements

To illustrate the empirical nature of the current research on low-frequency noise in MOS transistors we use Figure 20 as the template on which we project the impact of several factors. The impact of fabrication factors is usually investigated using a bottom-up approach by varying the factors in otherwise identical or similar structures. Since we have the data from many publications collected, we present here the alternative top-down case study on the impact of the factors on the values of the noise level, by intentionally breaking the link between the samples and formally using only the scaling rule $WLS_{V_G} \propto EOT^2$ of eq. (106), in order to see the overall picture for the low-frequency noise in MOS transistors, which is with large scattering in the values, as discussed above. We pick three factors, for which the bottom-up studies made strong cases. These are: noise in pMOS transistors is lower than in nMOS; noise in MOS transistors with metal gates is lower than in transistors with polysilicon gates; strain engineering (mechanical stress) increases the noise, whereas noise in SiGe transistors is lower. We will add one more factor – the fabricator. Our top-down study on published data is illustrated in Figure 22.

(i) pMOS versus NMOS.

Several studies of CMOS technology nodes suggested that pMOS transistors are less noisy than their nMOS counterparts from the same technology [95]. This has been observed earlier for S_{V_G} at lower gate bias voltages [91], while at higher bias, the noise levels were similar. In contrast, researchers observe the opposite situation – pMOS transistors are noisier [83, 98]. We collected published data in [150] and put the data together in the top-left plot of Figure 22. From this plot we observe only that the data for pMOS and nMOS transistor scatter in the same range, and the histograms earlier in Figure 21b do not argue much in favor of any of these transistors.

(ii) Gate material effect.

In an effort to reduce EOT, past investigations attempted metal and metal nitrides [50, 52, 57, 94, 99, 100, 109, 126, 128] and metal silicides [96, 106, 108, 128] for the gate conductor, in order to overcome the unavoidable depletion in polysilicon gates [151, 152]. Cumulative experimental research also indicates that there might be noise due to tunneling from polysilicon gate into high-k dielectrics or charge trapping and defects at this interface [97, 99, 126, 153, 154], or remote scattering caused by Coulomb coupling between gate trap and MOS

channel (oxide is thin and single excess charge at gate side of the oxide is “visible” as a local, e.g. 2-5 nm in diameter, potential “bump” in conduction channel on the other side of oxide). Therefore, the metal gate would probably reduce these effects (not yet theoretically explored for low-frequency noise, only qualitatively used in the so called N_t-profiling with LFN that we will discuss later in section IV.2). So, from the collected in [150] published data for noise in MOS transistors with metal and metal nitride gates [50, 52, 57, 94, 99, 100, 109, 126, 128] we have drawn the top-right plot in Figure 22, and we observe that the metal gate MOS transistors correspond to polysilicon gate transistors with higher noise, in contrast to the expectations deduced in the individual publications for metal gates.

(iii) Mobility degradation.

Another issue related to thin oxides and corresponding high electric fields in sub-100nm MOS transistors is the mobility degradation. Two approaches are taken to remedy the problem – introduction of strain (mechanical stress) in the surface lattice of the MOS transistor channel or addition of germanium Ge in the channel. The effect of the strain on the low-frequency noise was addressed in [82, 108, 126, 127, 155], since it is expected an increase of lattice defects and, therefore, increase of the noise levels. The addition of Ge in the channel of pMOS transistors is expected to increase the mobility [52, 57, 109], thus reducing the scattering and the noise. From the collected in [150] published data for noise in MOS transistors with strained lattice and Ge content in the channel we have drawn the bottom-left plot in Figure 22. We see in this plot that the noise is virtually unaffected in strained MOS transistors, as observed for nMOS transistors in [108, 126, 127, 155], since the diamonds are in the bottom half of the distribution, while an occurrence of Ge in the channel (or in the drain or source regions [126]) results in increase of the noise, since the triangles are in the upper half of the noise distribution.

(iv) The fabricator effect.

At this point we would like to make a comment. The model of eq. (104) is based on certain assumptions in a simple structure of Si-SiO_2 interface, while stacking, stretching and mixing of the materials in nm-scale is neglected in the model and violates the assumptions for uniformity and large statistical populations. The extrapolation of the model in empirical studies, as we did above using the scaling rule, is questionable and resulted in the large scattering of data from research samples, as shown with squares in the bottom-right plot of Figure 22. Fortunately, the constructive conservatism in commercial technologies prevents from the randomness in the empirical research and keeps the noise in the desired lower half of the distribution, as illustrated with diamonds in this plot. However, this conservatism results in a slowdown of device scaling [152], as was predicted in the past [3], because the costs of modifications for using metal gates are high and the predictability and repeatability of multilayer gate stacks are not certain, since no scalable model is currently available for these gate stacks. Thus, the risk is mostly moved in the research, and the commercial fabricators implement the modifications gradually and very carefully, as one can see in [152] for Fujitsu, for example, although the main reason is not the low-frequency noise, actually.

IV.1.4. Modeling factors to interpret the scattering of data in noise measurements

It is worth mentioning the conditions and assumptions at which the model for S_{V_G} in eq. (104) is derived. These are discussed in [56] and listed below.

i. Charge carriers only from surface channel of the MOS transistor are trapped in oxide or at semiconductor-dielectric interface. Bulk traps in semiconductor, trapping from gate or at any other interface (such as buried oxide in SOI) are neglected at assumption that they are “far” from the

conduction channel.

ii. The additional scatter around the trapped charge is small. This is remedied in the models with correlated mobility fluctuation.

iii. Charge exchange between channel, interface states and oxide traps occurs only at Fermi level, and other transitions, e.g. interchange with gate, bulk, between different energy levels, are neglected. This allows for using the Fermi-Dirac statistics in Shockley–Read–Hall process.

iv. The distributions are uniform and the populations of traps and carriers are large enough. This means that the traps are uniformly distributed both in space and energy, the conduction channel in MOS transistor is uniform (e.g. V_{DS} is low), the mobility and electric field are constant along the channel, they do not fluctuate, and the device sizes are large enough to have the approximation with averages valid.

v. The energy barrier for tunneling is constant and it is not affected by the electric field. The barrier is at the semiconductor-dielectric interface and the channel carriers are at this interface too. (For random walk model, the distance from carrier to interface state is constant)

vi. The charge captured in the oxide is approximately at the semiconductor-dielectric interface. The distance x_{ti} from interface, where the charge is trapped, is negligible as compared to the thickness t_{ox} of gate dielectric.

At these conditions, the trapping of one charge carrier causes an equivalent change in the gate voltage, with a magnitude given by

$$\delta_{V_G} = \frac{q}{WLC_{ox}} = \delta_{V_{FB}}, \tag{108}$$

and the fluctuation in the number and rate of charge trapping can be referred to the gate and attributed to noise in flat-band voltage V_{FB}, according to [156]

$$S_{V_G} \approx S_{V_{FB}} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT}{WL}\left(\lambda N_t + 0.1 D_{it}\right) \text{ at conditions i, ii, iii, iv, v, and vi satisfied,} \tag{109}$$

since the threshold voltage of MOS transistor is a linear function of V_{FB}. In modern MOS transistors from sub-100nm technology nodes, however, none of the above conditions is strictly valid anymore. Some of the deviations from these conditions are resolved, other – still not. This is now discussed.

Issues with condition i.

The oxides are very thin in modern MOS transistors and the tunneling of carriers from both sides of the oxide is observed [97, 99, 126, 153, 154]. The channel carriers are reasonable to be assumed as the source of charge trapping in the oxide, resulting in a capture rate of

$$\frac{1}{\tau_c} = \frac{1}{\tau_o}\exp\left(-\frac{x_{ti}}{\lambda}\right). \tag{110}$$

However, as shown in Figure 23a, the emission of the trapped charge can be arbitrary back to the channel or out to the gate depending on several factors. By taking the distance as the dominant factor in the tunneling, the emission probability is the sum of the probabilities for tunneling back to the channel and forth to the gate. Therefore, the emission rate is

$$\frac{1}{\tau_e}=\frac{1}{\tau_{e,ch}}+\frac{1}{\tau_{e,gt}}=\frac{1}{\tau_o}\exp\left(-\frac{x_{ti}}{\lambda}\right)+\frac{1}{\tau_o}\exp\left(-\frac{t_{ox}-x_{ti}}{\lambda}\right), \tag{111}$$

where t_{ox} is the physical thickness of the gate insulator (not EOT), and $1/\tau_{e,ch}$ and $1/\tau_{e,gt}$ are the emission rates to the channel and to the gate, respectively. In this way, the charge exchange rate with the oxide trap becomes as

$$\frac{1}{\tau}=\frac{1}{\tau_c}+\frac{1}{\tau_e}=\frac{1}{\tau_o}\exp\left(-\frac{x_{ti}}{\lambda}\right)\left[2+\exp\left(\frac{2x_{ti}-t_{ox}}{\lambda}\right)\right], \tag{112}$$

and the maximum value for the tunneling time constant is

$$\tau\le\tau_{max}=\frac{\tau_o}{2}\exp\left(\frac{t_{ox}}{2\lambda}\right), \text{ when the trap is in the middle of the oxide.} \tag{113}$$

To illustrate the impact of tunneling to the gate, we assume a conservative high value for $\tau_o=10^{-8}$ s (see after eq. (99)) and plot in Figure 23b the evolution of tunneling time constant τ with the position of the trap in the oxide for several cases relevant to current research of physical oxide thicknesses t_{ox}=3, 4 and 5 nm and λ=0.1nm and 0.21 nm for silicon and hafnium oxides, respectively.

The straight lines in Figure 23b are for very thick oxide (1000nm), for which the gate tunneling is negligible, as it is assumed in the derivations [28, 56, 156] of the model in eq. (109). These lines correspond to the equation

$$\tau=\frac{\tau_o}{2}\exp\left(-\frac{x_{ti}}{\lambda}\right)\left(\frac{t_{ox}-x_{ti}}{t_{ox}}\right)^2, \tag{114}$$

where the last term is for the correction that takes into account the reduced amplitude when the trap is far from the channel and close to the gate [56]. The correction is insignificant for thick oxides.

The peaking lines in Figure 23b are for the tunneling time τ constant in thin oxides, according to the equation

$$\tau=\tau_o\frac{\exp\left(\frac{x_{ti}}{\lambda}\right)}{2+\exp\left(\frac{2x_{ti}-t_{ox}}{\lambda}\right)}\left(\frac{t_{ox}-x_{ti}}{t_{ox}}\right)^2, \tag{115}$$

as follows from eq. (112) and also having the correction term for amplitude. The symbols are for the frequency $f=1/(2\pi\tau)$. One can see that the rising and falling slopes coincide well with exponential functions, and so, the distributions are close to $1/\tau$, and will produce 1/f noise for $f<1/(2\pi\tau_{max})$ – see again the discussion on eqs. (99), (100) and (101). However, the values for τ_{max} decrease when the oxide is thinner and the tunneling attenuation distance λ is larger. The silicon oxides with thickness t_{ox}>3nm still can produce 1/f noise down to 1Hz, but the high-k HfO_2 with a typical thickness of 4-5 nm cannot produce 1/f noise in the frequency range below 1kHz. On the other hand, there is no report that the noise in such samples levels off even at much lower frequencies of 10Hz [47, 49, 50, 87, 92, 99, 100, 109, 126, 128]. The discrepancy between the model prediction and experimental results is evident, and the physical interpretations of the values for N_t or application of scaling rules based on this model, are questionable for MOS transistors with thin high-k oxides.

Issues with condition ii.

The oxide trapping and random walk models neglect the carrier scattering as the source of noise. Therefore, eq.

(109) is for pure Δn fluctuation in carrier number n in the channel of MOS transistors – see again the discussion after eq. (5). At the increased electric fields in modern MOS transistors, however, it is observed that the mobility decreases, which is explained with intensive carrier scattering. So, the trapping in oxide and in interface states was recognized as factor that couples noise via changing the mobility, and the coupling coefficient K in eqs. (88) and (89) is different from 1. In most cases K>1, and varies with the bias of MOS transistor too.

The general assumption is that the trap occupancy changes the number n of charge carriers and their mobility μ simultaneously. Since the origin of the fluctuation is the same, the Δn and Δμ noise are correlated. It has been shown in [59] that the correlation can be derived from first principles and the trap screening due to high charge concentration in the inversion layer also takes place in mobility fluctuation, basically reducing it at higher gate bias in strong inversion. There are several equivalent forms for the equations for oxide trap noise with correlated mobility fluctuation.

One form is derived in [156] at the assumption that the Δn noise can be regarded as noise $S_{V_{FB}}$≈constant in flat band voltage, as given by eq. (109). The derivation uses the following approach. It assumes that the drain current I_D can be written as the product of two functions, one for carrier mobility $\mu(V_G-V_{FB})$ and another for the number of carriers $F_n(V_G-V_{FB})$ independent of mobility. That is

$$I_D = \mu F_n = \frac{\mu_o}{1+F_\mu(V_G-V_{FB})}F_n(V_G-V_{FB}) \approx \frac{\mu_o}{1+\theta(V_G-V_{FB})}F_n(V_G-V_{FB}), \tag{116}$$

where $F_\mu(V_G-V_{FB})=\theta(V_G-V_{FB})+F_2(V_G-V_{FB})\approx\theta(V_G-V_{FB})$ is a function that describes the mobility dependence, mostly degradation, when increasing the bias in the MOS transistor, θ~0.3-3 V^{-1}, but to the first order of approximation, $\partial F_\mu/\partial(V_G-V_{FB})\approx\theta$ and $|F_\mu|<1$. Then, for the number fluctuation

$$dI_D\big|_n = \frac{\partial I_D}{\partial F_n}\frac{\partial F_n}{\partial V_{FB}}dV_{FB} = \mu\frac{\partial F_n}{\partial V_{FB}}dV_{FB} = -\frac{\partial \mu F_n}{\partial V_G}dV_{FB} = -\frac{\partial I_D}{\partial V_G}dV_{FB} = -g_m dV_{FB}, \tag{117}$$

and for mobility fluctuation

$$\begin{aligned} dI_D\big|_\mu &= \frac{\partial I_D}{\partial \mu}\frac{\partial \mu}{\partial V_{FB}}dV_{FB} = F_n\frac{\partial \mu}{\partial V_{FB}}dV_{FB} = -\frac{\mu_o F_n}{(1+F_\mu)^2}\frac{\partial f}{\partial V_G}dV_{FB} \\ &= -\frac{I_D\theta}{1+F_\mu}dV_{FB} = -\frac{I_D\theta\mu_o}{(1+F_\mu)\mu_o}dV_{FB} = -I_D\theta\frac{\mu}{\mu_o}dV_{FB} \\ &\approx -I_D\theta dV_{FB}, \text{assuming } \mu\approx\mu_o. \end{aligned} \tag{118}$$

Combining the two components, one gets

$$\frac{dI_D\big|_{n+\mu}}{I_D} = \frac{dI_D\big|_n}{I_D} + \frac{dI_D\big|_\mu}{I_D} = -\left(\frac{g_m}{I_D}+\theta\frac{\mu}{\mu_o}\right)dV_{FB} \approx -\left(\frac{g_m}{I_D}+\theta\right)dV_{FB}, \tag{119}$$

and writes for the noise in MOS transistors with correlated Δn-Δμ fluctuation

$$\frac{S_{I_D}\big|_{n+\mu}}{I_D^2} = \left(1+\theta\frac{\mu}{\mu_o}\frac{I_D}{g_m}\right)^2\left(\frac{g_m}{I_D}\right)^2 S_{V_{FB}} \approx \left(1+\theta\frac{I_D}{g_m}\right)^2\left(\frac{g_m}{I_D}\right)^2 S_{V_{FB}}. \tag{120}$$

So, the coupling coefficient in eqs. (1), (4), (11), (125), (88) and (89) becomes

$$K=\left(1+\theta\frac{\mu}{\mu_o}\frac{I_D}{g_m}\right)^2\approx\left(1+\theta\frac{I_D}{g_m}\right)^2,\ \text{for MOS transistors with correlated } \Delta n\text{-}\Delta\mu \text{ fluctuation,} \quad (121)$$

where the mobility degradation parameter θ may vary in sign, magnitude and with bias, depending on type of trapping and scattering mechanism in MOS transistor, e.g. Coulomb, remote, surface, phonon scattering and screening of charge of the inversion layer [47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 72, 89, 110, 124, 137], but θ is usually taken as a constant in one transistor, because of the following reasons. At low gate bias, V_G-V_T<0.1V, the contribution of the correlated mobility fluctuation is low, because the ratio I_D/g_m<0.1, and since θ<1V^{-1}, then K≈1. At high gate bias, $I_D/g_m\propto(V_G-V_T)$>0.5V, K becomes a nearly quadratic function of (V_G-V_T), but the product $\theta(V_G-V_T)$ is usually less than 1, and variations in θ are difficult to inspect reliably from noise measurements data with an experimental uncertainty 1-3dB – see for example fig.4 in [89]. So, the experimental characterizations assume θ=constant; and the model of eq. (120) is very attractive since all quantities (S_{I_D}, I_D, g_m) can be estimated directly from the measured data without a need of device modeling. For example, the product $(S_{I_D}/I_D^2)(I_D/g_m)^2$ at low gate bias (V_G-V_T)<0.1V directly estimates S_{FB}, and then at higher gate bias (V_G-V_T)>0.5V, the value for θ can be obtain from the slope of the graph $\sqrt{S_{V_G}}$ vs. V_G, see for example [89], again using $(S_{I_D}/I_D^2)(I_D/g_m)^2=S_{V_G}$, since $I_D/g_m\propto(V_G-V_T)$. The direct parameter extraction by using the model of eq. (120) is a routine approach in noise characterization and most of the results presented here were obtained using this approach. Actually, it is argued in [133] that assuming Δn noise being regarded as noise $S_{V_{FB}}\approx$constant in flat band voltage, there is inherent error in the model that underestimates the noise level when the transistor is in saturation regime. The error is about 3dB, it is small, it is in the range of experimental inaccuracy, and using approximations for charge concentrations would cause similar uncertainty. So the model of eq. (120) is very attractive in experimental characterizations and it is widely used.

The situation, however, has changed when high-k dielectrics and high doping of the channel are used in order to reduce the channel length of the transistors below 100nm. The electric fields and scattering increased in these devices and the correlated mobility fluctuation becomes an issue not only because the contribution from the scattering reduces significantly the effective mobility in MOS transistors, but also because the noise models are developed at an assumption of dominant phonon scattering, while Coulomb screening by the inversion layer and roughness scattering [157, 158] also take place. Therefore, the alternative forms of the model with correlated Δn-Δμ fluctuations for MOS transistors are discussed below, because they are derived at physical assumptions for charge transport in MOS transistors, while the model of eq. (120) above was introduced formally by using the concept for coupled noise at the assumption that the oxide charge trapping can be referred as noise in the flat band voltage. These models are known as "unified" model for 1/f noise in MOS transistors and use charge fluctuation due to trapping in gate oxide accompanied with scattering due to Coulomb interaction around the trapped charge.

Unified model. The starting point in deriving the unified model is that in event of charge trapping in the oxide at distance x_{ti} from semiconductor interface at coordinate (w,l) in the channel region, in which the carrier areal concentration is $n'=\partial^2 n/\partial w\partial l$, the drain current changes as [54, 55, 56]

$$dI_D \propto (\mu dn' + q dn' d\mu) dw dl \Leftrightarrow \frac{dI_D}{I_D} = -\left(\frac{dn'}{n'} \pm \frac{d\mu}{\mu}\right), \tag{122}$$

where dn' and dμ are functions of trap occupancy, given by

$$\begin{aligned} dn' &= \frac{\partial n'}{\partial N_t} dN_t = R dN_t \text{ with } R = \frac{n'}{n' + n^*}, n^* = \frac{\varphi_t}{q}(C_{ox} + C_d + C_{it}), \text{and} \\ \frac{d\mu}{\mu} &= \frac{\partial \mu}{\partial N_t} dN_t = \alpha_s \mu dN_t \text{ with } \frac{1}{\mu} = \frac{1}{\mu_o} + \alpha_s N_t, \end{aligned} \tag{123}$$

where C_d and C_{it} are the capacitance per unit gate area due to depletion under the conductive channel in the MOS transistor and fixed interface states at the semiconductor-insulator interface, respectively, and α_s is a scattering coefficient, which takes into account for Coulomb interaction between oxide trapped charge and thus changing the mobility of the carriers in the inversion layer. In order to solve the integrals, several assumptions are made: the oxide trapping dN_t is negligible as compared to n', N_t, α_s and μ can be replaced with constants, and $d(N_t \alpha_s \mu) = -d(n' \alpha_s \mu / R)$. In this way, the general equation for the unified 1/f noise model for MOS transistor is written in several equivalent forms, one of which is

$$S_{I_D}\Big|_{\text{at constant } V_G \text{ and } V_B} = \frac{\lambda k T q I_D \mu}{f L^2} \int_{V_S=0}^{V_D} N^* \frac{R^2}{n'} dv, \tag{124}$$

where V_G, V_B, V_S and V_D (on which the areal carrier concentration n' in the channel of MOS transistor depends on) are the potentials of the gate, body, source and drain terminals, respectively, all other quantities are related to n', and also by using the approximation

$$N^* = A + Bn' + C(n')^2 = N_t(1 \pm \alpha_s \mu n' / R)^2, \tag{125}$$

where the parameters A=NOIA, B=NOIB and C=NOIC (being assumed constant fitting parameters in BSIM3 model) are approximately corresponding to $A=N_t$, $B=2\alpha_s \mu N_t / R \approx 2\alpha_s \mu N_t$ and $C=(\alpha_s \mu / R)^2 \approx (\alpha_s \mu)^2$. The sign ± is chosen by whether the trap is repulsive or attracting for the carriers in the channel, as mentioned in [48]. In the next step, the unified model is split into three integrals, which always have analytical solutions at A, B and C constant. These solutions are appropriate for compact modeling, since they capture the coupling of noise in the MOS channel by the non-uniform carrier concentration at various drain biasing, depending only on n' at drain and source sides, where the overdrive (V_G-V_T-V) with V=V_S or V_D, are known from the biasing. The model, however is slightly inconvenient for experimental characterizations, because it requires additional estimates for capacitances, the equations are long, see [48, 54, 55], and cannot be rewritten in a form so that A, B and C do not depend on each other. Nevertheless, the unified model converges to the flat band model of eq. (120) above in ohmic and sub-threshold operation of MOS transistor, and the derivations established for linear regime that

$$\begin{aligned} \theta(V_G - V_T) &= \alpha_s \mu n' / R = \alpha_s \mu \frac{C_{ox}(V_G - V_T) + \varphi_t(C_{ox} + C_d + C_{it})}{q} \\ &\approx \alpha_s \mu \frac{C_{ox}(V_G - V_T)}{q}, \text{if } (V_G - V_T) > (2...3)\varphi_t \end{aligned} \tag{126}$$

Therefore, since θ is directly obtained from measurements as discussed above, α_s can be easily evaluated from

$$\alpha_s = \frac{q\theta}{\mu C_{ox}} \sim 10^{-15}\ \text{Vs} \tag{127}$$

taking typical values for 180nm nMOS (C_{ox}~0.8 μF/cm^2, θ~1 V^{-1}, μ~200 cm^2/Vs).

Effects of different scattering mechanisms. The issue with correlated mobility is now addressed. It is due to the fact that neither α_s or μ are constants. This is because changing the gate bias, the electric field changes and there are several crossovers between different scattering mechanisms [157, 158]. These are illustrated in Figure 24.

The mobility is given by the Matthiessen rule, as

$$\frac{1}{\mu} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_r} + \frac{1}{\mu_{it}} + \frac{1}{\mu_b}, \tag{128}$$

where μ_{ph} is due to phonon scattering, μ_r is due to surface roughness scattering, μ_{it} is due to Coulomb scattering caused by interface states and oxide traps and μ_b is due to scattering with ionized impurities in the semiconductor. The investigations in [157, 158] have established that the different components have different dependences with the biasing of the MOS transistor.

- The phonon scattering mobility is given by

$$\frac{1}{\mu_{ph}} = a_{ph} T^{e_t} E_G^{0.3} \approx a_{ph} T^{1.75} \left(\frac{q\eta_E}{\varepsilon_{Si}}\right)^{0.3} \left(\frac{N_{dpl}}{\eta_E} + n'\right)^{0.3}, \tag{129}$$

with T being the absolute temperature, a_{ph}, e_t, η_E=0.5….0.3 being constants, ε_{Si} being the permittivity of semiconductor (~ 1.04×10^{-12} F/cm for silicon), E_G being the gate electric field, $E_G = q(N_{dpl} + \eta_E n')/\varepsilon_{Si}$, and N_{dpl} being the depletion charge per unit area, according to

$$N_{dpl} = \sqrt{\frac{4q}{\varepsilon_{Si}} N_{sub} \varphi_t \ln\left(\frac{N_{sub}}{n_i}\right)}, \tag{130}$$

where N_{sub} is the volume impurity concentration in semiconductor and n_i is the free charge concentration of intrinsic semiconductor (~10^{10} cm^{-3} for silicon at room temperature). The line labeled with “phonon scattering” in Figure 24 illustrates μ_{ph}. At fixed temperature, one can combine the constants into the phonon scattering parameter α_{ph} and rewrite eq. (129) as

$$\frac{1}{\mu_{ph}} = \alpha_{ph} \left(\frac{N_{dpl}}{\eta_E} + n'\right)^{0.3} = \alpha_{ph} \left(n_\eta + n'\right)^{0.3}, \tag{131}$$

where n_η is a scaled version of N_{dpl}.

- The surface roughness scattering mobility is given by

$$\frac{1}{\mu_r} = a_r E_G^{e_r} \approx a_r \left(\frac{q\eta_E}{\varepsilon_{Si}}\right)^{e_r} \left(\frac{N_{dpl}}{\eta_E} + n'\right)^{(2\pm1)}, \tag{132}$$

where a_r and e_r are approximately constants. The line labeled with “surface roughness scattering” in Figure 24 illustrates μ_r, showing that this mobility degradation is observed when the electric field is high, e.g. above

0.5MV/cm. Again combining the constants into a scattering parameter α_r, one writes

$$\frac{1}{\mu_r} = \alpha_r \left(\frac{N_{dpl}}{\eta_E} + n' \right)^{e_r} = \alpha_r \left(n_\eta + n' \right)^{(2\pm1)}. \tag{133}$$

- The Coulomb scattering is prominent when the inversion layer is with low concentration of carriers. The inversion layer screens the Coulomb scattering at higher gate bias. For the Coulomb scattering due to ionized impurities in the semiconductor, the mobility μ_b is given by

$$\frac{1}{\mu_b} = a_b \frac{N_{sub}}{n'} \approx \frac{\alpha_b}{n'}, \tag{134}$$

where α_b is the corresponding scattering parameter. The curve labeled with "impurity screening" in Figure 24 illustrates μ_b, showing that this mobility degradation is observed when the electric field is low, e.g. below 0.2MV/cm. At a little higher gate biasing, as shown with the curve "interface screening" in Figure 24, the Coulomb scattering caused by interface states and oxide traps takes place. Respectively, the mobility μ_{it} is

$$\frac{1}{\mu_{it}} = a_{it} \frac{D_{it} + b_{it} N_t}{\sqrt{n'}} \approx \frac{\alpha_{it}}{\sqrt{n'}}, \tag{135}$$

where α_{it} is the corresponding scattering parameter and b_{it} refers the oxide traps as apparent areal density at the semiconductor-oxide interface.

- Combining all scattering mechanisms, one gets

$$\frac{1}{\mu} = \frac{\alpha_b}{n'} + \frac{\alpha_{it}}{\sqrt{n'}} + \alpha_{ph} \left(n_\eta + n' \right)^{0.3} + \alpha_r \left(n_\eta + n' \right)^{(2\pm1)} = \frac{1}{\mu_o} + \alpha_s (n') \times N_t, \tag{136}$$

where the terms are written in a sequence as they are dominant from low to high gate biasing, resulting in a bell-shaped curve for μ, labeled with "effective mobility" in Figure 24. The last equation implies that the effective scattering parameter α_s varies with the gate biasing, and α_s is high at low and high areal carrier density n' in the inversion layer in MOS transistor channel, as shown with circles in Figure 24. Despite these variations, the term

$$1 + \theta(V_G - V_T) = 1 + \alpha_s \mu n' / R \approx 1 + \alpha_s \mu \frac{C_{ox} (V_G - V_T)}{q} \tag{137}$$

for correlated mobility fluctuation does not change very much, as shown with squares in Figure 24, because the product $\alpha_s\mu$ is constant, if one scattering mechanism is dominant. This can be seen from eq. (136) by neglecting $1/\mu_o$. Usually, the term $[1+\theta(V_G-V_T)]\approx[1+\theta I_D/g_m]$ can be fitted approximately with a linear function within the experimental inaccuracy, especially in silicon MOS transistors with not very thin (>4nm) and well processed SiO_2 for gate insulator. In these transistors, since the transistor channel is L>0.3 μm, then the phonon scattering usually dominates, because $N_{sub}<3\times10^{16}$ cm^{-3} and $E_G<0.6$ MV/cm.

A solution at dominant Coulomb scattering due to interface states and oxide traps is deduced in [159], proposing a modification of eqs. (124) and (125). The modified equation is

$$\frac{S_{I_D}}{I_D^2} = \frac{kT\lambda N_t}{fWL} \left(\frac{1}{n'} + \frac{\mu}{\mu_{C0}\sqrt{n'}} \right)^2 \equiv \frac{kT\lambda N_t}{fWL} \left(\frac{1}{n'} + \frac{\alpha_{it}\mu}{N_t\sqrt{n'}} \right)^2, \tag{138}$$

where the parameter $\mu_{C0} \equiv N_t/\alpha_{it}$ reflects the parameter for screened Coulomb scattering at interface states and oxide traps in eq. (135). This scattering mechanism would be pronounced in sub-100 nm MOS transistors, since $N_{sub} \approx 5\times10^{17}$ cm^{-3}, and $E_G < 0.8$MV/cm. Assuming $\mu \approx \mu_{it}$ in eq. (136), then $\alpha_{it}\mu_{it}(n')^{-0.5} \approx$ constant $\approx 1/N_t$, which is a paradox of cancelling the mobility when the other scattering mechanisms are neglected.

In fact, the identification of scattering coefficients for MOS transistors from sub-100nm MOS is difficult, because $N_{sub} > 7\times10^{16}$ cm^{-3} in these transistors in order to compensate for DIBL (drain induced barrier lowering that affects the threshold voltage V_T), resulting in $E_G > 0.6$ MV/cm necessary to invert the channel conductance and control the inversion layer. These are accompanied with crossover between different scattering mechanisms, and the linear approximation is not precise anymore. Two examples for crossover are shown in Figure 25. In both examples, the correlated mobility noise due to Coulomb scattering is screened and ceases with increasing the gate overdrive voltage. In the second example in Figure 25b, the noise of phonon or roughens scattering causes a rising correlated mobility noise. These data have been explained in other manner in [57], and the crossover between different scattering mechanisms is not analyzed in details for the case of low-frequency noise, although some suggestions are available in [58, 159].

Issues with condition iii.

This condition states that the charge exchange between channel, interface states and oxide traps occurs only at Fermi level and the charge trapping is a Shockley–Read–Hall (SRH) process. The assumption allows using Fermi-Dirac statistics in the superposition of the fluctuations of individual traps, by integration over energy. For example, the gate referred noise voltage is obtained in [54, 55, 56] from

$$S_{V_G} = \frac{q^2}{WLC_{ox}^2} \int_0^{t_{ox}} \left[\int_{-\infty}^{+\infty} \frac{4N_t\tau}{1+(\omega\tau)^2} f(E)(1-f(E))dE \right] dx, \tag{139}$$

using the probability function f(E) of Fermi-Dirac statistics, given by

$$f(E) = \frac{1}{1+\exp\dfrac{E-E_F}{kT}}, \tag{140}$$

where E_F is the quasi Fermi level in the inversion layer of MOS transistor channel. So, the factor f(1-f) provides the term $\tau_e\tau_c/(\tau_e+\tau_c)^2$ of the emission and capture time constants of charge trapping in the limit of Shockley–Read–Hall process. The term $\tau_e\tau_c/(\tau_e+\tau_c)^2$ participates in the expression for Lorentzian spectrum of a generation-recombination process. The factor f(1-f) is sharply peaking function of E, which allows the inner integral of eq. (139) to be solved assuming all other quantities unchanged. Since,

$$\int_{-\infty}^{+\infty} f(E)(1-f(E))dE = kT, \tag{141}$$

then the absolute temperature T occurs in the final expressions for the noise, as given by eqs. (104), (105), (109), (124), (138) and other derived from them. These equations suggest that the 1/f noise in MOS transistors should be proportional to the absolute temperature even assuming tunneling in gate oxides. The issue is that this proportionality is not observed experimentally [160], rather, the 1/f noise is found to be temperature independent in nMOS transistors [91] (with a small variation in the slope of the spectrum), while the tunneling model for

noise should be valid for nMOS transistors. Thus, the use of Shockley–Read–Hall process might be incorrect for the tunneling noise from oxide traps, since it is questionable whether the traps are in equilibrium. On the other hand, the capture and emission time constants in RTS noise in MOS transistors are found to follow Shockley–Read–Hall process [68, 74, 107, 112, 148, 161, 162]. Other issues related to the assumptions for superposition of tunneling events in gate oxide in creating 1/f noise are discussed in [4].

Issues with condition iv.

This condition for Δn models of MOS transistors states that the trapping centers are uniformly distributed both in space and energy, the populations of traps and carriers are large enough to be assumed continuous and approximated with averages, and the mobility and electric field do not fluctuate.

Let us inspect the numbers for a sub-100 μm MOS transistor from the L=65nm technology. Assume W=3L=195nm, WL=2.1×10^{-10} cm^2, EOT=2nm, C_{ox}=1.8μF/cm^2, (V_G-V_T)=0.5V, inversion layer carrier density n'=$C_{ox}(V_G-V_T)/q=5.5\times10^{12}$ cm^{-2}, number of carriers WLn'≈700. Using ITRS past predictions for the 65nm technology node [3], we would have WLS_{V_G}(at 1 Hz)=$FOM_{S_{VG}}=1.6\times10^{-10}$ $\mu m^2V^2/Hz$, which would result in $N_t=3.68\times10^{17}$ $cm^{-3}eV^{-1}$, assuming tunneling attenuation distance λ=0.2nm close to that of HfO_2. The traps only within ±3kT around quasi Fermi level fluctuate, the other are either occupied or empty. For a noise measurement in 3 frequency decades, the traps are located in a slice of oxide thickness $\Delta t_{ox}=\lambda\ln(10^3)$=1.38nm. So, the number of traps observed in the measurement would be

$$6kT\Delta t_{ox}N_tWL\approx1.01,\text{ for W/L=195/65 nm transistor in 3 decades of frequency.} \quad (142)$$

Evidently, the population of traps is small in sub-100nm MOS transistors and the assumption in the Δn models for 1/f noise in MOS transistors for continuous distributions approximated with averages is not valid. One should (and do) observe Lorentzian spectra and RTS noise, instead of 1/f noise, in these devices. However, 1/f noise is still present in sub-100nm devices, and it is perhaps from the intrinsic (Hooge or mobility) noise, because the number of carriers is in the range of several hundred carriers, which are at least 10 carriers per frequency decade. So, the assumption, as made in the Δn models for 1/f noise in MOS transistors, that the mobility (electric field or other quantity related to the intrinsic properties of carrier transport) can be neglected, is not valid for sub-100nm MOS transistors. Nevertheless, the Δn model converges well with the observed RTS in small-area MOS transistors, as it will be shown later in section IV.3 , and the model should not be "retired"; rather, the model will be properly analyzed and extended to describe the noise variation when the populations of traps and carriers are small.

In fact, the statistical modeling of noise based on the Δn model has begun, both empirically, analytically and by means of simulations. In this paper, we introduced the problem of statistical noise modeling in the section for BJT – please see again eqs. (80), (81), (82), (83), (84), (85) and the discussion after eq. (85). More discussions on averaging techniques are given in section VIII. "**Outlook for the LFN**" after eq.(469).

Issues with condition v.

This condition for Δn models of 1/f noise in MOS transistors states that the barrier for tunneling in the oxide (or for trapping in random walk model) is constant and not affected by the electric field. Also, the barrier is at the semiconductor-dielectric interface and the channel carriers are at this interface too.

Certainly, these assumptions were good for gate oxides with thickness t_{ox}>10nm until the gate electric field was

less than $E_G \approx (V_G - V_T)/t_{ox} < 1V/10nm = 1MV/cm$, c.f. nodes with minimum gate length $L_{min} > 0.35\mu m$, which was the case when the Δn models were developed. In a thin insulator and at high fields, however, the approximations with constant parameters become rough. Consider Figure 26 for a MOS transistor with gate insulator stack, such as HfO_2 with interfacial layer of SiO_2. The steep slope of the potential in the semiconductor creates a potential well near the dielectric. Owing to quantum effects, the energy levels of electrons increase with approximately $\Delta\Phi_q \sim 0.2eV$ at electric field greater than 0.5MV/cm [51, 163]. One effect is that the tunneling barrier effectively decreases with $\Delta\Phi_q$, which is also accompanied with a departure from rectangular barrier, leading to another barrier lowering with $\Delta\Phi_e$, especially when the oxide is thinner than 1.5 nm [164]. Consequently, the tunneling attenuation distance λ increases according to from Wentzel-Kramer-Brillouin (WKB) approximation of eq. (23). The barrier lowering was modeled in [51] using the Schottky limit for emission over barrier, given by

$$\Delta\Phi = \sqrt{\frac{q^3 E_{ox}}{4\pi\varepsilon_{ox}}}, \tag{143}$$

Other quantum effect at high electric field is that the centroid of the inversion layer is moved 1.5-2.5 nm in the depth of semiconductor, as shown with Δx_q in Figure 26, which is 2-3 times deeper than for the case of using Poisson equation alone, without considering quantum effects. Similar effect due to depletion of polysilicon gates occurs at the gate side [151], shown with $\Delta x_d \sim 2mn$ in Figure 26. There are consequences for Δn model of 1/f noise in MOS transistors.

The first consequence is that Δx_q and Δx_d result in potential drops. For Δx_q the increase of the surface potential is

$$\Delta\psi_s = q\frac{n'\Delta x_q}{\varepsilon_{Si}} \approx \frac{n'}{3\times 10^{13} cm^2} 1V \sim 0.1\text{-}0.3V \text{ in strong inversion.} \tag{144}$$

The order of magnitudes is similar for Δx_d. Thus, the voltage across gate dielectric is approximately

$$\begin{aligned} V_{ox} &= V_G - V_T - \Delta\psi_s - \Delta\psi_{xd} \approx V_G - V_T - 1.5\Delta\psi_s \\ &\sim V_G - V_T - 1.5q\frac{n'\Delta x_q}{\varepsilon_{Si}} \approx V_G - V_T - \frac{n'}{2\times 10^{13} cm^2} 1V \sim V_G - V_T - (0.2 \ldots 0.5)V \end{aligned} \tag{145}$$

For transistors with thin oxide, e.g. nodes 90nm and below, the supply voltage is about 1V, $V_T \sim 0.2V$, and ⅓ to ½ of overdrive voltage $(V_G - V_T)$ is "lost" in quantum effects and depletion of polysilicon gate. For overdrive in the range of 0.8V, the corresponding $E_{ox} = V_{ox}/t_{ox} < 0.6V/3nm \sim 2MV/cm$, and the barrier lowering $\Delta\Phi$ due to E_{ox} is in the range 0.2-0.25eV, according to eq. (143). This is about 6% barrier lowering for electrons at SiO_2 interface, which would result in small increase of 3% for the tunneling attenuation distance λ, according to Wentzel-Kramer-Brillouin (WKB) approximation by eq. (23). So, the 1/f noise $S_{V_G} \propto \lambda$ due to Δn fluctuation would not be affected hardly by barrier lowering, when the gate dielectric is SiO_2. Taking an average number from Table 2, the same calculation for HfO_2 suggests a barrier lowering of about 0.25eV/1.25eV~20%, or increase of 10% for $S_{V_G} \propto \lambda$. On the other hand, the barrier lowering is found pronounced in thin dielectrics, suggesting that WKB approximation is not accurate for these dielectrics [164], and the noise in MOS transistors with high-k dielectrics is relatively high. We see here an open question on how to implement the barrier lowering into Δn fluctuation model for 1/f noise in MOS transistors. The use of constant value for tunneling attenuation distance λ estimated from WKB in Δn fluctuation model is not precise when the physical oxide thickness is less than 5nm. Also, for

the case of stacking several dielectrics in the gate insulator, the assumption for one value for λ is rough. A suggestion is given in [49] for how one can modify the Δn fluctuation model when two materials are used in the gate insulator stack. The suggested modification uses weighting functions A and B, and it is given by

$$(\lambda N_t)_{eff} = A(f, t_{ox1})\lambda_1 N_{t1} + B(f, t_{ox1}, t_{ox2})\lambda_2 N_{t2}, \tag{146}$$

where for each dielectric layer in the stack, λ_1 and λ_2 are the corresponding tunneling attenuation distances, and N_{t1} and N_{t2} are the corresponding oxide trap densities. The functions A and B depend both on the thicknesses t_{ox1}, t_{ox2} of the dielectric layers and the frequency f. More details will be given shortly in section IV.2 when discussing the gate dielectric profiling.

The second consequence for Δn fluctuation model, owing to Δx_q and Δx_d, is that the gate capacitance is bias dependent and it is less than the capacitance C_{ox} of gate insulator stack. Thus, the derivation for S_{V_G} based on eqs. (139), c.f. eqs. (89), (98), (101), (104) and (109), and the assumption R=n'/(n'+n*) in eq. (123) become approximate, since the depletion capacitance of polysilicon gate and the "capacitance" arisen from the distance of the inversion layer centroid are not very large (ε_{Si}/1nm~10μF/cm^2) when compared to the gate insulator capacitance larger than 1μF/cm^2. This consequence also reflects in the next issue.

Issues with condition vi.

This condition for Δn models of 1/f noise in MOS transistors states that the charge captured in the oxide is approximately at the semiconductor-dielectric interface. The distance x_{ti} from interface, where the charge is trapped, is negligible as compared to dielectric thickness. This condition is clearly stated in [56], where also is given that

$$\delta n = \left(\frac{t_{ox} - x_{ti}}{t_{ox}} \right) \delta N_t, \tag{147}$$

where t_{ox} is the physical thickness of the gate insulator, x_{ti} is the distance from semiconductor-insulator interface to the position of the trap in the oxide, δN_t is variation in the number of trapped charges at x_{ti} and δn is the change in the number of carriers in the channel of MOS transistor. Eq. (147) arises from Coulomb ("image charge") balancing of the trapped charge between capacitances from the position of the trap to inversion layer and gate, where the charge is "mirrored". Consider that the charge is trapped between the two dielectric layers in the gate stack in Figure 26, where the right-hand layer is SiO_2 interfacial layer with thickness x_{ti} and permittivity ε_{SiO_2}, and the left-hand layer is HfO_2 with thickness (t_{ox}–x_{ti}) and permittivity ε_{HfO_2}. The capacitance (per unit area) C_{xti} from the trap to the inversion layer is

$$\frac{1}{C_{xti}} = \frac{1}{C_{SiO_2}} + \frac{1}{C_{Si,xq}} = \frac{x_{ti}}{\varepsilon_{SiO_2}} + \frac{\Delta x_q}{\varepsilon_{Si}} \approx \frac{x_{ti}}{\varepsilon_{SiO_2}} + \frac{1}{10\mu F/cm^2}, \tag{148}$$

including the capacitance $C_{Si,xq}$ due to the displacement of the inversion layer centroid from the semiconductor-insulator interface. The corresponding capacitance to the gate conductor is

$$\frac{1}{C_{gti}} = \frac{1}{C_{HfO_2}} + \frac{1}{C_{Si,xd}} = \frac{t_{ox} - x_{ti}}{\varepsilon_{HfO_2}} + \frac{\Delta x_d}{\varepsilon_{Si}} \approx \frac{t_{ox} - x_{ti}}{\varepsilon_{HfO_2}} + \frac{1}{10\mu F/cm^2}, \tag{149}$$

including the capacitance $C_{Si,xd}$ due to depletion of polysilicon gate. From the Coulomb balancing, we have

$$\frac{q\delta N_t}{C_{xti}+C_{gti}}=\frac{q\delta n}{C_{xti}}=\frac{q\delta n_g}{C_{gti}}. \tag{150}$$

The fluctuation in the channel charge, therefore, is

$$\delta n=\frac{C_{xti}}{C_{xti}+C_{gti}}\delta N_t=R_{xti}\delta N_t\le\delta N_t, \tag{151}$$

where R_{xti} has the same meaning as R in eqs. (123) and (124) of coupling between oxide trap and channel charges, but R_{xti} depends on the position of the oxide trap, and therefore, from the time constant τ of the trap, since $\tau=\tau_o\exp(x_{ti}/\lambda)$, according to eq. (99). For the example taken above, we have

$$R_{xti}=\frac{C_{xti}}{C_{xti}+C_{gti}}=\frac{1/C_{gti}}{1/C_{gti}+1/C_{xti}}=\frac{t_{ox}-x_{ti}+\dfrac{\varepsilon_{HfO_2}}{10\mu F/cm^2}}{t_{ox}-x_{ti}\left(\dfrac{\varepsilon_{HfO_2}}{\varepsilon_{SiO_2}}-1\right)+\dfrac{\varepsilon_{HfO_2}}{10\mu F/cm^2}}, \tag{152}$$

which reduces to the expression in the brackets of eq. (147), neglecting the depletion in polysilicon gate and quantum effects in the channel and taking uniform dielectric ($\varepsilon_{HfO_2}=\varepsilon_{SiO_2}$). The actual expression for R_{xti} is more complicated, if one considers that the trap is not at the boundary between dielectrics in the gate insulator stack. Also, R_{xti} is a function of bias and tunneling time constant. For the simple case of uniform dielectric and depletion in polysilicon gate and quantum effects neglected, one has

$$R_{xti}(\tau)=\frac{C_{xti}}{C_{xti}+C_{gti}}=1-\frac{\lambda\ln(\tau/\tau_o)}{t_{ox}}, \tag{153}$$

and R_{xti} has to be included in the evaluation of the integral for superposition of Lorentzian spectra, since it multiplies R in the general expression, c.f. eq.(124) of the unified 1/f noise model for MOS transistor. Thus,

$$S_{I_D}\approx\frac{\lambda kTqI_D\mu}{L^2}\int_{V_S=0}^{V_D}N^*\frac{R^2}{n'}\left[\int_{\tau_o}^{\tau_{max}}\frac{4R_{xti}^2d\tau}{1+(2\pi f\tau)}\right]dv, \tag{154}$$

but the integral in the square brackets is solved analytically only for the case $R_{xti}=1$, to the best of our knowledge. Provided that the position of x_{ti}~2 nm of the slow traps inside the oxide is a significant portion of thin oxides t_{ox}~3-5nm, and also the depletion of polysilicon gate and quantum effects are not explicitly considered in the coupling parameters of the unified model, then an issue arises that these has to be included, in order to preserve the physical consistence of the model when applied to aggressively down-scaled MOS transistors.

Conclusions to issues (i)-(vi).

Owing to the above issues, the accuracy of the Δn-$\Delta\mu$ model for 1/f noise in modern transistors with thin and stacked gate dielectrics is not very large. However, when carefully calibrated, the model is convenient for compact modeling, circuit simulation, and it is vital in providing information for comparisons in a qualitative manner for the properties and quality of gate stacks. One interesting class of techniques based on Δn model for characterization of the oxide trap profile is now discussed.

IV.2. Charge trap profiling of gate dielectrics

The Δn model for 1/f noise in MOS transistors, see eqs. (104) and (105), uses approximation with uniform distributions, both in energy and space, for oxide traps and interface states. Such distributions result in $g(\tau)\propto 1/\tau$ distribution for the time constants of the traps, e.g. eq. (100) for oxide traps, and produce 1/f noise from superposition of the Lorentzian spectra of the fluctuation of the individual traps – see again eqs. (72) to (75) and Figure 19. The deviation of noise power spectrum density from 1/f is used to evaluate the departure from uniform trap distribution, and in this way, provides profiling of the traps, by means of energy or distance, since either of them modifies the $1/\tau$ distribution for the time constants of the traps [73]. The duality of capture energy and tunneling distance is addressed in [147] and earlier by eqs. (102) and (103), and the separation of spatial and energy profiles of the traps is made at assumptions for physical consistence at pre-determined characterization model [73].

IV.2.1. Spatial profiling of trap density

Simple approach

The simplest approach for spatial profiling of trap density $N_t(x_{ti})$ in the oxide depth at distance x_{ti} from semiconductor-insulator interface is used in [140] for a MOS transistor with gate stack of 2.1nm SiO_2 interfacial layer at semiconductor interface and 5nm HfO_2 on top of it. The assumption is that at given frequency $f=f_i$, the traps with time constant $\tau_i=1/(2\pi f_i)$ are dominant in the 1/f noise S(f), since the Lorentzian spectrum of these traps has maximum contribution in the quantity $f_i\times S(f_i)$. At this assumption, one writes from eqs. (99) and (101) that

$$\tau_o \exp\left(\frac{x_{ti}}{\lambda}\right)=\tau_i=\frac{1}{2\pi f_i}\Rightarrow x_{ti}=\lambda\ln\left(\frac{1}{2\pi f_i\tau_o}\right),\ \text{with } (2\pi f_i\tau_o)<<1, \tag{155}$$

$$S_{V_G}=\frac{1}{f_i}\left(\frac{q}{C_{ox}}\right)^2\frac{kT}{WL}\lambda N_t(x_{ti})\Rightarrow N_t(x_{ti})=\left(\frac{C_{ox}}{q}\right)^2\frac{WL}{kT}\frac{f_i S_{V_G}(f_i)}{\lambda}, \tag{156}$$

and taking $\lambda\approx$constant$\approx$0.1nm, both x_{ti} and $N_t(x_{ti})$ can be obtained, as shown in Figure 27 for the abovementioned MOS transistor. Note that the oxide trap profiling by using this simple approach is qualitative, as mentioned in [140], since details are not elaborated when N_t is not uniform, when two materials with different tunneling attenuation distances $\lambda_{SiO_2}\neq\lambda_{HfO_2}$ are used, and when the overlap in the spectra of traps with different time constants is neglected. The later caused that the increase of N_t in Figure 27 is detected at x_{ti}=1.8nm rather than at 2.1nm, where the interface SiO_2-HfO_2 was. This is similar to the broadening of step changes in trap densities when measured by charge pumping techniques [147]. In the origin of this broadening is the integral form that describes the superposition of Lorentzian spectra of individual traps with similar time constants, c.f. eqs. (72) to (75). It is shown in [68] that the superposition integrals are equivalent to a convolution between the trap distributions in space and energy, since both distributions affect the distribution of trap time constants, which is used as integration variable. The effect of the convolution is that the slope of 1/f noise in respect to the frequency may vary, rather than a step in the level of 1/f noise magnitude at particular frequency to be observable [2, 56, 68]. Mathematically, the reason is that the function $f\tau/[1+(2\pi f\tau)^2]$ is not sharply peaking at $2\pi f=1/\tau$, in order to produce a step when integrating it. To demonstrate the problem, we adopt the approach from [49], as follows.

Gate stacks

In the approach for spatial trap profiling in [49], one from the assumptions above is not postulated, in particular, that the traps with time constant $\tau_i=1/(2\pi f_i)$ are dominant in the 1/f noise at $f=f_i$. This allows inspecting the evolution of 1/f noise in gate stacks made of different materials, with different trap densities ($N_{t,IL}$, $N_{t,Hk}$) and different tunneling attenuation distances (λ_{IL}, λ_{Hk}) at relaxed assumption that the materials are otherwise uniform. The index "IL" stands for interfacial dielectric layer of thickness t_{IL} between semiconductor and main dielectric of thickness t_{Hk}. The index "Hk" stands for the main gate dielectric, which is usually with high permittivity (high-k). We also make an approximation that the interface states are at semiconductor-insulator boundary and have narrow distributions (delta functions) for the cross-section and barrier energy, so that the prefactor τ_o for the tunneling time is constant for the entire gate stack. Consequently, we assume that the conditions i, ii, iii, iv, v, and vi for number fluctuation model, which are discussed above, are satisfied. Therefore, we can split the superposition integral for the gate referred voltage noise into two parts corresponding to interfacial layer and main dielectric, resulting in

$$S_{V_G} \approx S_{V_{FB}} = a\left[\int_0^{t_{IL}} \frac{N_{t,IL}\,4\tau(x)}{1+(2\pi f)^2\tau^2(x)}dx + \int_{t_{IL}}^{t_{Hk}+t_{IL}} \frac{N_{t,Hk}\,4\tau(x)}{1+(2\pi f)^2\tau^2(x)}dx\right] \tag{157}$$

where $a=(q/C_{ox})^2(kT)/(WL)$ is regarded as device constant at particular temperature T, and the tunneling time constant $\tau(x)$ increases exponentially with the distance x from semiconductor-dielectric interface as

$$\tau(x)=\begin{cases}\tau_o \exp\left(\dfrac{x}{\lambda_{IL}}\right), \text{for } x \le t_{IL} \text{ in IL}\\ \tau_o \exp\left(\dfrac{t_{IL}}{\lambda_{IL}}\right)=\tau_{IL}, \text{for } x = t_{IL} \text{ at IL - Hk interface}\\ \tau_{IL} \exp\left(\dfrac{x-t_{IL}}{\lambda_{Hk}}\right), \text{for } t_{IL} \le x \text{ in main Hk dielectric}\end{cases} \tag{158}$$

The solution of the integrals in eq. (157) gives

$$S_{V_G} = \frac{2a}{\pi f}\left\{\lambda_{IL}N_{t,IL}\,\text{artg}\left[2\pi f\tau_o \exp\left(\frac{x}{\lambda_{IL}}\right)\right]\Bigg|_{x=0}^{t_{IL}} + \lambda_{Hk}N_{t,Hk}\,\text{artg}\left[2\pi f\tau_{IL} \exp\left(\frac{y}{\lambda_{Hk}}\right)\right]\Bigg|_{y=0}^{t_{Hk}}\right\} \tag{159}$$

For a low-frequency noise measurement in a bandwidth (f_{max}–f_{min})~(100kHz–1Hz), assume that $2\pi f_{max}\tau_o<<1$, so that $\text{artg}(2\pi f_{max}\tau_o)\approx 0$, and the thickness $t_{ox}=(t_{Hk}+t_{IL})$ of the gate oxide stack is much larger than the tunneling attenuation distance $\lambda_{Hk}>\lambda_{IL}$, so that $[2\pi f_{min}\tau_o\exp(t_{IL}/\lambda_{IL})\exp(t_{Hk}/\lambda_{Hk})\,]>10>>1$ even for the minimum frequency of interest f_{min}~1Hz [see after eq. (115) for critical discussion on this assumption]. Then, the substitution of the limits in eq. (159) provides

$$S_{V_G} \approx \frac{a}{f}\left[\lambda_{Hk}N_{t,Hk} - \frac{\lambda_{Hk}N_{t,Hk}-\lambda_{IL}N_{t,IL}}{\pi/2}\,\text{artg}(2\pi f\tau_{IL})\right], \text{with } \tau_{IL}=\tau_o\exp\left(\frac{t_{IL}}{\lambda_{IL}}\right) \tag{160}$$

Evidently from this equation, the 1/f noise in MOS transistors with stack of gate dielectrics depends not only on the properties of the materials via the products λN_t, but also the 1/f noise magnitude is a function of the thickness of the layers and the frequency of measurement via the term $\text{artg}[2\pi f\tau_o\exp(t_{IL}/\lambda_{IL})]$. At low frequency, $2\pi f_{low}\tau_o\exp(t_{IL}/\lambda_{IL})\approx 0$, the 1/f noise depends on the main gate dielectric, with $S_{V_G}\approx a\lambda_{Hk}N_{t,Hk}/f$. At high frequency,

in contrast, $2\pi f_{hi}\tau_o\exp(t_{IL}/\lambda_{IL})>10>>1$, and the interfacial layer dielectric determines the 1/f noise with $S_{V_G}\approx a\lambda_{IL}N_{t,IL}/f$, since $\text{artg}[2\pi f_{hi}\tau_o\exp(t_{IL}/\lambda_{IL})]\approx\pi/2$. At intermediate frequencies around $2\pi f_{med}\tau_o\exp(t_{IL}/\lambda_{IL})\sim1$, both dielectrics contribute to the 1/f noise, causing a transition region in the noise spectrum with a slope different from 1/f, as shown in Figure 28. Usually $\lambda_{Hk}N_{t,Hk}>\lambda_{IL}N_{t,IL}$ (solid symbols in the figure), and one observes noise spectrum with a slope steeper than 1/f in a range of about two frequency decades, which corresponds to dielectric thickness of about $(4–5)\lambda\sim1$nm around the IL-Hk interface. In contrast, if $\lambda_{Hk}N_{t,Hk}<\lambda_{IL}N_{t,IL}$ (open symbols in the figure), then noise spectrum levels off in the transition region with a slope less than 1/f. The width of the transition region is proportional to the ratio $\lambda_{Hk}N_{t,Hk}/\lambda_{IL}N_{t,IL}$. Thus, one would characterize a gradual change in N_t, if applying the simplest approach of using eqs. (155) and (156). The tail of this gradual change is seen in Figure 27, although the IL-Hk interface perhaps is much abrupt and located above 2nm. Nevertheless, by comparing the results for the product fS(f) in 4-5 frequency decades, the simplest characterization approach of eqs. (155) and (156) can be useful to determine whether N_t is higher in the main dielectric or in the interfacial layer, since $\text{artg}[2\pi f\tau_o\exp(t_{IL}/\lambda_{IL})]$ is either zero or $\pi/2$ at the ends of the frequency range. Such spatial profiling of oxide traps is demonstrated in [128], and it was estimated that the gate conductor also affects the low-frequency noise, which generally leads to the issues with conditions i and vi for the Δn model of 1/f noise in MOS transistors with thin gate dielectrics. These issues were discussed earlier by the help of eqs. (110) to (115) and (147) to (154), respectively.

In the above discussion, we have assumed the oxide trap density changes abruptly at a depth t_{IL} in the oxide. This would result in slopes $1/f^2$ or zero only in the transition region. However, one may observe in low-frequency noise measurements a slope different from 1/f in the entire frequency range of 4-5 frequency decades, which suggests that the trap density changes gradually either along the depth of the oxide or the traps have non-uniform energy distribution. The effect of gradual trap distributions in the oxide depth was addressed in [56], and in [2] an analytical model was suggested for the case when the energy distribution of the traps is broad. The slope of the flicker noise power spectrum is affected in both cases, and it deviates from the 1/f slope [73].

Gradual oxide trap distribution

To inspect the effect of gradual oxide trap distribution on the slope γ of the $1/f^{\gamma}$ flicker noise, we rewrite $\omega\equiv2\pi f$ and $\xi\equiv x/\lambda=0\ldots\xi_{max}$ with $\xi_{max}=t_{ox}/\lambda>>1$ in the superposition integral and integrate it by parts, as

$$\begin{aligned}
S_{V_G} &\approx bkT\int_{x=0}^{t_{ox}} N_t(x)\frac{\tau_o\exp\left(\frac{x}{\lambda}\right)}{1+(\omega\tau_o)^2\exp^2\left(\frac{x}{\lambda}\right)}dx \\
&= bkT\lambda\int_{0}^{\xi_{max}} N_t(\lambda\xi)\frac{\tau_o\exp(\xi)}{1+(\omega\tau_o)^2\exp^2(\xi)}d\xi \\
&= \frac{bkT\lambda}{\omega}\left\{\begin{array}{l} N_t(\lambda\xi_{max})\text{artg}[\omega\tau_o\exp(\xi_{max})]-N_t(0)\text{artg}(\omega\tau_o) \\ -\int_{0}^{\xi_{max}}\text{artg}[\omega\tau_o\exp(\xi)]\frac{\partial N_t}{\partial\xi}d\xi \end{array}\right\}=\frac{bkT\lambda}{\omega}F,
\end{aligned} \tag{161}$$

where $b=a/(kT)=(q/C_{ox})^2/(WL)$, and we denote the function in the large brackets with $F=\{\ldots\}$. Since the

integration is not on the frequency, then we use the Leibniz rule for differentiating under the integral sign, and obtain

$$\frac{\partial S_{V_G}}{\partial \omega} = bkT\lambda F \frac{\partial (1/\omega)}{\partial \omega} + \frac{bkT\lambda}{\omega}\frac{\partial F}{\partial \omega} \quad \text{with} \quad \frac{\partial (1/\omega)}{\partial \omega} = -\frac{1}{\omega^2} \quad \text{and}$$

$$\frac{\partial F}{\partial \omega} \approx \left\{ \begin{array}{l} \dfrac{N_t(\lambda \xi_{max})/\omega}{\omega \tau_o \exp(\xi_{max})} \\ -\dfrac{N_t(0)}{\omega}\omega\tau_o \\ -\displaystyle\int_0^{\xi_{max}} \frac{\partial N_t}{\partial \xi}\frac{\tau_o \exp(\xi)}{1+(\omega\tau_o)^2 \exp^2(\xi)} d\xi \end{array} \right\} \tag{162}$$

Here, we will use that $\omega\tau_o\exp(\xi_{max}) >> 1 >> \omega\tau_o \approx 0$, since the measurement frequency range is well within the range between minimum and maximum values of tunneling time constants, and one can omit the first and second terms in the last large brackets. By multiplying the nominator and denominator of the expression under the integral sign with N_t, we get

$$\begin{aligned} \frac{\partial S_{V_G}}{\partial \omega} &= -\frac{1}{\omega}\left[\frac{bkT\lambda}{\omega}F\right] - \frac{bkT\lambda}{\omega}\int_0^{\xi_{max}} \frac{\partial \ln(N_t)}{\partial \xi}\frac{N_t \tau_o \exp(\xi)}{1+(\omega\tau_o)^2 \exp^2(\xi)} d\xi \\ &= -\frac{S_{V_G}}{\omega} - \frac{\partial \ln(N_t)}{\partial \xi}\frac{1}{\omega}\left[bkT\lambda \int_0^{\xi_{max}} \frac{N_t \tau_o \exp(\xi)}{1+(\omega\tau_o)^2 \exp^2(\xi)} d\xi\right] \\ &= -\frac{S_{V_G}}{\omega} - \frac{\partial \ln(N_t)}{\partial \xi}\frac{S_{V_G}}{\omega}, \end{aligned} \tag{163}$$

where the expressions in the square brackets are the same as eq. (161), and we have assumed that $\partial\ln(N_t)/\partial\xi \approx$ constant, which suggests searching for exponential dependence of N_t on the distance x from semiconductor-insulator interface [73]. By dividing both sides of eq. (163) on S_{V_G}/ω and using $\partial z/z = \partial \ln(z)$, we get for the slope γ of the $1/f^{\gamma}$ flicker noise that

$$\gamma = -\frac{\partial \ln(S_{V_G})}{\partial \ln(\omega)} = -\frac{\partial \ln(S_{V_G})}{\partial \ln(f)} = 1 + \frac{\partial \ln(N_t)}{\partial \xi} = 1 + \frac{\partial \ln(N_t)}{\partial (x/\lambda)} \tag{164}$$

Therefore, at constant slope γ of the $1/f^{\gamma}$ flicker noise, the oxide trap density evolves as an exponential function of the distance x from semiconductor-insulator interface, given by

$$\begin{aligned} \frac{N_t(x - x_{ti})}{N_t(x_{ti})} &= \exp\left[(\gamma - 1)\frac{x - x_{ti}}{\lambda}\right] \\ &= \exp\left\{-\left[\frac{\partial \ln(S_{V_G}(f_i))}{\partial \ln(f)} + 1\right]\frac{x - x_{ti}}{\lambda}\right\}, \end{aligned} \tag{165}$$

where x_{ti} and $N_t(x_{ti})$ are the depth and trap density in the oxide, according to eq. (155) and (156), respectively, probed by measuring the magnitude and slope γ of the $1/f^{\gamma}$ flicker noise at frequency f_i. Provided that the power

spectrum density of the flicker noise in MOS transistor has a slope within $\gamma=1\pm0.2$, then the characteristic oxide thickness t_{exp} for gradual variation of N_t is

$$t_{exp} = \frac{\lambda}{|\gamma - 1|} \geq 5\lambda \sim 0.5 \ldots 1 \text{ nm, for } \lambda=0.1 \text{ nm } (SiO_2) \text{ to } 0.2\text{nm } (HfO_2), \text{ see Table 2,} \qquad (166)$$

which suggests that one decade change in N_t for a distance larger than $2.3t_{exp}\sim1.15\ldots 2.3$ nm will produce 1/f-like noise. The value for t_{exp} is somewhat small. Previous estimate reported in [73] is even smaller, $t_{exp}\approx0.154$ nm, which is about two-three atomic distances for one decade change in N_t, and it is unrealistic number for gradual variation of oxide trap density, since it suggests delta-like distribution for N_t in contrary to the assumption in the Δn model for uniform distribution of N_t. Nevertheless, the ideas in [73] for dependence of 1/f noise in MOS transistors simultaneously on energy and spatial distributions of oxide trap density are interesting and have to be followed up.

General rules

To summarize the above three methods for trap density profiling in the depth of gate insulator by means of low-frequency noise measurement, we provide the following rules.

The simple profiling of oxide trap density N_t based eqs. (155) and (156) is accurate only for regions in flicker noise spectrum with constant frequency slope, which is very close to 1/f. If the slope in the spectrum is constant, but different from 1/f, then an exponential dependence of N_t on oxide depth x is expected, as given by eqs. (164) and (165). If a discontinuity in 1/f spectrum occurs, then an interface inside the gate insulator with abrupt change of the product (λN_t) at this interface is expected, and the noise spectrum is given by eq. (162). In this case, the transition region in the low-frequency noise spectrum with slope different from 1/f begins at frequency f_i corresponding to the depth x_{ti} of the internal interface and the noise level $[f_iS(f_i)]$ is determined by the material with higher value of product $(\lambda N_t)_{high}$. The frequency bandwidth of the transition region is a function of the ratio $(\lambda N_t)_{high}/(\lambda N_t)_{low}$ in the two materials. The above rules do not apply, if the spectrum is constant (slope zero) at $f<f_i$ and the slope is $1/f^2$ at higher frequency $f>f_i$. In this case, the spectrum is dominated by a generation-recombination or random telegraph noise with a specific time constant $\tau_i=1/(2\pi f_i)$, resulting in a Lorentzian shape of low-frequency noise. Overall, the frequency range of flicker noise measurements is not large enough to reliably separate uniform, gradual and step profiles for oxide trap density N_t. This is mostly due to the convolution inherent for the superposition integrals, on which the Δn models for noise in MOS transistors are based on. A good de-convolution scheme is currently lacking.

IV.2.2. Energy profiling of trap density

We have mentioned several times that there is a duality (and convolution) between capture barrier energy E_B and distance x_{ti} of the oxide traps – see again eqs. (102), (103) and (104). Above we have discussed the trap profiling in terms of distance x_{ti} from semiconductor-dielectric interface. However, equations (155) and (156) can be rewritten in terms of energy. At an assumption for (apparent at semiconductor-dielectric interface) areal trap density $D_t=0.1D_{it}$ or $D_t=\lambda N_t(x_{ti}=0)$, as follows from eq. (104), it is shown in [2, 73, 165] that one can deduce energy profile $D_t(E_B)$ in MOS transistor, given by

$$\tau_{ph}\exp\left(\frac{E_B}{kT}\right) = \tau_i = \frac{1}{2\pi f_i} \Rightarrow E_B = kT\ln\left(\frac{1}{2\pi f_i \tau_{ph}}\right), \text{ with } (2\pi f_i\tau_{ph})<<1, \qquad (167)$$

$$S_{V_G} \approx \frac{1}{f_i}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT}{WL} D_t(E_B) \Rightarrow D_t(E_B) = \left(\frac{C_{ox}}{q}\right)^2 WL \frac{fS_{V_G}(f)}{kT}, \tag{168}$$

where τ_{ph}~10^{-15}–10^{-14} s [2, 165] is the reciprocal of phonon attempt frequency at semiconductor-dielectric interface. It is observed that $D_t(E_B)$ is a peaking function at E_B~1eV in metal films, since many random processes occur with activation energies around this energy value [2]. In MOS transistors, different peaks are observed in the distribution of activation energies, which allows for the characterization of defects and traps that are the origin of the noise [165]. Also, an interesting relation between the temperature dependence of the magnitude $S(f_i,T)$ and the slope $\gamma(f_i,T)$ of the $1/f^{\gamma}$ flicker noise at given frequency f_i and temperature T is deduced in [2]. This relation is

$$\gamma(f_i,T) = 1 - \frac{1}{\ln(2\pi f_i \tau_{ph})}\left\{\frac{\partial \ln[S(f_i,T)]}{\partial \ln(T)} - 1\right\} = 1 + \frac{kT}{E_B}\left\{\frac{\partial \ln[S(f_i,T)]}{\partial \ln(T)} - 1\right\}, \tag{169}$$

and it is known as Dutta-Horn equation for the slope of flicker noise. The relation is valid at a condition that D_t varies slowly with E_B for an energy interval of kT, that is $|\partial \ln(D_t)/\partial E_B|kT<<1$, in order to neglect higher-order derivatives in the expansion of D_t in Taylor series. This condition is rarely checked in experiments, although is noted in [165] that this is a mandatory step, and in contrast to the results in [165], it is argued in [70] that for "regular" traps in silicon with a peak barrier energy of up to 0.3eV there might be not enough room to provide broad distribution for D_t varying slowly for many multiples of kT.

Now, we shall obtain the impact of barrier energy on the noise from traps located inside the gate insulator, in order to analyze the temperature dependence of flicker noise in MOS transistors, in which the tunneling mechanism for Δn noise is dominant. We will also show that the Dutta-Horn equation (169) follows from the corresponding equation (164) for gradual trap profiles. So, with respect to eq. (102) that $\tau \propto \exp(E_B/kT)\times\exp(x/\lambda)$, we write the superposition integral in more general form of

$$S_{V_G} = \left(\frac{q}{C_{ox}}\right)^2 \frac{1}{WL} \iint_{x,E_B} N_t \frac{4\tau_{ph}\exp\left(\frac{x}{\lambda}+\frac{E_B}{kT}\right)}{1+(2\pi f\tau_{ph})^2 \exp^2\left(\frac{x}{\lambda}+\frac{E_B}{kT}\right)} dx\, dE_B \tag{170}$$

Evidently, one cannot discriminate between distance x/λ and energy E_B/kT, and the integral convolutes them. However, one can define $\xi_B = x_B/\lambda = E_B/kT$ that transforms the energy and temperature into "apparent distance" x_B. Since ξ_B and x_B are not real distances, then they are constants by the integration in eq. (161) and scale the results via τ_o and integration limits, but otherwise preserving the form of the results. So, we use eq. (164) to derive Dutta-Horn equation.

The experimental conditions for the Dutta-Horn equation (169) are that the flicker noise of a device is measured at different temperatures, that is, T is changed, but the frequency of noise measurement is fixed to $f=f_i$=constant, and all other conditions are repeated unchanged at each temperature, e.g. the biasing, device parameters and experimental accuracy are virtually the same. According to eq. (104), $S_{V_G} \propto kTN_t/f$. Therefore

$$\ln(S_{V_G}) = \ln(N_t) + \ln(T) - \ln(f_i) + \text{constant}, \tag{171}$$

where f_i is also a constant in Dutta-Horn equation, as stated above. Therefore,

$$\partial\ln(N_t) = \partial\ln(S_{V_G}) - \partial\ln(T)\,. \tag{172}$$

Since f_i is a constant, then

$$\frac{1}{2\pi f_i} = \tau_i = \tau_{ph}\exp\left(\frac{x}{\lambda} + \frac{E_B}{kT}\right) = \text{constant}. \tag{173}$$

As far as the function $\dfrac{\tau_{ph}\exp(\xi+\xi_B)}{1+(2\pi f\tau_{ph})^2\exp^2(\xi+\xi_B)}$ is peaking at $f=f_i$, it is the main contribution in the superposition integral of eq. (161), and thus, in the spectrum of the flicker noise at $f=f_i$, as explained after eq. (156) and in [2]. Therefore,

$$\partial\left(\frac{x}{\lambda}\right) = \partial\xi = -\partial\xi_B = -\partial\left(\frac{E_B}{kT}\right) = \frac{E_B}{kT}\partial\ln(T)\,, \tag{174}$$

showing that the noise at f_i is probed from another distance in the oxide, when the temperature is changed. Substituting eqs. (172) and (174) into eq. (164)

$$\gamma = -\frac{\partial\ln(S_{V_G})}{\partial\ln(f)} = 1 + \frac{\partial\ln(S_{V_G}) - \partial\ln(T)}{\frac{E_B}{kT}\partial\ln(T)} = 1 + \frac{kT}{E_B}\left[\frac{\partial\ln(S_{V_G})}{\partial\ln(T)} - 1\right], \tag{175}$$

which is the same as the Dutta-Horn equation (169). The difference from the Dutta-Horn equation, which is for bulk materials and thin metal films, is following from the fact that for the case of oxide traps, the distance x from the semiconductor-dielectric interface also takes place, according to eq. (173). In particular, when substituting from eq. (155), one gets

$$\frac{E_B}{kT} = -\ln\left[2\pi f_i\tau_{ph}\exp\left(\frac{x}{\lambda}\right)\right] = -\ln\left(\frac{\tau_{ph}}{\tau_o}\right) \quad \text{, and} \tag{176}$$

$$\gamma = -\frac{\partial\ln(S_{V_G})}{\partial\ln(f)} = 1 + \frac{1}{\frac{E_B}{kT}}\left[\frac{\partial\ln(S_{V_G})}{\partial\ln(T)} - 1\right] = 1 + \frac{1}{\ln(\tau_o/\tau_{ph})}\left[\frac{\partial\ln(S_{V_G})}{\partial\ln(T)} - 1\right] \tag{177}$$

Since the tunneling time constant prefactor τ_o is in the range $10^{-7}s^{-1}$ [134] to $10^{-10}s^{-1}$ [49, 55, 140], while τ_{ph}~10^{-15}–10^{-14} s [2, 165], then $\ln(\tau_o/\tau_{ph})$~9…18, and is not so large as compared the original estimate in [2] of $|\ln(2\pi f_i\tau ph)|$~35…40 for metals measured in frequency range between 10Hz and 10kHz. Thus, traps with barrier capture energy E_B<0.3eV may actually cause variations of the slope of 1/f noise in MOS transistors via number fluctuation, despite the concerns in [70] for limited room for the distribution of barrier energy of traps in silicon. Indeed, when recalling the duality between tunneling distance and energy, the variation of 1/f noise slope can be due to spatial non-uniformity of oxide trap density N_t, rather than in distribution in energy, although the former will cause the temperature variation. The random telegraph noise in MOS transistors discussed in the next section also leads to this conclusion. Furthermore, it is shown in [73] that the energy distribution can be inspected by varying the gate bias, instead of temperature, although in a very narrow range of few meV and with very low value for E_B~50meV.

IV.3. RTS noise in MOS transistors

As the size of MOS transistors is becoming smaller and smaller, then the charge capture and emission process by individual traps is increasingly distinguishable. This process results in a random bistable fluctuation in time domain, such as Random Telegraph Signal (RTS), and therefore the random bistable fluctuation is also called RTS noise. The RTS noise in the drain current of MOS transistors is usually measured and then referred to the gate terminal as a voltage by using the transconductance g_m of the transistor [74]. It is cumulatively observed that the amplitude of the gate referred voltage from individual RTS noise matches well with addition and removal of one elementary electronic charge q to the gate oxide capacitance [74], that is,

$$\frac{\Delta I_D}{g_m} = \Delta V_G = \frac{q}{C} = \frac{q}{WLC_{ox}} \quad . \tag{178}$$

IV.3.1. Time constants of RTS noise

The time constants of the two states of individual RTS match with the predictions of Shockley–Read–Hall theory for generation-recombination process [68, 74, 107, 112, 148, 166]. Consequently, the superposition of larger number of individual RTS, each of which having a Lorentzian spectrum, is found to coincide with 1/f noise in larger-area MOS transistors [68, 74, 112, 148]. The above findings are made over a period of about 50 years and has been reviewed in [68] at the time of entering the sub-micrometer technologies. In studies of MOS transistors with very thin oxides, RTS noise is observed also in the gate leakage current [148].

The above summary of findings implies a coherent picture for RTS noise in MOS transistors. However, looking at the details or individual measurements, one will observe significant deviations and will meet with difficulties to manage long time records, and to link them to spectra and models for noise. In fact, there is no standard procedure for analysis and compact model for RTS noise in MOS transistors. The RTS noise is also non-monotonically bias- and temperature dependent [58, 68, 137, 161, 167], it varies between different time records captured from one sample [68, 74], and between nominally identical samples [137, 166]. On the other hand, the amplitudes of RTS noise can be large in sub-micron area MOS transistors [74, 112, 137] and RTS noise becomes important issue that the designs have to overcome, e.g. in CMOS imagers with correlated double sampling [162].

Generally, three parameters describe the individual RTS noise component. For MOS transistors, these are amplitude ΔI_D (or ΔV_G), and capture and emission time constants of the trapping GR center, respectively. The studies in [68, 74, 107, 112, 148, 166] confirmed that the time constants follow very well the Shockley–Read–Hall (SRH) statistics, which established eq. (63). In most cases, the emission time constant τ_e is a weak function of the biasing (it increases when increasing V_G-V_T), whereas the capture time constant τ_c decreases when the biasing (V_G-V_T) increases, owing to the increase of the carrier concentration n' in the channel of the MOS transistors. There is a possible exception in sub-threshold regime of operation of the MOS transistor when charges from the bulk are trapped, and the bias dependence of the time constants can be inverted for this case [161]. There are experimental difficulties by obtaining the values for time constants since a processing of long records captured in time domain is necessary, especially when two or more RTS are present simultaneously [68, 74, 112], but overall there is no other significant issue with the RTS time constants. When the values are obtained, they match well the predictions of the theory and with the corner frequency of Lorentzian spectrum related to RTS, and given by

$$S_{I_D} = \frac{4\Delta I_D^2 F(1-F)\tau}{1+(2\pi f\tau)^2} \text{ or } S_{V_G} = \frac{4\Delta V_G^2 F(1-F)\tau}{1+(2\pi f\tau)^2}, \text{ with } \frac{1}{\tau} = \frac{1}{\tau_e} + \frac{1}{\tau_c} \text{ and } F(1-F) = \frac{\tau_e \tau_c}{(\tau_e + \tau_c)^2}. \quad (179)$$

where F is the trap occupancy factor, given by Fermi-Dirac statistics. Therefore, we do not extend the discussion for the time constants of RTS further.

IV.3.2. Amplitude of RTS noise

There are some discrepancies between values for the RTS amplitudes ΔI_D and ΔV_G, measured in time domain, estimated from spectrum using eq. (179) and predicted by eq. (178). We focus on this issue, because the amplitudes of RTS noise will increase as the device size decreases with every next generation of technology nodes of minimum feature gate length L_{min}, since RTS due to individual traps will dominate the low-frequency noise and, according to eq. (178), ΔV_G and ΔI_D will increase as $1/(WL) \propto 1/L_{min}^2$, because C_{ox} cannot be increased proportionally furthermore [3].

The characterization of MOS transistors usually uses eq. (178) rewritten in normalized form for the RTS current, given by

$$\frac{\Delta I_D}{I_D} = \frac{g_m}{I_D}\Delta V_G = \frac{g_m}{I_D}\frac{q}{C} = \frac{g_m}{I_D}\frac{q}{WLC_{ox}}, \quad (180)$$

where $\Delta V_G = q/(WLCox) = \Delta V_{FB}$ is regarded as modulation flat band voltage V_{FB} in MOS transistor owing to trapping of single electron at semiconductor-dielectric interface [58, 156]. This equation corresponds to the generic expression for coupling of noise, being a square-rooted version of eq. (1) with coupling coefficient $\sqrt{K}=1$. The direct application of this equation against measured data usually shows some discrepancies, as illustrated in Figure 29. One can see in the figure that the overall evolution of the RTS amplitude is captured by eq. (180), but the amplitude is underestimated at low bias, whereas it is overestimated at high bias. This implies that the coupling between trap and conduction layer involves also other factors, and the coupling coefficient $\sqrt{K} \neq 1$ is also bias dependent. The reasons of this discrepancy correspond to the issues with the Δn model for 1/f noise in MOS transistors, since the origin of the models is the same.

Correlation between mobility variation and trap occupancy

One effect neglected in eq. (180) is the mobility variation correlated to the trap occupancy. This correlation was investigated in [58] for RTS amplitude, considering Coulomb and phonon scattering at lower and higher biasing of MOS transistor. The essence of this investigation is that, increasing the bias from weak to strong inversion, the channel carrier mobility μ increases at low bias, since μ is limited by Coulomb scattering [see eqs. (134) and (135)], whereas the channel carrier mobility μ decreases at high bias, since μ is limited by phonon and roughness scattering [see eqs. (131) and (133)]. Consequently, via the surface potential, at low bias the derivative $\partial\mu/\partial\Delta V_{FB} > 0$ and adds to the change $\partial n'/\partial\Delta V_{FB} > 0$ in carrier concentration n' in the MOS channel, causing the RTS amplitude to be higher than that given by eq. (180), whereas at high bias the derivative $\partial\mu/\partial\Delta V_{FB} < 0$ and subtracts from $\partial n'/\partial\Delta V_{FB} > 0$, causing the RTS amplitude to be lower than that given by eq. (180). At a crossover bias ($V_{G,cr}$, $I_{D,cr}$, $g_{m,cr}$), when $\partial\mu/\partial V_G \approx 0$, eq. (180) matches the measurement. Noticeably, we can observe qualitatively exactly the same behavior in Figure 29, as well as in fig. 8 in [107], and to the first order of approximation we can suggest an empirical expression for coupling coefficient $\sqrt{K_\mu}$ for correlated mobility contribution to RTS amplitude that is in the form of eq. (121) for the coupling coefficient for 1/f noise in MOS

transistors with correlated Δn-Δμ fluctuation.

$$\sqrt{K_\mu} \approx 1+\theta\left(\frac{I_D}{g_m}-\frac{I_{D,cr}}{g_{m,cr}}\right)\rightarrow\frac{\Delta I_D}{I_D}=\frac{g_m}{I_D}\Delta V_{FB}\sqrt{K_\mu}=\frac{g_m}{I_D}\frac{q}{WLC_{ox}}\sqrt{K_\mu}\rightarrow\Delta V_G=\Delta V_{FB}\sqrt{K_\mu} \quad (181)$$

For physical justification, the ratio $I_{D,cr}/g_{m,cr}\propto C_{ox}(V_{G,cr}-V_T)\propto n'_{cr}$ corresponds to carrier concentration n'_{cr} in the channel at which crossover between Coulomb and phonon scattering occurs and $\partial\mu/\partial V_G\approx 0$.

Position of the trap along the channel and depth of the trap in the oxide

A second detail neglected in eq. (180) is the position of the trap in the oxide and in the channel.

Position of the trap along the channel. Considering the channel conduction at different spots under the gate, it is deduced in [167] that the RTS amplitude is function not only on channel carrier concentration n', but also on the lateral electric field at the position of the trap along the channel. Assuming that one carrier is trapped at a particular spot in the conductive channel of the MOS transistor, the relative RTS amplitude is then given by [167]

$$\frac{\Delta I_D}{I_D}=\frac{\Delta G_D}{G_D}=\frac{\mu E^2}{\mu_{avg}E_{avg}^2}\frac{1}{n}\ , \quad (182)$$

where $G_D=I_D/V_D$ is the channel conductance between drain and source terminals, ΔI_D and ΔG_D are the RTS amplitudes, μ and E are local carrier mobility and lateral electric field in the channel at the spot of charge trapping, μ_{avg} and E_{avg} are average values for carrier mobility and lateral electric field carrier, and n is the total number of carriers in the channel, respectively. By inspection of MOS transistor equations, it can be shown that $1/n\approx g_m q/(I_D WLC_{ox})$. Therefore, eq. (182) can be rewritten in the form of eqs. (180) and (181), as

$$\frac{\Delta I_D}{I_D}=\sqrt{K_E}\frac{g_m}{I_D}\frac{q}{WLC_{ox}}=\frac{\mu E^2}{\mu_{avg}E_{avg}^2}\frac{g_m}{I_D}\frac{q}{WLC_{ox}},\ \text{with}\ \sqrt{K_E}\approx\frac{\mu E^2}{\mu_{avg}E_{avg}^2} \quad (183)$$

being a coupling coefficient corresponding to variation of the lateral electric field E in MOS channel. Obviously, the lateral electric field at the source side of the channel is lower than E_{avg} and traps located at this side will produce RTS with amplitude lower than that predicted by eq. (180). In contrary, the lateral electric field at the drain side of the channel is higher than E_{avg} and traps located at the drain side will produce RTS with amplitude higher than that predicted by eq. (180), especially when the MOS transistor is in saturation mode of operation. Perhaps this is the reason why the RTS amplitudes from different traps scatter in one sample and between identical samples, as illustrated in Figure 30 from [68].

In this figure, 58 different RTS are shown from 12 nominally identical nMOS transistors biased at the same condition $(V_G-V_T)=(3.8-1)$V, $V_D=0.1$V, $V_S=0$, $I_D=6.7$μA. The transistors are from 2.5μm technology node, but the mask was W=L=2μm being below the minimum feature size. From DC measurements, it was estimated that the effective size of the samples was W=0.5μm and L=0.75μm, but tolerances are not reported in [68]. The circles in Figure 30 present the relative RTS amplitudes $\Delta I_D/I$ organized in [Fig. 16 in 68] in a scatter plot sorted by increased values of $\Delta I_D/I_D$ and formally numbered from 1 to 58, as shown in the bottom horizontal axis of Figure 30. We assume that the traps are uniformly distributed along the channel length L, so number 0 corresponds to source edge of the channel and number 59 corresponds to drain edge of the channel. In this way,

we scale the top horizontal axis of Figure 30, using z/L=(Number of RTS)/59, where (z) represents the position of the trap along the channel length L. Next assumption is that the carrier concentration in the channel is nearly constant, since $(V_G-V_T)\approx 2.8V$, while $(V_D-V_S)=V_D=0.1V$ and the transistors were operating deep in ohmic regime. Therefore, we expect linear increase of lateral electric field E from source to drain with an average value of $E_{avg}=V_D/L\approx 1.3kV/cm$. According to eqs. (182) and (183), the square root $\sqrt{\Delta I_D/I_D} \propto E/E_{avg}$, and by multiplying with the value above for E_{avg}, we obtain the lateral electric field E for each data point, as shown with squares in Figure 30, which are fitted with the linear function $E\propto z/L$. This linear-fit function was used in eq. (183) to calculate $\Delta I_D/I_D\propto(E/E_{avg})^2$, as shown with thick line passing through the circles in Figure 30, estimating also $C_{ox}\approx 95nF/cm^2$ and $EOT\approx 37nm$, which are not stated in [68], but are reasonable for 2.5μm technology node. Apart from 4 data points of high RTS amplitudes on the right side of Figure 30, the agreement between measured (circles) and calculated (thick line) from eq. (183) values for the relative RTS amplitudes $\Delta I_D/I_D$ is good, despite the many assumptions stated above. The agreement leads to the conclusion that the variation of RTS amplitude in MOS transistors is vulnerable to the position of the trap along the channel, since the lateral electric field is different at different positions. The 4 data points that deviate from the rest of the data in Figure 30 are probably from one of the 12 devices. The sizes of this device probably deviate 20%-30% from the nominal values, which is reasonable for the case of transistors with effective sizes L and W being about 1/4-1/5 of the minimum feature size of the technology node. Unfortunately, the information in [68] is aggregated among all 12 samples, and we cannot inspect the details further.

The analysis above for relation between trap position along the MOS transistor channel and amplitude of RTS noise is in general agreement with other works. It is argued in [133] that the flat band perturbation, which is in the origin of eq. (180) via $\Delta V_G=q/(WLC_{ox})=\Delta V_{FB}$, neglects the local perturbation in the surface potential and the associated charge carrier density at a particular position in the channel, and thus underestimates both the RTS amplitude and 1/f noise as a superposition of RTS, when the transistor operates in strong inversion and saturation, i.e. inversion layer charge densities at source and drain sides are $n'_S>2\varphi_t C_{ox}/q$ and $n'_D<<n'_S$. At this condition, as mentioned earlier, the 1/f noise magnitude is underestimated about 2 times by the flat band perturbation model. For RTS amplitude, the difference can be larger, and this was used to locate positions of traps in the channel in [166]. The methods in [166] require knowledge for details in the transistor and employ formulation in terms of surface potential and numerical iterations. This approach, although accurate, is too complicated for the purpose of experimental noise characterization, in which the measured data scatter, and in this way, do not allow for obtaining very precise modeling. The advantage of the methods in [166] is that the position of the trap along the channel can be found simultaneously with the depth of the trap in the oxide. Now we discuss the latter.

Depth of the trap in the oxide. Based on earlier works [168], it is suggested in [112] that the relative RTS amplitude $\Delta I_D/I_D$ can be given as

$$\frac{\Delta I_D}{I_D}=\eta\frac{g_m}{I_D}\frac{q}{WLC_{ox}}\left(1-\frac{x_{ti}}{t_{ox}}\right), \tag{184}$$

where t_{ox} is the physical thickness of the gate insulator, x_{ti} is the distance from semiconductor-insulator interface to the position of the trap in the oxide. The factor $(1-x_{ti}/t_{ox})=(t_{ox}-x_{ti})/t_{ox}$ is a coupling coefficient $\sqrt{K_{ti}}$ and it is the same as in eq. (147). The parameter η corresponds here to the product of coupling coefficients $\sqrt{K_\mu}\sqrt{K_E}$ in eqs.

(181) and (183) above, but it is left as a fitting parameter in [112]. Therefore, in order to obtain information for $(t_{ox}-x_{ti})/t_{ox}$, the ratio of the capture τ_c and emission τ_e time constants is used in [112], since τ_c and τ_e depend on x_{ti}, see eq. (112), and on bias via Fermi level, see eq. (63). The explicit expressions are given in [166], but the overall bias dependence is [112]

$$\frac{\partial \ln(\tau_c/\tau_e)}{\partial V_G} = -\frac{1}{\varphi_t}\frac{x_{ti}}{t_{ox}} - \frac{1}{\varphi_t}\left(1-\frac{x_{ti}}{t_{ox}}\right)\frac{\partial \psi_s}{\partial V_G}, \tag{185}$$

where $\varphi_t=kT/q\approx 0.026V$ is the thermal voltage at room temperature T=300K, V_G is the gate bias voltage and ψ_s is the surface potential in the MOS channel. The surface potential ψ_s is a complicated function of V_G (and other biasing voltages applied to MOSFET), and the precise evaluation of the derivative $\partial\psi_s/\partial V_G$ is inconvenient and not necessary when the spread in the values for τ_c and τ_e is large, which is usually the case of noise measurements. Therefore, for practical cases of characterization, one can obtain an approximate value for the derivative $\partial\psi_s/\partial V_G$, using several general relations in MOS transistors from [169], as follows, in which inversion charge Q_{inv}, oxide capacitance C_{ox} and depletion capacitance C_d are given per unit area.

Assume that the measurement was when the MOS transistor operated either in weak (sub-threshold) or in strong (above threshold) inversion regimes, that is $|V_G-V_T|>0.1V$.

In weak inversion, when V_G is below the threshold voltage V_T, the inversion charge Q_{inv} is negligible and the gate bias is spread across the series connection of oxide capacitance C_{ox} and depletion capacitance C_d. Therefore [pp.75 and 86 in 169]

$$\frac{\partial V_G}{\partial \psi_s} = 1+\frac{\gamma}{2\sqrt{\psi_s}} = \alpha = 1+\frac{C_d}{C_{ox}} \approx \text{constant} \approx 1...1.5\text{, in weak inversion,} \tag{186}$$

neglecting the interface states' capacitance C_{it}. The parameter α can be obtained from sub-threshold slope $\partial V_G/\partial \log_{10}(I_D)=2.3\alpha\varphi_t$ [p.175 in 169] of the transfer characteristic $I_D(V_G)$. The parameter γ is the body bias coefficient and can be obtained from $\partial V_G/\partial V_B$, where V_B is a bias voltage applied to the bulk (body) under the MOS transistor channel. However, γ is not needed in this analysis.

At the same time, the ratio g_m/I_D between transconductance g_m and drain current I_D in MOSFET in weak inversion also depends on α, and it is given by [p. 173 in 169]

$$\frac{g_m}{I_D} = \frac{1}{\alpha\,\varphi_t} \tag{187}$$

So, from eqs.(186) and (187), we get that

$$\frac{\partial \psi_s}{\partial V_G} = \frac{1}{\alpha} = \varphi_t\frac{g_m}{I_D} \approx \text{constant} \approx 0.6...1\text{, when the MOSFET is in weak inversion regime.} \tag{188}$$

For strong inversion regime, according to [p.87 in 169], we write

$$\frac{\partial V_G}{\partial \psi_s} = \frac{\partial V_G}{\partial \ln(Q_{inv})}\frac{\partial \ln(Q_{inv})}{\partial \psi_s} = Q_{inv}\frac{\partial V_G}{\partial Q_{inv}}\frac{1}{2\varphi_t} \tag{189}$$

$$\frac{\partial \psi_s}{\partial V_G} = \frac{2\varphi_t}{Q_{inv}}\frac{\partial Q_{inv}}{\partial V_G} \tag{190}$$

in linear (ohmic) mode, when $(V_D-V_S)<<(V_G-V_T)$, and where $V_{DS}=(V_D-V_S)$ is drain-source bias voltage and V_T is threshold voltage of MOS transistor. Also, the inversion layer charge Q_{inv} is approximately constant along the channel, and it is

$$Q_{inv} \approx C_{ox}\left(V_G - V_T\right) = \frac{I_D}{\mu}\frac{L}{W}. \tag{191}$$

where μ is the carrier mobility in the transistor channel of length L and width W. So, eq. (190) becomes

$$\frac{\partial \psi_s}{\partial V_G} = \frac{2\varphi_t}{\frac{I_D}{\mu}\frac{L}{W}}\frac{\partial}{\partial V_G}\left(\frac{I_D}{\mu}\frac{L}{W}\right) = \frac{2\varphi_t}{I_D}\frac{\partial I_D}{\partial V_G}, \text{ or} \tag{192}$$

$$\frac{\partial \psi_s}{\partial V_G} = 2\varphi_t \frac{g_m}{I_D} \approx \varphi_t \frac{2}{V_G - V_T}, \text{ in strong inversion and ohmic mode.} \tag{193}$$

To obtain an expression for $\partial\psi_s/\partial V_G$ in at strong inversion regime saturation mode, we look closer at the charge sheet model for the drain (to source) current I_D with neglected diffusion, given by [p.157 in 169]

$$I_D = \frac{W}{L}\frac{\mu}{2\alpha C_{ox}}\left(Q_{inv,S}^2 - Q_{inv,D}^2\right), \tag{194}$$

in which the inversion layer charge densities $Q_{inv,S}$ and $Q_{inv,D}$ are at the source and drain sides of the channel, respectively, and $Q_{inv,S}$ and $Q_{inv,D}$ are given with the bias voltages V_S and V_D applied to these sides, as [p.121 in 169]

$$\begin{aligned} &Q_{inv,S} = \alpha C_{ox}\left(V_P - V_S\right), \text{if } V_S < V_P, \text{otherwise } Q_{inv,S} = 0, \text{ and} \\ &Q_{inv,D} = \alpha C_{ox}\left(V_P - V_D\right), \text{if } V_D < V_P, \text{otherwise } Q_{inv,D} = 0. \end{aligned} \tag{195}$$

The obvious approximation for the pinch-off voltage V_P in the equation above is given by [p.158 in 169]

$$V_P \approx \frac{V_G - V_T}{\alpha}. \tag{196}$$

The transistor is in saturation mode, if $V_D>V_P$, and in ohmic mode, if $V_D<V_P$. Note, I_D, $Q_{inv,S}$ and $Q_{inv,D}$ depend on V_G only via V_P, and α is assumed constant, because the depletion capacitance C_d in eq.(186) does not change significantly with V_G, especially in strong inversion.

In saturation mode, $(V_D>V_P)$ and $Q_{inv,S}>>Q_{inv,D}\approx 0$. Therefore, eq. (194) is reduced, and using $V_S=0$, which is the normal case when the body and source of MOSFET are tied together, one gets

$$Q_{inv} \approx Q_{inv,S} = \sqrt{\frac{I_D}{\mu}\frac{L}{W}2\alpha C_{ox}}, \tag{197}$$

$$\frac{\partial Q_{inv}}{\partial V_G} = \frac{\partial}{\partial V_G}\left(\sqrt{\frac{I_D}{\mu}\frac{L}{W}2\alpha Cox}\right) = \frac{1}{2}\sqrt{\frac{2\alpha Cox}{\mu I_D}\frac{L}{W}}\frac{\partial I_D}{\partial V_G} = \frac{g_m}{2}\sqrt{\frac{2\alpha Cox}{\mu I_D}\frac{L}{W}}, \tag{198}$$

and

$$\frac{\partial Q_{inv}}{\partial V_G} = \frac{g_m}{2I_D}\sqrt{\frac{I_D}{\mu}\frac{L}{W}2\alpha Cox} = \frac{g_m}{2I_D}Q_{inv}. \tag{199}$$

This means that one can substitute in eq. (190) the quantity

$$\frac{1}{Q_{inv}}\frac{\partial Q_{inv}}{\partial V_G}=\frac{g_m}{2I_D}, \tag{200}$$

and to obtain

$$\frac{\partial \psi_s}{\partial V_G}=2\varphi_t\frac{g_m}{2I_D}=\varphi_t\frac{g_m}{I_D}\approx\varphi_t\frac{2}{V_G-V_T}\text{, in strong inversion and saturation mode.} \tag{201}$$

By comparing eqs.(188), (193) and (201), one can summarize that

$$\frac{\partial \psi_s}{\partial V_G}\approx\varphi_t\frac{g_m}{I_D}\text{, in any mode of operation of MOSFET,} \tag{202}$$

$$\frac{\partial \psi_s}{\partial V_G}\approx 0.6\ldots 1\text{, in weak inversion regime (sub-threshold, } V_G \text{ is below } V_T\text{),} \tag{203}$$

and $\partial\psi_s/\partial V_G$ gradually decreases at gate biasing above threshold, as

$$\frac{\partial \psi_s}{\partial V_G}=\frac{2\varphi_t}{V_G-V_T}\text{, in strong inversion regime (above threshold, } V_G \text{ above } V_T\text{).} \tag{204}$$

The accuracy of eqs. (202), (203) and (204) is within a factor of 2 (6dB), which is sufficient for analysis of RTS noise measurement data, which scatter usually more. So, eq. (185) can be rewritten to useful for experimental characterization forms of

$$\frac{\partial \ln(\tau_c/\tau_e)}{\partial V_G}=-\frac{1}{\varphi_t}\frac{x_{ti}}{t_{ox}}-\left(1-\frac{x_{ti}}{t_{ox}}\right)\frac{g_m}{I_D}\text{, in any mode of operation of MOSFET, or} \tag{205}$$

$$\frac{\partial \ln(\tau_c/\tau_e)}{\partial V_G}=-\frac{1}{\varphi_t}\frac{x_{ti}}{t_{ox}}-\frac{1}{\varphi_t}\left(1-\frac{x_{ti}}{t_{ox}}\right)\frac{1}{\alpha}\approx-\frac{1}{\varphi_t}\left(1+2\frac{x_{ti}}{t_{ox}}\right)\text{, in weak inversion, and} \tag{206}$$

$$\frac{\partial \ln(\tau_c/\tau_e)}{\partial V_G}=-\frac{1}{\varphi_t}\frac{x_{ti}}{t_{ox}}-\left(1-\frac{x_{ti}}{t_{ox}}\right)\frac{2}{V_G-V_T}\text{, in strong inversion.} \tag{207}$$

The equations include parameters that can be easily obtained from DC measurements of the MOSFET transfer characteristic (V_T, g_m, α from sub-threshold slope). So, once the capture and emission times are characterized from RTS noise measurements at several bias points for V_G, then the derivatives in the left-hand side of the equations can be found by a simple fitting of slopes in semi-log plot of $\ln(\tau_c/\tau_e)$ vs. V_G. Then, by substituting in one of eqs. (205), (206) or (207), the ratio x_{ti}/t_{ox} of the distance x_{ti} from the trap to semiconductor-insulator interface to the physical thickness t_{ox} of the gate insulator can be estimated, and the coupling coefficient $\sqrt{K_{ti}}$ for the trap position in the oxide can be found as

$$\sqrt{K_{ti}}=\left(1-\frac{x_{ti}}{t_{ox}}\right). \tag{208}$$

Note that the above equations are appropriate for approximate experimental characterization and volume tests when the scattering in the data is large. If a precision analysis of few samples is required, one should use rigorous models, such as this in [166], which can also locate the position of the trap along the channel of

MOSFET. The price is, of course, that the surface potential ψ_s has to be obtained prior to analysis of the oxide trap, and this requires specific information, numerical simulations and optimizations, which might be not always available, or affordable, owing to time, expertise or other constraint in the experimental practice.

Amplitude of RTS (concluding remarks). To summarize, the analysis of amplitude of RTS noise in MOS transistor requires large number of waveforms captured in time domain, from several samples and various biasing. The waveforms need to be processed so that the evolution of the amplitudes and time constants of RTS from individual traps with biasing and over the samples is observed. Then, the data have to be fitted to

$$\frac{\Delta I_D}{I_D} = \sqrt{K}\frac{g_m}{I_D}\frac{q}{WLC_{ox}} = \sqrt{K}\frac{1}{n}\ , \tag{209}$$

where the coupling coefficient $\sqrt{K}=\sqrt{K_\mu}\sqrt{K_E}\sqrt{K_{ti}}$ accommodates several dependences, and n is the total number of carriers in the channel of the MOS transistor, with $n\approx WLC_{ox}(V_G-V_T)/q$ in the ohmic regime of operation of the MOS transistor at $V_D<<(V_G-V_T)$, and twice smaller $n\approx ½WLC_{ox}(V_G-V_T)/q$ in the saturation regime of operation of the MOS transistor at $V_D\geq(V_G-V_T)$.

The coupling coefficient $\sqrt{K_\mu}$ takes into account for the contribution of mobility variation, $\sqrt{K_\mu}$ is given by eq. (181) and the parameters in this equation should be chosen so that the modeled and measured relative RTS amplitudes $\Delta I_D/I_D$ are proportional when the gate bias voltage V_G is varied. As the initial values in eq. (181), one can use the first order mobility degradation coefficient $\theta\sim 0.3\ldots 1\ V^{-1}$ and the ratio $I_{D,cr}/g_{m,cr}\propto C_{ox}(V_{G,cr}-V_T)$ of the current to transconductance at gate bias corresponding to maximum mobility in strong inversion regime ($V_{G,cr}$ above threshold V_T).

The coupling coefficient $\sqrt{K_E}$ takes into account for the variation of the lateral electric field and carrier velocity in the channel of MOS transistor, and $\sqrt{K_E}$ is given by eq. (183). If data from many samples measured in strong inversion and ohmic modes are available, then the value for $\sqrt{K_E}$ can be obtained from a scatter plot of relative RTS amplitudes $\Delta I_D/I_D$ as discussed above, and the position of the traps along the channel length can be estimated. Otherwise, knowledge for the lateral electric field and mobility has to be provided by other means, e.g. simulation of structure, in order to obtain surface potential along the channel [166]. Other approaches are also possible, e.g. swapping drain and source, to discriminate between traps located closer to source or closer to drain sides of the channel – see references in [166].

The coupling coefficient $\sqrt{K_{ti}}$ takes into account for the depth x_{ti} of the trap in the gate insulator of thickness t_{ox}. At assumption for tunneling mechanism for charge exchange between the trap and channel carriers, the ratio x_{ti}/t_{ox}, can be estimated using eqs. (205), (206) and (207) and DC parameters of MOS transistor from the gate bias dependence of capture τ_c and emission τ_e time constants of individual RTS, by the slope of $\partial\ln(\tau_c/\tau_e)/\partial V_G$. Then, $\sqrt{K_{ti}}= (1-x_{ti}/t_{ox})$, as given by eq. (208).

The advantage of the procedure above is that it is based only on DC measurements and waveform captures, and device simulation is not needed. However, the procedure requires large number of measurements and extensive processing of waveforms by establishing relation between data from different measurements and keeping track of the evolution of RTS parameters of individual traps. Nevertheless, tools and methods for semi-automated extraction of RTS parameters are reported (in [74, 107, 112, 166] using correspondence between ΔI in time domain and low-frequency plateau of Lorentzian spectrum in frequency domain, in [58, 74, 107] using histograms of drain current, in [28, 29] using discontinuity of waveform), although they should be further

organized to make the tests feasible for applications in the volume production in semiconductor industry.
The above analysis of RTS noise is based on charge trapping. Alternative suggestion for RTS noise from mobility fluctuation is given in [170] for diodes, at condition when $I<\Delta I<q/\alpha_H\tau$ where τ is minority carrier life time and α_H is the Hooge parameter – see eq. (6). Situation $I<\Delta I$ of DC current smaller than RTS amplitude has been observed in carbon nanotube pMOS-like field-effect transistors [171].

IV.4. Figures of merit for MOS transistors

The intensive research on MOS devices resulted in many figures of merit (FOM) that have been used to compare different technologies and explore the scaling of these devices. Each FOM was suggested in order to emphasize particular feature of the devices or circuit, which used these devices, or to facilitate a model or design. Consequently, the values from different FOM became difficult to compare each to other. In this section, we will discuss several FOM that are commonly used for the low-frequency noise in MOS transistors in attempt to relate the different FOM each to other.

IV.4.1. Definition

To set up the discussion, first it is helpful to state what FOM is and how it differs from device parameters or physical quantities. In principle, FOM is a customized expression that combines several device parameters and physical quantities according to particular model or targeting particular application. For example, the DC value I_D of the drain current in MOS transistor and the power spectrum density S_{I_D} of drain current are physical quantities, but the normalized noise S_{I_D}/I_D^2 is a figure of merit for the ratio noise to DC in the transistor, and S_{I_D}/I_D^2 may (or may not) vary with frequency and bias, depending on what noise in a transistor is addressed. If the flicker noise is the concern, then one assumes a model with 1/f scaling rule for S_{I_D} and can evaluate the SPICE parameter $K_F=fS_{I_D}/I_D^2$ according to eq. (2), but K_F is also a FOM, since it is derived from another figure of merit, by using additional scaling rule for the frequency dependence of S_{I_D} and the 1/f dependence is canceled in K_F just for convenience. Certainly, K_F is a good figure of merit for flicker noise in BJT, since in most cases the 1/f noise is coupled from IFO via the transconductance g_m, as discussed in section III.2. "**Differences in BJT fabrication, IFO**", and $g_m/I_C\approx 1/\varphi_t\approx$constant in wide range of biasing conditions. However, K_F is not the best choice for MOS transistors in strong inversion regime, since $g_m/I_D\propto 1/(V_G-V_T)$ and the bias dependence of 1/f noise is not cancelled in K_F. On the other hand, if the shot noise in BJT is addressed at higher frequencies, then the appropriate figure of merit for normalized noise is $S_{I_C}/I_C\sim 2q$, which is frequency independent in principle, but not K_F. Moving further to RF range, one usually uses the so-called "Noise Figure" or "Noise Factor", which is a completely different FOM in its basis, and RF Noise Figure is a function of the ratio between device and thermal (Nyquist) noise at certain conditions for impedance matching. Thus, the different FOM depend on device, models and ranges, and one should explicitly state the expressions, the origin and the purpose of the normalization used for particular case of interest, since all normalizations are FOM, they are valid at certain conditions, and even the most popular FOM have counterparts.

IV.4.2. Input and output referred noise – scalability of normalized noise

MOS transistor

In both cases, the word is for the 1/f noise S_{I_D} in the drain current I_D, which is at the output terminal (drain) of the MOS transistor, and it should be clearly stated that noise S_{I_G} in the gate leakage I_G of MOS transistors with

ultra thin oxides is not included in the input and output FOM, and it is separately analyzed, if present, with one exception in [110].

The practice is that output noise S_{I_D} is referred to the input (gate) terminal as a "gate voltage" S_{V_G} by using the most general expression for coupling from input to output via transconductance g_m, which is

$$S_{I_D} = g_m^2 S_{V_G} , \tag{210}$$

and it follows from eq. (1). Consequently, S_{V_G} is a FOM, because it is a derived quantity, it cannot be directly measured as a voltage, but it is convenient, since it is weakly dependent on the bias of MOS transistor, it scales properly with the area WL of the transistor and it has been well explained in terms of oxide trapping, oxide capacitance, mobility degradation, etc., as discussed in previous sections. To compare different MOS transistors, the most popular FOM is the product WLS_{V_G} at low gate overdrive (V_G-V_T)~0.1V and frequency 1Hz, which is used in ITRS [3] – see eq. (106) for more details. At these conditions and at lower gate overdrive, the number fluctuation due to trapping at the oxide dominates, and $S_{V_G} \approx S_{FB}$, where S_{FB} is regarded as flat band voltage noise – see eq. (109). At higher gate overdrive, $(V_G-V_T)>0.2$V, the mobility fluctuation may significantly contribute to the noise. As explained in [156], if the mobility is correlated to the oxide trapping, then S_{V_G} becomes a quadratic function of $(V_G-V_T) \propto I_D/g_m$, and an additional FOM=$(S_{V_G})^{0.5}$ vs. (V_G-V_T) is used to obtain the carrier scattering coefficient [57, 89, 92, 139, 156] – see eq. (121) and the paragraph after it. If the mobility noise is not correlated to oxide trapping, then $S_{V_G} \propto (V_G-V_T)$ and the slope of this dependence is used to obtain the Hooge parameter α_H, which follows from to eqs. (11) and (88), and it will be discussed further in IV.4.4. "***Physical figures – trap density, Hooge parameter, scattering parameter***" along with other FOM derived from S_{V_G}.

The normalization of output noise is a FOM of form

$$\frac{S_{I_D}}{{I_D}^2} = \left(\frac{g_m}{I_D}\right)^2 S_{V_G} \quad . \tag{211}$$

It is widely used for MOS transistors as the intermediate step in analyses, such as for inspection of number fluctuation, or obtaining S_{V_G} and other quantities. However, this normalized noise is rarely used to obtain the SPICE parameter K_F except for cases when comparing noise in MOS and BJT at similar biasing current targeting specific circuit application, e.g. low power RF oscillators [22, 23], BiCMOS circuits [83] and radiation resistance or sensitivity [102, 103, 104, 165].

Comparison BJT-MOS transistors

As mentioned above, K_F is an essential FOM for BJT, because it varies a little in wide range of biasing, but the problem is that K_F in MOS transistors is a strong function of biasing above threshold voltage V_T, and it is approximately a reciprocal function of gate overdrive (V_G-V_T). Strictly speaking, comparisons based on K_F in MOS transistors are valid only at fixed gate overdrive. The better way to compare MOS and BJT is to refer K_F in BJT to the base terminal as a voltage noise S_{V_B}, and then to compare S_{V_B} and S_{V_G} as suggested in ITRS [3]. The conversion of K_F into S_{V_B} is straightforward for BJT, because $g_m/I_C \approx 1/\varphi_t \approx$constant – see again eq. (3) and the discussion between eqs. (28) and (30), because

$$\frac{S_{I_C}}{{I_C}^2} = \frac{K_F}{f} = \left(\frac{g_m}{I_C}\right)^2 S_{V_B} \approx \frac{S_{V_B}}{\varphi_t^2} \Rightarrow S_{V_B}\left(\text{at } f = 1\text{Hz}\right) \approx \frac{(26\text{mV})^2 \times K_F}{1\text{Hz}} \text{ at room temperature.} \quad (212)$$

We have used this conversion widely in this work.

To compare with the predictions in ITRS [3], we write here the values for $FOM_{S_{VB}}$ and $FOM_{S_{VG}}$ for the input referred 1/f voltage noise in BJT and MOS transistors, as evaluated from the trends in the experimental data shown in Figure 15.

The trend for 1/f noise in npn BJTs in Figure 15 is

$$FOM_{S_{VB}} = A_E \frac{f}{1\text{Hz}} S_{V_B} = 3.8\times10^{-12}\,\mu\text{m}^2\text{V}^2/\text{Hz} \text{ , with standard deviation } \sigma_{dB}=3.4\text{dB} \quad (213)$$

The prediction in ITRS is between 10^{-11}µm²V²/Hz to 10^{-12}µm²V²/Hz for the period from year 2001 to 2020 – see Figure 1a. The above $FOM_{S_{VB}}=3.8\times10^{-12}$µm²V²/Hz is within this interval, and thus, it is a representative value for 1/f noise in analog BJTs.

The trend for 1/f noise in MOS transistors in Figure 15 is

$$FOM_{S_{VG}} = WL\frac{f}{1Hz} S_{V_G} = 1.0\times10^{-9}\,\mu m^2V^2/Hz \text{ , with standard deviation } \sigma_{dB}=8.8\text{dB} \quad (214)$$

The prediction in ITRS is between 10^{-9}µm²V²/Hz to 1.8×10^{-10}µm²V²/Hz for the period 2001-2020 for analog MOS – see Figure 1a. The above $FOM_{S_{VG}}=1.0\times10^{-9}$µm²V²/Hz is slightly high. Nevertheless, this value is still representative for the 1/f noise in analog MOS transistors, considering the large spread in the experimental data, because the lower boundary $FOM_{S_{VG}}/10^{(\sigma_{dB}/10dB)}$ is 1.3×10^{-10}µm²V²/Hz, and it includes the prediction in ITRS. Moreover, when analyzing publications in the period 2007-2020 only, this value is reduced to 4.1×10^{-10}µm²V²/Hz, well inside the range predicted in ITRS. Note also that $FOM_{S_{VG}}$ is much larger than $FOM_{S_{VB}}$. The trends in Figure 15 imply $FOM_{S_{VG}}/FOM_{S_{VB}}\approx300$, which is in agreement with the prediction in ITRS [3] for 1/f noise in MOS transistors for analog applications – see again Figure 1a.

Flicker noise in gate leakage current

It was mentioned above that the noise from gate leakage has its own FOM, which is additively included in the output and input FOM of MOS transistors. The review [164] on gate tunneling current implies that the DC value of gate leakage current I_G usually follows Fowler–Nordheim tunneling with some enhancement at low field. Nevertheless, the power spectrum density S_{I_G} of the gate current noise is found to scale with the square of the I_G, as shown in the top-left plot of Figure 31, and S_{I_G} is reasonably well described by the simple SPICE equation for 1/f noise, given by

$$\frac{S_{I_G}}{{I_G}^2} = \frac{K_{FG}}{f} = \frac{FOM\,1}{f}. \quad (215)$$

Thus, the generic FOM 1 for S_{I_G} is the SPICE parameter K_{FG} itself. The data from [101, 110, 149] imply that $K_{FG}\sim10^{-9}$ for gate dielectrics based on silicon oxides, the nitridation of oxides virtually does not impact the noise and it can be three orders of magnitude higher when hafnium is added in order to increase the dielectric

permittivity. Note, FOM 1 does not use any rule for scaling with device size.

We can construct a performance FOM 2, which relates the noise to gate leakage current density $J_G=I_G/(WL)$. One may find FOM 2 handy for designs and quick comparison of different gate stacks, since the gate leakage is usually reported as a current density, while the scaling rule for S_{I_G} is not certain. Thus, we arrange the data for S_{I_G} vs. J_G as illustrated in bottom-left plot of Figure 31, and

$$\frac{S_{I_G}}{J_G{}^2} = \frac{(WL)^2 K_{FG}}{f} = \frac{FOM\ 2}{f}. \tag{216}$$

The observation by using of FOM 2=K_{FG}×Area² is that not only Hf based, but also any composite material for gate dielectric causes 3 decades higher noise in the leakage current, which is somehow in contradiction with the observation made by using FOM 1=K_{FG}.

Further, we can use FOM 3=Area×K_{FG}, which is the regular scaling rule for (reciprocal) areal dependence of noise, as follows from eqs. (2), (14) and (15). As mentioned above, the gate leakage is usually given in terms of current density, so we write

$$\frac{(WL)S_{I_G}}{I_G{}^2} = \frac{(WL)S_{I_G}\big/(WL)^2}{J_G{}^2} = \frac{(WL)K_{FG}}{f} = \frac{FOM\ 3}{f}, \tag{217}$$

and plot $S_{I_G}/(WL)$ vs. J_G as illustrated in bottom-right plot of Figure 31. In this plot we observe that FOM 3 spreads from low value for SiO_2 to high value for HfO_2, suggesting that the noise relatively increases when the dielectric constant increases. The quantity “noise density”, S_I/Area, vs. current density, J=I/Area, was used in [83] to present the noise in BJTs.

Thus, with these three FOM, which indeed are not unique choice, we demonstrated that there is a room for investigations and speculations for the noise in gate leakage, both observed in publications. The different FOM lead to different conclusions and the modeling of the noise in gate leakage current is not certain. Based on temperature dependence of S_{I_G} in the device with HfO_2, it is argued in [101] that the gate leakage current is due to trap-assisted tunneling or Poole–Frenkel conduction, whereas the conduction mechanism of the leakage is owing to direct tunneling in SiO_2 and HfSiON with low Hf content, showing much weaker temperature dependence. The 3 orders of magnitude higher flicker-noise for the case of HfO_2 with respect to the SiO_2 dielectric by means of FOM 1 and FOM 3 is found to be in contradiction with the smaller difference (~1.5 decades) for the noise in the drain current and it is argued in [101] that S_{I_G} is “sensitive to all the defects” in the oxide, whereas the drain current depends only on the fewer interfacial defects. However, the modeling in terms of flat band fluctuation for both S_{I_G} and drain current noise implies the opposite in [101]. In order to inspect whether the low-frequency noise in the gate and drain currents has the same origin, we have measured the coherence between S_{I_G} and S_{I_D} in [149], see eq. (27). The correlation between gate and drain noise currents is weak, the coherence is less than 30%, as shown in Figure 32, suggesting an independence between these noise currents. This is in agreement with the former observations in [101] that the gate current noise is a result from the fluctuation in the insulator leakage process, while the drain current noise reflects only the charge trapping in the gate insulators. Consequently, the later gate leakage current noise model based on flat-band fluctuation [101]

is not applicable, since the coherence between gate and drain noise currents is low, whereas it should be high, e.g. >60%, if the origin of these noise currents is the same. On the other hand, the model in [101] is able to explain biasing variations in the normalized noise S_{I_G}/I_G^2. Nevertheless, the good overlap in the bottom-left plot of Figure 31 suggests a handy relation between S_{I_G} and J_G, but this relation breaks the rule for areal dependence of noise and it is questionable in physical significance. Again, the flicker noise in gate leakage current is an interesting topic, the investigations are performed in a lack of mature theory, and there are speculations and unresolved issues for the scaling rules in this noise.

IV.4.3. Noise factor, noise resistance, noise temperature

These FOM represent the noise in a device in ratio to the thermal noise, which has frequency independent "white" spectrum density of the energy $S_{th}=4kT$, where kT is the product of Boltzmann constant k and absolute temperature T; and S_{th} is the fundamental lower limit for noise set by thermodynamic considerations in physical systems [172, 173]. In sensors, the reference value might not be the thermal noise, but a "unit value" of the sensed quantity – please see Sec. VII.3. **Noise in sensors**.

Provided that the thermal noise is also electrical, then the noise current i_n and voltage v_n are assumed fully correlated with the impedance Z_n (or its reciprocal, the admittance $Y_n=1/Z_n$), and

$$S_{th} = 4kT = i_n v_n = \frac{S_{Ith}}{Y_n} = \frac{S_{Vth}}{Z_n}, \tag{218}$$

where one of the spectra i_n or v_n is taken complex conjugated in order to obtain power spectrum densities $S_{Ith}=i_{th}i_{th}^*$ for noise current and $S_{Vth}=v_{th}v_{th}^*$ for noise voltage across the impedance $Z_n=1/Y_n$.

At RF range, which is currently in the focus of wide research [174], Z_n is a complex quantity and the imaginary part of $S_{Vth}=S_{th}Z_n$ is not measurable directly. Many publications [175, 176, 177, 178, 179, 180, 181, 182, 183 184, 185, 186, 187] imply that the RF noise originates from resistances of gate conductors and gate-channel resistance, providing also models and characterization methods, and the RF noise is affected by the impedance matching to Z_n and by the frequency. Looking chronologically, the earlier works [175, 176, 177, 178, 179, 180, 181, 182] address the RF noise from the perspective of the general theory of RF networks, replicating impedance mismatches in test setups with impedance tuners, while the later publications [183, 184, 185, 186, 187] relate the RF noise closely to device parameters and circuit applications, involving also simpler test setups to obtain the device noise parameters in cost-effective manner. However, we do not pursue a discussion on RF noise resistance further, since we deal with low-frequency noise in this work, where the imaginary part of Z_n is negligible and $Z_n \equiv R_n$. In this case, eq. (218) is rewritten as

$$S_{Vth} = 4kT_nR_n \Leftrightarrow S_{Ith} = 4kT_n/R_n, \tag{219}$$

where S_{Vth} (or S_{Ith}) is measured noise power spectrum density of voltage (or current), which might be not due to thermal noise, R_n is equivalent noise resistance at assumption $T_n=T$ being the actual temperature, or T_n is equivalent noise temperature at assumption $R_n=R$ being the actual resistance of the device or circuit impedance at a node of interest (e.g. output impedance of signal source). Consequently, R_n and T_n are FOM, since they might not correspond directly to physically existing electrical resistance or temperature in the device.

Example: Noise temperature of shot noise

One example that demonstrates difficulties of using the FOM noise temperature is given in Figure 33 for the shot

noise in gate leakage current, which is due to electrons overcoming the insulator barrier, rather than thermal motion of charge carriers. Consequently, the measured power spectrum density S_{I_G} of the white noise (filled squares in left-hand figure) obeys the relation for shot noise $S_{I,SH}$ (line through squares), given by

$$S_{I_G} = S_{I,SH} = 2qI_G\,, \tag{220}$$

and $S_{I_G} \neq S_{Ith}=4kT\cdot g_G$ does not follow the relation for thermal noise S_{Ith} (circles), where $g_G=\partial I_G/\partial V_G$ is the dynamic (AC) conductance of the gate insulator (open diamonds). From discussions, such as in [188], the noise temperature T_{SH} of the shot noise can be estimated from a chain of relations, given by

$$\begin{aligned}\frac{T_{SH}}{T} &= \frac{4kT_{SH}\cdot g_G}{4kT\cdot g_G} = \frac{S_{I,SH}}{S_{Ith}} = \frac{2qI_G}{4kT\cdot g_G} \\ &= \frac{I_G/g_G}{2kT/q} = \frac{I_G\dfrac{\partial V_G}{\partial I_G}}{2\varphi_t} = \frac{I_{DC}\times r_{AC}}{2\varphi_t} = \frac{1}{2}\frac{\text{"voltage drop"}}{\text{"thermal voltage"}} = \text{FOM}\end{aligned}\,. \tag{221}$$

Thus, the noise temperature of shot noise is a FOM, and it is 50% of the ratio of "voltage drop" of DC current on dynamic resistance r_{AC} to thermal voltage $\varphi_t=kT/q$. If the "voltage drop" is higher than $2\varphi_t$, then the noise temperature of shot noise is larger than the device temperature, as illustrated with circles in the right-hand plot of Figure 33. For pn-junctions biased in forward, the shot noise temperature is nearly 50% of the device temperature, since $r_{AC}\approx\varphi_t/I_{DC}$, and so, the shot noise is "colder" than the junction. In contrast, the shot noise in the collector current is "hot", because

$$r_{AC} = \frac{\partial V_{CE}}{\partial I_C} \approx \frac{V_{CE}+V_{EA}}{I_C} >> \frac{\varphi_t}{I_C}\,, \tag{222}$$

where V_{EA}~100V>10V is the Early voltage, then $V_{EA}/2\varphi t>400$ and $T_{SH}>400\times300K>10^4$ K and there is no physical meaning for these high noise temperatures in this case, although one can see even higher values for the so-called "hot temperature T_{hot}" of solid state noise sources with excess noise ratio $ENR=10dB\times\log10(T_{hot}/290K-1)>20dB$. Furthermore, the noise temperature also depends on the resistance of bias circuit [189], since $r_{AC}=R_B\|r_G$ is a parallel connection of device r_G and bias R_B resistances, as depicted above the right-hand plot in Figure 33. If $R_B>>r_G$, then the device determines the noise temperature, as shown in right figure for $R_B=\infty$. By reducing R_B from 20MΩ to 0.2MΩ, the circuit noise temperature due to shot noise in gate leakage current decreases, and the shot noise becomes "cold" once $I_G\times R_B<2\varphi_t$. Again, the noise temperature is a FOM, it might not correspond directly to physically existing electrical resistance or temperature in the device, although one may need to evaluate the noise temperature in particular applications, such as instrumentation for radiometry and cryogenic amplifiers [190, 191, 192], or to get insight on carrier energy in short channel MOS transistors and shot noise in the reverse biased junction from body to source in MOSFET, as reviewed in [174].

Noise temperature of carrier motion in the MOS transistor channel.

Nevertheless, when the dominant source is associated with carrier motion in the MOS transistor channel, then the FOM "noise temperature T_n" and "noise resistance R_n", respectively, are very useful, because the values for R_n match the model predictions both at medium frequency range 10-100MHz [105], as well as in RF frequency range above 1GHz. For the latter, a good correction for the channel noise temperature $T_{n,ch}$ [193] is

$$T_{n,ch} = \left(1 + \frac{g_m}{g_d}\right) T, \tag{223}$$

For cryogenic temperatures (T<10K) and millimeter waves (f>30GHz), the quantum fluctuations may also cause a difference between physical T and noise T_n temperatures, since the relation between them is [192]

$$\frac{T_n}{T} = \frac{\frac{hf}{kT}}{\exp\left(\frac{hf}{kT}\right) - 1} \le 1, \tag{224}$$

where $h=6.63\times10^{-34}$ Js is Planck constant and $k=1.38\times10^{-23}$ J/K is Boltzmann constant. It follows from this equation that the vacuum noise temperature is $T_{n,vac}\approx hf/(2kT)\sim3K$ at f=100GHz and T=5K, but at room temperature T=300K and f<1GHz, the relative difference between T_n and T is less than 0.01%, and it is negligible for the noise measurements of much larger inaccuracy. Worth mentioning, eq. (224) follows from the original derivation in [173], where the thermal voltage is given as

$$S_{V_{th}} = 4\frac{hf}{\exp\left(\frac{hf}{kT}\right) - 1} R \approx 4kTR \text{ , since } hf/kT<<1, \text{ and } \exp\left(\frac{hf}{kT}\right) - 1 \approx \frac{hf}{kT} \quad . \tag{225}$$

Further details for the quantum limit of noise temperature in amplifiers can be found in [192, 194, 195].

Noise Figure or Noise Factor

Among all other FOM that relate the noise in devices and circuits to the thermal noise, the Noise Figure or Noise Factor, abbreviated as NF, plays a crucial role both in device characterization and circuit design. This is because NF represents one transducer, e.g. amplifier or device that will be used in amplification circuit, how close to the ultimate floor of thermal noise is, since the generic definition of NF is the ratio of noise S_{in} "seen" at transducer input to the thermal noise S_{th} of the signal source. Thus, NF is

$$NF = \frac{S_{in}}{S_{th}} \text{ in linear units, or } NF_{dB} = 10dB \cdot \log_{10}(NF) \text{ in logarithmic units.} \tag{226}$$

Since S_{th} is always present at the transducer input, then $S_{in}\ge S_{th}$, and the ideal, noiseless transducer, has NF=1 or NF_{dB}=0dB. The real transducer "adds" noise with magnitude $S_{in\_ref}=S_{th}(NF-1)$ as referred at its input. So, devices and amplifiers with smaller NF are desired.

From this generic definition for NF, many other definitions are derived, in order to obtain expressions suitable for particular application in physical analyses, device characterization and measurements, or design procedures. Now we will present several of them.

Using the definition for spectral density for the power of the thermal noise, $S_{th}=4kT$ for the signal source and $S_{in\_ref}=4kT_{in\_ref}$ for the input referred noise power of the amplifier, then the NF can be rewritten in terms of noise temperatures in a straightforward manner as

$$NF = \frac{S_{in}}{S_{th}} = \frac{S_{in\_ref} + S_{th}}{S_{th}} = \frac{T_{in\_ref}}{T} + 1. \tag{227}$$

The problem for electrical circuits, however, is the determination of T_{in_ref}, since T_{in_ref} does not follow directly

from electrical parameters of the devices, as we have demonstrated above for shot noise. One can use eqs. (218) and (219) to write in terms of electrical quantities for impedance and power spectrum density of voltage, S_V, and gets

$$\mathrm{NF}=\frac{S_{in}}{S_{th}}=\frac{S_V+S_{Vth}}{S_{Vth}}=\frac{4kTZ_n}{4kTZ_s}+1\approx\frac{R_n}{R_s}+1, \tag{228}$$

where $R_s\approx Z_s$ is the resistance of the signal source, assumed with only thermal noise, but the transducer noise impedance $R_n\approx Z_n$ is usually different from its electrical input impedance, since the noise in the transducer is added from other noise sources and mechanisms not necessarily related to the thermal noise in the input electrical impedance.

Example. To illustrate this, we use the circuit in Figure 34 with nMOS transistor, which electrical resistance is large at the input terminal (>1V/100pA=10GΩ), while the noise resistance obtained from the last equation above is finite and less than 1MΩ, as estimated from the minimum NF later. We also address some issues with the fact that the Noise Figure (NF), the Input Referred Noise (IRN≡S_V below) and the Signal to Noise Ratio (SNR) are different measures for the noise, and they are conditionally related each to other.

In the circuit of Figure 34, the MOS transistor operates in ohmic mode, having also electrostatic discharge (ESD) protection diode, when acquiring from the signal source of resistance R_s via a "cable" with capacitance C_s=10pF. The 1/f noise of the nMOS transistor is modeled in terms of number fluctuation according to eq. (101) and the white noise is taken as the channel thermal noise with spectral density [169, 174]

$$S_{V_G}g_m^2=S_{I_D}=4kTg_d\text{, white noise of MOS transistor in linear regime,} \tag{229}$$

according to eq. (219), where $g_d=\partial I_D/\partial V_D\approx(W/L)\mu C_{ox}(V_G-V_T)$ is the channel conductance when (V_G-V_T)=1V>>V_D=0.1V, and no correction for excess noise due to short channel effects, body leakage, gate induced or avalanche noise is made [174], since the channel length is large (L=1μm) and the drain voltage and the frequency are low. The transistor and condition parameters relevant to the following calculations are given in Figure 34. Also, both 1/f and channel noise are referred as noise voltage S_{V_G} at gate terminal, and S_{V_G} is shown with checker-board patterned line in the left-hand plot of Figure 35.

The reverse (leakage) current I_{ESD} of the ESD protection diode produces shot noise current with power spectrum density $2qI_{ESD}$, which is converted into voltage noise at the input plane as $S_{V_{SH}}=2qI_{ESD}Z_s^2$, where the impedance Z_s of the circuit input node is obtained from $1/Z_s=1/R_s+j2\pi fC_s$. Since Z_s is a function of the frequency f, then $S_{V_{SH}}$ is frequency dependent, having a plateau $S_{V_{SH}}=2qI_{ESD}R_s^2$ at low frequencies that scales with the resistance R_s of the signal source, and $S_{V_{SH}}$ is attenuated at high frequencies by $[1+(2\pi fC_sR_s)^2]$, as shown with thin lines in the left-hand plot of Figure 35.

The thermal noise voltage $4kTR_s$ of the signal source is at the signal plane "inside" the signal source, but it appears attenuated at the input plane, by an attenuation factor of $[1+(2\pi fC_sR_s)^2]$ from signal plane to input plane for the power of the signals. Therefore, the thermal noise at the input plane is $S_{V_{TH}}=4kTR_s/[1+(2\pi fC_sR_s)^2]$, and $S_{V_{TH}}$ is shown with diagonal patterned lines in the left-hand plot of Figure 35 as the bottom boundaries of the shaded areas, because $S_{V_{TH}}$ is the reference against which the noise figure NF is determined. According to the generic definition in eq. (226), the noise figure is calculated from $\mathrm{NF}=S_V/S_{Vth}=1+(S_{V_G}+S_{V_{SH}})/S_{Vth}$, where the

total noise $S_V = (S_{V_G}+S_{V_{SH}}+S_{V_{th}})$ is the sum of all noise contributions to the input plane, as "seen" at the input of the amplifier (the gate terminal of the MOS transistor), and S_V is shown with thick black lines in the left-hand plot of Figure 35 as the upper boundaries of the shaded areas. The shaded areas in this figure is the NF, and the values for NF [converted in dB – see again eq. (226)] as function of frequency f and R_s are given in Figure 35 on the right-hand side by the contour plot, in which the dashed lines correspond to shaded areas in the left-hand plot of Figure 35.

Increasing the signal source resistance R_s, several observations can be made in the in the left-hand plot of Figure 35, in which the noise spectra are shown for three values of signal source resistance $R_s=\{10k\Omega,1M\Omega,100M\Omega\}$.

- First, at low frequency, f<10Hz, the 1/f noise of the MOS transistor dominates in the total input noise S_V, as along as $R_s<100M\Omega$, and S_V is virtually unchanged, although NF decreases (shaded areas become smaller at higher R_s), because the reference thermal noise voltage increases. Thus, the reduction in NF at low frequencies at high R_s does not imply reduction of noise, and since the noise level is constant, then the signal to noise ratio is also unchanged. At $R_s\sim200M\Omega$, the thermal noise dominates in the total noise S_V at 10Hz, and NF reaches local minimum, but S_V begins increasing with R_s. At higher R_s, the shot noise takes over at low frequencies, and both NF and S_V increase. For clarity in the figure, this situation is not shown.

- Second, at high frequency, f>100kHz, the total noise S_V changes a little with R_s, since the thermal noise and shot noise are attenuated, owing to the low impedance of the parasitic capacitance C_s at the input plane. Thus, S_V is constant or even decreases with R_s to the level of the white noise in S_{V_G} of the transistor, but NF increases, since the reference thermal noise is attenuated by approximately $(2\pi fR_sC_s)^2$. While a constant $S_V\sim S_{V_G}$ set by the noise of MOS transistor is similar to the first case at low frequency, the signal to noise ratio at high frequency decreases with R_s, since any signal from the signal source is attenuated in the same proportion $(2\pi fR_sC_s)^2$ as the thermal noise, which is different from the first case.

- Third, at medium frequencies of few kHz, the 1/f noise of MOS transistor is lower, and the thermal noise S_{Vth} from R_s can reach levels above 1/f noise before being attenuated by the square of capacitance conductance $(2\pi fC_s)^2$. In this case, the input noise $S_V\sim S_{Vth}$ follows closely the thermal noise, S_V increases with R_s, while NF reaches minimum values. The global minimum is known as Minimum Noise Figure NF_{min}, and it is in the locus of the contour plot in the right-hand side of Figure 35. NF_{min} is a conditional figure of merit, because it depends on biasing, frequency and impedance Z_s. The condition for Z_s suggests noise impedance matching, at which the noise from the transducer has the least relative contribution to the total noise, and NF_{min} estimates this contribution in comparison to the thermal noise. From the contour plot in the right-hand side of Figure 35, we estimate $NF_{min}<1.02=0.1dB$ at $R_s\approx2M\Omega$ and f=10 kHz. Any deviation from these conditions increases NF, as depicted in the contour plot in the right-hand side of Figure 35 with labeled arrows for the factors that cause increase in NF. The substitution in eq. (228) with the values for NF_{min} and the corresponding R_s implies that the noise resistance of the MOS transistor is $R_n<2M\Omega\times2\%\approx40k\Omega$, which is much lower than any electrical impedance of the gate terminal of the transistor at 10kHz, e.g. the impedance $1/(2\pi fC_g)\sim20M\Omega$ for the gate capacitance $C_g\sim0.8pF$ in the example.

Also, it should be noted again that the input referred noise level and SNR might be not at the optimum at the conditions for NF_{min}, because the thermal noise also varies with the impedance Z_s of circuit at its input plane, the low-frequency applications consider several frequency decades at low loading of signal source, thus do not care for power matching between source and load, and the SNR usually is not determined in respect to thermal noise,

since other types of noise dominate both in the signal source and in the amplifier. In RF applications, in contrast, the thermal noise is dominant in signal sources, the frequency range is “narrow”, e.g. few to 30 percent around the central frequency even in the so-called ultra wide band systems, the impedances are predetermined in a range between 30Ω and 600Ω, usually close to the characteristic impedances of 50Ω or 75Ω, and the impedance matching is very important in order to prevent from electromagnetic wave reflections, standing waves and to obtain power gain at high frequency. Therefore, NF is convenient and of high importance for RF applications.

Definition of NF for RF applications. To meet the objectives mostly in applications such as low noise RF preamplifiers, another definition for NF is derived from the generic definition of eq. (226). First of all, the impedance matching between signal source impedance Z_s and amplifier input impedance Z_i are taken into accout and the reference thermal noise power is not the full power 4kT, but the available power from the signal source at matched condition $Z_i=Z_s^*$, where Z_s^* is complex conjugated of Z_s; and the available thermal noise power from the signal source is [196]

$$S_{th,avbl}=\frac{S_{Vth}}{4R_s}=\frac{4kTR_s}{4R_s}=kT\text{ , available thermal noise power from RF signal source,} \tag{230}$$

where R_s is the real component of Z_s, and the noise voltage from the source is divided by 2, that is equally, between Z_s and Z_i at the matched condition $Z_i=Z_s^*$. This reduction of reference power in RF noise figures to ¼ of thermal noise is clearly stated in [197], but rarely mentioned in recent publications.

So, some RF noise figures are, as following. The input referred noise figure is

$$NF_{in}=\frac{S_{in,avbl}}{S_{th,avbl}}=1+\frac{S_e}{kT}=1+\frac{T_e}{T}, \tag{231}$$

where $S_{in,avbl}$ is the noise in the amplifier being referred as available power from signal source, $S_e=S_{in,avbl}-S_{th,avbl}$ is the excess noise added from the amplifier, also referred as available power from signal source, and $T_e=S_e/k$ is the equivalent excess noise temperature, corresponding to S_e. The standard reference temperature T_o is 290K, and if the physical temperature is different, then a correction with ratio T/T_o is made, as discussed in [197]. The excess noise ratio ENR, for example for noise sources, is given in respect to T_o, and $ENR=10dB\times\log10(S_e/kT_o)=10dB\times\log10(T_e/290K)$.

Since $S_{in,avbl}$ is input referred (actually, source referred), it does not exist as a physical signal that can be measured at the input of the amplifier. Therefore, one usually uses the output referred noise figure NF_{out}, which is obtained from the input referred by multiplying the nominator an denominator of eq. (231) with the so-called Transducer Power Gain G_T, and NF_{out} is

$$NF_{out}=\frac{G_TS_{in,avbl}}{G_TS_{th,avbl}}=\frac{S_{out}}{G_TS_{th,avbl}}=\frac{S_{out}}{G_TkT}, \tag{232}$$

where S_{out} is the measured noise power at the output of the amplifier, and $G_TS_{th,avbl}$ is the reference value for the output noise power that corresponds only to the thermal noise from signal source. The transducer power gain G_T is the ratio of the power delivered to load with impedance Z_L in the output of the amplifier to the power available from the signal source with impedance Z_s at the input of the amplifier, and G_T can be obtained from S-parameter measurements of the source, amplifier and load, according to

$$G_T = \frac{|S_{21}|^2\left(1-|\Gamma_s|^2\right)\left(1-|\Gamma_L|^2\right)}{\left|(1-S_{11}\Gamma_s)(1-S_{22}\Gamma_L)-S_{12}S_{21}\Gamma_s\Gamma_L\right|^2}, \tag{233}$$

where S_{11}, S_{12}, S_{21}, S_{22} are the S-parameters of the amplifier, $\Gamma_s=(Z_s-Z_o^*)/(Z_s+Z_o)$="$S_{22s}$" is the reflection coefficient of the signal source ("S_{22}" of the source), $\Gamma_L=(Z_L-Z_o^*)/(Z_L+Z_o)$="$S_{11L}$" is the reflection coefficient of the load at the output of the amplifier, all these measured with a network analyzer with characteristic impedance $Z_o=Z_o^*=50\Omega$, for example. It is important to note that there are several noise figures for RF. For example, the publications usually report minimum noise figure [175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187], which is bias-frequency dependent and for optimum impedance of the RF network for the lowest noise figure. As mentioned above for the noise resistance, the earlier works [175, 176, 177, 178, 179, 180, 181, 182] address the RF noise from the perspective of the general theory of RF networks, replicating impedance mismatches in test setups with impedance tuners, while the later publications [183, 184, 185, 186, 187] relate the RF noise closely to device parameters and circuit applications, involving also simpler test setups to obtain the device noise parameters in cost-effective manner. The driving force for simplification of the RF noise analyses is normally that the RF noise is just one of many other performances that one needs to investigate, e.g., during reliability analyses by hot-carrier stress of RF amplifiers [198, 199].

We should also note that the optimum impedance for the minimum noise figure is neither the impedance Z_o of the network nor the impedance for maximum power gain. Thus, the noise figure of an amplifier or mixer in a given RF application might be a lot larger (in the range 4-20 dB) compared to the minimum noise figure (in the range from fraction to few dB) of the used transistors.

Signal to noise ratio. The most popular definition for noise figure is in terms of decrease of signal to noise ratio (SNR) in the output of the amplifier, SNR_{out}, as compared to SNR_{in} at amplifier input, or precisely, SNR_s in the output of signal source. This form of NF_{SNR} was historically first introduced [200] due to its clear meaning for the practice, when the noise in the signal source is thermal, e.g. antenna of receiver. NF_{SNR} can be obtained from the generic definition for NF by assuming a signal with an arbitrary power level P_s provided from the signal source and amplified to output power $P_L=G_TP_s$ on the load of the amplifier. Since the gain $G_T=P_L/P_s$ is the same also for the noise, then NF_{SNR} is

$$NF_{SNR} = \frac{S_{in,avbl}}{S_{th,avbl}} = \frac{P_s/S_{th,avbl}}{(G_TP_s)/(G_TS_{in,avbl})} = \frac{SNR_s}{P_L/S_{out}} = \frac{SNR_{in}}{SNR_{out}}. \tag{234}$$

Evidently, NF_{SNR} does not explicitly state what is the reference noise power at the input, and one may carry out a procedure for evaluation of NF_{SNR} by measuring SNR first at the source plane and then at the output of amplifier. Since NF_{SNR} is formally derived from the generic definition of NF, then one may decide that NF_{SNR}=?=NF and to substitute in the generic definition for NF in order to evaluate the input referred noise S_V, getting

$$\frac{SNR_{in}}{SNR_{out}} = NF_{SNR} = ? = NF = 1+\frac{S_V}{S_{Vth}} = \begin{cases} 1+\dfrac{S_{V,LF}}{4kTR_s}, \text{for LF noise, or} \\ 1+\dfrac{S_{V,RF}}{kTR_s}, \text{for RF noise,} \end{cases}, \tag{235}$$

from which the power spectrum density of the input referred noise voltage S_V of the amplifier will be estimated

as

$$S_V = \frac{S_{V,LF}}{4} \text{ or } S_{V,RF} = kTR_s(1-NF) = ? = kTR_s(1-NF_{SNR}) = kTR_s\left(1-\frac{SNR_{in}}{SNR_{out}}\right) = S_{V,SNR}\,. \quad (236)$$

We put the question in the equations, because the equality is valid only if SNR_{in} is measured in respect to the thermal noise, which is not possible in the practice directly, because the thermal noise is the ultimate noise floor, and both the signal source and the measurement instrument for noise power are expected to have higher noise levels. Consequently, S_V determined from NF is the correct, whereas $S_{V,SNR}$ calculated from NF_{SNR} has a value different from S_V. This is because, taking reference noise level $S_{ref} \neq kTR_s$, we have

$$S_{V,SNR} = kTR_s\left(1-\frac{SNR_{in}}{SNR_{out}}\right) = kTR_s\left[1-\frac{\frac{P_s}{S_{ref}}}{\frac{G_T P_s}{G_T(S_{ref}+S_V)}}\right] = \frac{kTR_s}{S_{ref}}S_V \neq S_V\,, \quad (237)$$

and, since the signal source used to supply with the signal P_s has most likely noise $S_{ref} > S_{Vth} = kTR_s$, then the use of NF_{SNR} will underestimate the actual value for the input referred noise of the amplifier under test with the ratio S_{Vth}/S_{ref}.

FOMs for non-thermal sources of noise

In summary, the above discussion on noise temperature, noise resistance and NF implies that these are conditional figures of merit that relate the noise to the thermal noise, and thus, they are applicable when the origin of the noise in the signal source is thermal. There is a wide range of applications in the medium frequency range, e.g. material characterization, ultrasound imaging and electronic identification tags (RFID), where these figures of merit are essentially useful. The noise figure is particularly important for RF applications and currently intensive research is undergoing to resolve many issues with matching and characterization uncertainty. Furthermore, the thermal noise is dominant in the channel noise of MOS transistors, and therefore, the figures of merit that use the thermal noise as the reference are important. However, when the origin of the noise is not thermal, the above figures of merit are not very suitable. These are cases for BJT, reverse biased and avalanche diodes and non-resistive leakages (due to tunneling in insulators, especially in SOI transistors), for example, in which the shot noise is the major concern. As for the low-frequency range, the dominant noise is not originating from thermal noise, both in signal sources and in the devices, and the input referred noise is usually given with a pair S_{Veq} and S_{Ieq} of equivalent voltage and current noise sources, respectively. For MOS transistors, the input referred voltage noise is dominant, and it is given by $S_{Veq} = S_{V_G}$ in most cases, as discussed earlier in section IV.1. "**Models and predictability**", while for BJT the input referred current noise is dominant, and it is given by $S_{Ieq} = S_{I_B}$, as discussed in details in section III.3. "**Crossover between different noise sources in BJT**". Most of the input referred noise voltages are physically non-existing quantities that represent equivalently noise sources inside the devices, while the most of input referred noise currents are actually fluctuations in biasing or leakage currents and are physically existing quantities in the input terminal of the devices and circuits. Interestingly, from the ratio S_{Veq}/S_{Ieq} we can derive other figures of merit, R_{eq} and T_{eq}, given by

$$\frac{S_{Veq}}{S_{Ieq}} = R_{eq}^2 \Rightarrow \frac{S_{Veq}}{R_{eq}} = S_{Ieq}R_{eq} = 2kT_{eq}\,. \quad (238)$$

The significance of R_{eq} is that it determines whether the voltage or current input noise will dominate at particular resistance R_s of signal source or circuit impedance. If $R_s<R_{eq}$, then the voltage noise is dominant, whereas if $R_s>R_{eq}$, then the current noise is dominant. At $R_s=R_{eq}$, both sources have equal contribution to the input referred noise, which can be referred to equivalent noise temperature T_{eq} and equivalent power $4kT_{eq}$ of the input referred noise. By expressing in terms of voltage noise and using eq. (228), for example, we can relate to noise figure, as

$$\begin{aligned} NF &= 1+\frac{S_V}{S_{Vth}} = 1+\frac{S_{Veq}+R_s^2 S_{Ieq}}{4kTR_s} \\ &= 1+\frac{2kT_{eq}R_{eq}+R_s^2\,2kT_{eq}/R_{eq}}{4kTR_s} \\ &= 1+\frac{T_{eq}}{T}0.5\left(\frac{R_{eq}}{R_s}+\frac{R_s}{R_{eq}}\right)\geq 1+\frac{T_{eq}}{T}\geq 1+\frac{\min\left(T_{eq}\right)}{T}=NF_{min} \end{aligned} \tag{239}$$

Taking into account that S_{Veq} and S_{Ieq} vary with the frequency, for example 1/f noise at low frequencies, the significance of last line of expressions is the following. The left-hand expression shows that NF has a local minimum when $R_s=R_{eq}(f)$, and the local minimum $[1+T_{eq}(f)/T]$ is given with the second expression. For illustration, consider again the contour plot in the right-hand side of Figure 35 for a frequency about 200Hz. At low R_s=1kΩ, the 1/f noise voltage S_{V_G} is dominant, but increasing R_s to 1GΩ, the shot noise current $2qI_{ESD}$ takes over. At $R_s=R_{eq}$~30MΩ, the 1/f noise and the shot noise have equal contributions, NF reaches local minimum of about 0.5dB=1.12, and since the curves are for room temperature, then T~300K and T_{eq}(f=200Hz)~36K. Increasing the frequency, the local minimum decreases, and for f~10kHz and R_s≈2MΩ, the global minimum for $NF=NF_{min}$<0.1dB=1.02 is reached, which corresponds to $\min(T_{eq})$<7K. This is given with the left-hand expressions in eq. (239). In this way, we demonstrate that the ratio between voltage and current noise is related to the noise resistance R_{eq} and noise temperature T_{eq} by eq. (238), and at condition $R_s=R_{eq}$, the contribution of voltage and current noise are equal, the noise figure has a minimum with value of $(1+T_{eq}/T)$, as given with eq. (239).

Recalling that the original derivation for RF noise figure has extracted NF_{min} from input referred voltage and current noise sources [201], then NF_{min} should have similar meaning in the case of frequency dependent impedances, as the discussed above for low frequencies. We did not find recent work that provides deep insight on the physical significance of the four RF noise parameters NF_{min}, r_n and magnitude and angle of Γ_{opt}, that participate in the popular equation for RF noise figure

$$NF_{RF} = NF_{min}\left(\Gamma_{opt}(f)\right)+4r_n\frac{\left|\Gamma_s-\Gamma_{opt}\right|^2}{\left(1-\left|\Gamma_s\right|^2\right)\left|1+\Gamma_{opt}\right|^2} = NF_{min}\left(Y_{opt}(f)\right)+\frac{R_n}{G_s}\left|Y_s-Y_{opt}\right|^2, \tag{240}$$

where at given frequency f, the RF noise figure NF_{RF} has a minimum in respect to signal source matching Γ_{opt}, or admittance Y_{opt}, and when the reflection coefficient Γ_s or admittance $Y_s=G_s+jB_s$ of the signal source deviates from Γ_{opt} or Y_{opt}, then NF_{RF} increases with a "rate" given by the parameter $r_n=R_n/Z_o$, or R_n/G_s, respectively. The relation between power $Y_s=Y_{in}$* and noise matching $Y_s=Y_{opt}$ is still not well elaborated for implementation in design procedures.

Another issue for RF noise figure is that the reference noise is solely attributed to the real component $R_s=1/G_s$ of the impedance of the signal source, and all reactances are assumed noiseless. This is somewhat in contradiction with the general expression for thermal noise in eq. (218), which does not discriminate complex impedances and suggests looking at the input signal plane not discriminating between source and load at it. Thus, along with the many technical difficulties in measuring RF noise, some more general research is expected in near future, since the RF applications reached maturity in millimeter wavelengths, and the reactances and distributed loss dominate in these circuits, while the equation for noise uses lumped parameters. Again, the thermal noise is not dominant in the low-frequency noise, and all figures of merit related to thermal noise are lacking of physical significance from the perspective of non-linear device physics at low frequency. Therefore, other figures of merit are usually used for low-frequency noise, and these are discussed next for MOS transistors.

IV.4.4. Physical figures – trap density, Hooge parameter, scattering parameter

For MOS transistors, three physical parameters are usually used as figures of merit for the 1/f noise. These are oxide trap density N_t for number fluctuation model, scattering parameter for models with correlated mobility fluctuation and Hooge parameter α_H for uncorrelated mobility fluctuation and noise from contact resistance. The physical significance and the issues that have arisen with these parameters were discussed in section IV.1. "**Models and predictability**". Here we mention that the most popular procedure to discriminate number and uncorrelated mobility fluctuation is that in [156], which analyzes the behavior of normalized noise S_{I_D}/I_D^2 against the behavior of $(g_m/I_D)^2$ versus bias. Consider eq. (88). If $S_{I_D}/I_D^2 \propto (g_m/I_D)^2$, then the 1/f noise in MOS transistor can be referred as gate voltage noise S_{V_G} with origin number fluctuation according to eq. (104), whereas, if $S_{I_D}/I_D^2 \propto (g_m/I_D)$, then the uncorrelated mobility fluctuation is in the origin of 1/f noise, since in strong inversion $(g_m/I_D) \propto 1/(V_G - V_T) \propto 1/n_{eff}$ is inversely proportional to number of carriers n_{eff} in MOS channel, while both in strong and in weak inversion regimes $I_D \propto n_{eff}$, and so, $S_{I_D}/I_D^2 \propto \alpha_H/n_{eff} \propto 1/I_D$.

Obviously, the noise is referred to the gate terminal of MOS transistors, and when including the correlated mobility fluctuation, the gate referred 1/f noise voltage to the first order of approximation is

$$S_{V_G} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT\lambda N_t}{WL}\left[1+\theta\left(V_G - V_T\right)\right]^2 + \frac{1}{f}\left(\frac{q}{C_{ox}}\right)\frac{\alpha_H}{WL}\left(V_G - V_T\right), \tag{241}$$

as follows from eqs. (88), (90), (104), (120) and (121) for strong inversion regime of operation of MOS transistor above threshold voltage V_T in linear mode, for example. The tunneling attenuation distance λ and gate capacitance per unit area C_{ox} depend on the gate dielectric, and the mobility degradation coefficient θ depends on the electric field, but to the first order of approximation they can be taken constants for the purpose of comparison in terms of figure of merit. Taking the values for silicon MOS transistor with SiO_2 gate insulator of thickness t_{ox}=EOT and with permittivity $\varepsilon_e=\varepsilon_{SiO_2}=8.85\times10^{-14}$F/cm tunneling attenuation distance λ_e=0.1nm, one can rewrite the last equation as

$$S_{V_G} = \frac{1}{f}\left\{\frac{qEOT}{\varepsilon_e}\right\}^2\left\{\frac{kT\lambda_e}{WL}\right\}N_t\left[1+\theta\left(V_G - V_T\right)\right]^2 + \frac{1}{f}\left\{\frac{qEOT}{\varepsilon_e WL}\right\}\alpha_H\left(V_G - V_T\right), \tag{242}$$

where the quantities in the large brackets are taking care for device scaling, and N_t, θ and α_H can be used as figures of merit, which have physical meaning of equivalent trap density, scattering parameter and Hooge

parameter, respectively, for number fluctuation, correlated mobility fluctuation and uncorrelated mobility. Note that, if the scattering parameter θ is constant, then the correlated mobility results in quadratic dependence on gate overdrive (V_G-V_T), while the uncorrelated mobility suggests a linear dependence, if α_H is bias independent. This difference in the bias dependences is suggested in [156] as the criterion to discriminate correlated and uncorrelated mobility fluctuations. Many publications suggest that the Coulomb scattering causes bias variation of correlated mobility fluctuation, as discussed earlier by the help of eqs. (116) to (138).

The figure of merit for the equivalent oxide density N_t is shown in Figure 36. For nMOS transistors, the data are collected from [22, 47, 48, 49, 50, 51, 72, 82, 88, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 149, 145, 202, 203, 204, 205, 206] for silicon transistors and from [109, 155, 207] for transistors that use germanium in the structure, e.g. to provide strain in the lattice. For pMOS transistors, the data are collected from [47, 52, 88, 94, 95, 96, 97, 100, 102, 104, 105, 109, 124, 125, 126, 128] for silicon transistors and from [52, 57, 79, 109, 208] for SiGe and SiGeC pMOS transistors. The data are stored in numerical form in [209]. While in the past, the uncorrelated mobility fluctuation was found to dominate in pMOS transistors, later publications imply the opposite, when EOT<6nm. The second observation in Figure 36 is that there is a crossover for N_t at EOT<4–5nm. In "thick" oxide transistors with EOT>4nm, the equivalent oxide trap density is low with an average of $N_t \sim 10^{17} cm^{-3} eV^{-1}$, as shown by the right-hand histogram in Figure 36. In "thin" oxide transistors, however, N_t (and thus, the 1/f noise) increases inversely with EOT with a steep slope of m=2.5 in magnitude. The scattering in the data is large and the slope was estimated by adjusting the symmetry in the left-hand histogram for the quantity $N_t EOT^m$. Since the slope m>2, then the trend at low EOT suggests that the $1/C_{ox}^2$ scaling rule for 1/f noise is not followed anymore for transistors with high-k gate dielectrics of high permittivity.

Many reasons for higher N_t are suggested in the above references, e.g. larger tunneling attenuation distance λ in high-k dielectrics, non-uniformity of dielectric structure, and other. One of the other is the increased doping in the channel of MOS transistor, which is necessary in order to compensate for drain induced barrier lowering. The higher doping results in intensive Coulomb scattering at low bias, which effectively increases gate referred 1/f noise voltage at $(V_G-V_T)\sim 0.1V$, as discussed in section IV.3. "**RTS noise in MOS transistors**". On the other hand, this biasing condition is usually assumed for extraction of the value for N_t, since the term $\theta(V_G-V_T)<<1$ is assumed in eq. (242). However, as follows from eq. (138), the relative contribution of correlated mobility fluctuation due to Coulomb scattering increases when the bias is reduced, and the figure-of-merit form in eq. (242) may overestimate the oxide trap density. Interestingly, the slope of increase of N_t at low EOT in Figure 36 is with value 0.5 higher than 2, which one can expect from the term for Coulomb scattering in eq. (138).

To avoid the Coulomb scattering effects, we have used in Figure 37 only data for high overdrive voltage $(V_G-V_T)>0.3V$. At this condition, one expects that the effective parameter θ for correlated mobility fluctuation reflects phonon and surface scattering, according to eqs. (136) and (137); and θ is given by

$$\theta = \left\{ \begin{array}{l} \alpha_s \mu \dfrac{C_{ox}}{q}, \text{ with } \alpha_s \sim 10^{-15} Vs \\ \alpha_s \mu C_{ox}, \text{ with } \alpha_s \sim 10^{4} Vs/C \end{array} \right\} \propto \mu C_{ox} \propto \frac{1}{EOT \times N_t} \qquad (243)$$

depending on whether the electron charge q is included in the definition for the scattering parameter α_s, that is, whether the number fluctuation model uses concentration n' of channel carriers or their charge concentration

q×n'. In eq. (243), the gate oxide capacitance per unit area C_{ox} suggests that $\theta \propto 1/EOT$, and according to eq. (136), the carrier mobility μ suggests $\theta \propto 1/N_t$. None of these dependences can be observed in Figure 37 from the published data in [47, 48, 49, 50, 51, 95, 100, 106, 149] for nMOS transistors, in [155] for strained on SiGe layer and control nMOS transistors, in [47, 52, 88, 94, 95, 100, 105, 124, 126, 128] for pMOS transistors, and in [52, 57, 79, 208] for SiGe and SiGeC pMOS transistors. The data are stored in numerical form in [209].

The above discussion clearly indicates that there are problems in the characterization of 1/f noise in terms of physical figures of merit. One partial solution to the problems could be, if the expression for correlated mobility fluctuation is split into two terms, one for Coulomb scattering at low bias and another for phonon and roughness scattering at high bias. The resulting equation for the gate referred 1/f noise voltage is

$$S_{V_G} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT\lambda_e N_t}{WL}\left[1+\theta_C\sqrt{V_G - V_T}+\theta(V_G - V_T)\right]^2 + \frac{1}{f}\left(\frac{q}{C_{ox}}\right)\frac{\alpha_H}{WL}(V_G - V_T), \tag{244}$$

but this equation has a problem when the parameter θ_C associated with the Coulomb scattering dominates, since the equation reduces to

$$S_{V_G} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT\lambda_e N_t}{WL}\theta_C^2(V_G - V_T) + \frac{1}{f}\left(\frac{q}{C_{ox}}\right)\frac{\alpha_H}{WL}(V_G - V_T), \tag{245}$$

and the correlated and uncorrelated mobility fluctuations cannot be discriminated experimentally, because they have the same bias dependence. Nevertheless, both fluctuations in this case are mobility fluctuations and perhaps it is not so critical to know their contributions separately. If this is acceptable, then one can define "apparent" Hooge parameter α_{HC} from the parameter θ_C for correlated mobility fluctuation caused by Coulomb scattering by the equation

$$\alpha_{HC} = \frac{qkT\lambda_e N_t}{C_{ox}}\theta_C^2 = kT\lambda_e N_t\left(\frac{\mu}{\mu_{C0}}\right)^2, \text{ with } \theta_C = \frac{\mu}{\mu_{C0}}\sqrt{\frac{C_{ox}}{q}} \approx \frac{\mu_o}{\mu_{C0}}\sqrt{\frac{\varepsilon_{SiO_2}}{qEOT}} \tag{246}$$

for strong inversion regime, where the mobility $\mu_o \approx \mu$ can be regarded as the maximum value for carrier mobility at low electric fields [57]. By taking typical values $\mu \approx 100 cm^2/Vs$, $\mu_{C0} \approx 10^8 cm/Vs$ and $C_{ox} \approx 10^{-6} F/cm^2$ for transistors with EOT~3.5nm, then $\theta_C \approx 2.5\ V^{-0.5}$, and 1/f noise due to Coulomb scattering dominates at $(V_G - V_T) > 0.16V$, since $\theta_C\sqrt{V_G - V_T} > 1$. Therefore, eqs. (245) and (246) are applicable for $(V_G - V_T) \sim 0.2V$, and since kT=0.026eV and $\lambda_e \approx 0.1nm$, then from eq. (246) we get $\alpha_{HC} \approx 2.6\times10^{-4}$ for $N_t \sim 10^{18} cm^{-3}eV^{-1}$, the latter taken from the trend in Figure 36 for EOT~3.5nm. Interestingly, while "apparent", α_{HC} is in the range of meaningful values for Hooge parameter, which shows one more time that there is a convergence between different models for 1/f noise.

IV.4.5. <u>Performance figures – RF to LFN</u>

Denormalization rules for $FOM_{S_{VG}}$

The mostly accepted performance figure of merit for 1/f noise in MOS transistor is $FOM_{S_{VG}}$, it is related to the gate referred voltage noise S_{V_G}, as given by eq. (106), and it originates from number fluctuation model. $FOM_{S_{VG}}$ is the power spectrum density of S_{VG} at 1Hz in a transistor with gate area of $1\mu m^2$ and at low gate overdrive

voltage (V_{GS}-V_T)≈0.1V. $FOM_{S_{VG}}$ takes into account the general scaling rule for reciprocal dependence between noise level and device active area, given by eq. (14), and therefore allows for examination other factors and scaling rules that impact the 1/f noise and for comparisons between devices, technologies and so on, as has been illustrated in Figure 20 and Figure 22. The details related to $FOM_{S_{VG}}$ are discussed in section IV.1. "**Models and predictability**". Here, we briefly give the denormalization rules for $FOM_{S_{VG}}$ by the help of the following equation

$$S_{V_G} = \frac{FOM_{S_{VG}}}{f}\frac{1\mu m^2}{WL} \text{ and } S_{I_D} = S_{V_G} g_m^2 \text{, at specified bias condition for } FOM_{S_{VG}}. \quad (247)$$

Note that the bias dependence of the noise is neglected in $FOM_{S_{VG}}$ at an assumption that the bias dependence of the noise in relative units is similar in different MOS transistors, and also, it is assumed that the only one size dependence is the scaling rule with the reciprocal of the area of otherwise identical devices and the noise is exactly 1/f. Evidently from the discussions in preceding sections, the bias, frequency and size dependences vary several decades, and the models are devoted to capture these dependences. There are many issues when denormalizing $FOM_{S_{VG}}$ with bias and when the noise is not 1/f. Therefore, $FOM_{S_{VG}}$ is just a helpful figure of merit with significance for general comparisons, while, in practice, the actual realization of the low-frequency noise in particular MOS transistor and circuit made of it will be different from $FOM_{S_{VG}}$.

Corner frequency f_c between flicker (1/f) noise and white noise (FOM_{fc})

Other figure of merit is the corner frequency f_c between flicker (1/f) noise and white noise, which we denote here as $FOM_{fc}\equiv f_c$. FOM_{fc} is used often in the practice in the fields of electronic oscillators, frequency converters and sampling systems, because of two reasons. First, the design equations are given in terms of f_c, and second, the white noise in MOS transistors follows very closely the theoretical derivations made on assumption for channel thermal noise. The power spectrum density of the channel thermal noise is given by [169, 174]

$$S_{V_G,wh}\, g_m^2 = S_{I_D,wh} = 4kT\left[\frac{W}{L}\mu C_{ox}\left(V_G - V_T\right)\right], \text{ MOS transistor white noise in linear regime,} \quad (248)$$

$$= 4kT\left[\frac{2}{3}\frac{W}{L}\mu C_{ox}\left(V_G - V_T\right)\right], \text{ MOS transistor white noise in saturation regime,} \quad (249)$$

$$= 2qI_D, \text{ MOS transistor white noise in sub-threshold regime,} \quad (250)$$

where $S_{I_D,wh}$ is the white noise in drain current, $S_{V_G,wh}$ is the gate referred voltage white noise, which is "coupled" back for convenience via the transconductance g_m, according to eq. (1). These equations are for long-channel approximation, and one usually applies multiplicative correction factor γ for excess noise due to short channel effects, body leakage, gate induced or avalanche noise [174], but we omit it for clarity in analyses of low-frequency noise, since the variations in flicker noise are much larger than γ, while γ plays essential role in RF range.

We rewrite eqs. (248), (249) and (250) in terms of gate referred voltage white noise $S_{V_G,wh} = S_{I_D,wh}/g_m^2$, because we have the expressions for the gate referred flicker noise S_{V_G} in previous sections, e.g., eq. (241) above. The resulting equations for $S_{V_G,wh}$ are

: in linear mode $g_m = \frac{W}{L}\mu C_{ox} V_D$. So

$$S_{V_G,wh} = \frac{S_{I_D,wh}}{g_m^2} = \frac{4kT}{\mu C_{ox}} \frac{L}{W} \frac{(V_G - V_T)}{V_D^2}, \text{ MOS transistor white noise in linear regime,} \quad (251)$$

in saturation mode $g_m = \frac{W}{L}\mu C_{ox}(V_G - V_T)$. So

$$S_{V_G,wh} = \frac{2}{3} \frac{4kT}{\mu C_{ox}} \frac{L}{W} \frac{1}{(V_G - V_T)}, \text{ MOS transistor white noise in saturation regime,} \quad (252)$$

in sub-threshold mode $g_m = \frac{I_D}{kT/q} = \frac{I_D}{\varphi_t}$. So

$$S_{V_G,wh} = \frac{2q}{I_D}\left(\frac{kT}{q}\right)^2 = \frac{2q}{I_D}\varphi_t^2 \approx \frac{1}{4} \frac{4kT}{\mu C_{ox}} \frac{L}{W} \frac{1}{\varphi_t \exp\left(\frac{V_G - V_T}{\varphi_t}\right)}, \text{ white noise in sub-threshold regime,} \quad (253)$$

when using $I_D \approx \frac{W}{L}\mu C_{ox}(2\alpha)\varphi_t^2 \exp\left(\frac{V_G - V_T}{\alpha\varphi_t}\right)$ for the MOS transistor current in sub-threshold regime [p.177 in 169] with $\alpha \approx 1$ by neglecting depletion capacitance – see eq. (186).

The corner frequency $FOM_{fc} \equiv f_c$ between flicker noise S_{V_G} and white noise $S_{V_G,wh}$ is defined obviously as the frequency at which $S_{V_G} = S_{V_G,wh}$, because at $f < f_c$ the flicker noise dominates S_{V_G} in the power spectrum density of the noise, while $S_{V_G,wh}$ dominates at $f > f_c$. To obtain expressions for $FOM_{fc} \equiv f_c$, we use the last three eqs. (251), (252) and (253) for the white noise $S_{V_G,wh}$ in different regimes of operation of MOS transistor, and the components for the 1/f noise S_{V_G} for different noise mechanisms in eq. (241).

- In sub-threshold regime, we use the number fluctuation model Δn for the flicker noise S_{V_G}, and FOM_{fc} is found using eq. (253) for the white noise $S_{V_G,wh}$, as

$$S_{V_G} = \frac{1}{f_c}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT\lambda N_t}{WL} = \frac{1}{4} \frac{4kT}{\mu C_{ox}} \frac{L}{W} \frac{1}{\varphi_t \exp\left(\frac{V_G - V_T}{\varphi_t}\right)} = S_{V_G,wh}, \ \Delta n \text{ in sub-threshold.} \quad (254)$$

$$FOM_{f_c}\Big|_{V_G < V_T}^{\Delta n} = f_c = \frac{q^2 \lambda N_t}{C_{ox}} \mu \frac{\varphi_t \exp\left(\frac{V_G - V_T}{\varphi_t}\right)}{L^2} \propto \exp\left(\frac{V_G}{\varphi_t}\right) \propto I_D. \quad (255)$$

- When using eq. (252) for the white noise $S_{V_G,wh}$ in saturation regime above threshold, the number fluctuation Δn for the flicker noise S_{V_G} results in FOM_{fc}, given by

$$FOM_{f_c}\Big|_{sat}^{\Delta n} = f_c = \frac{3}{8}\frac{q^2\lambda N_t}{C_{ox}}\mu\frac{(V_G - V_T)}{L^2} \propto (V_G - V_T) \propto \sqrt{I_D}\ . \tag{256}$$

At high gate bias, the correlated mobility fluctuation $\Delta\mu>\Delta n$ dominates when $\theta(V_G-V_T)>1$, then $[1+\theta(V_G-V_T)]^2\approx\theta^2(V_G-V_T)^2$, and FOM_{fc} is given by

$$FOM_{f_c}\Big|_{sat}^{\Delta\mu > \Delta n} = f_c = \frac{3}{8}\frac{q^2\lambda N_t}{C_{ox}}\mu\theta^2\frac{(V_G - V_T)^3}{L^2} \propto (V_G - V_T)^3 \propto I_D\sqrt{I_D}\ , \text{ if } \theta\approx\text{constant.} \tag{257}$$

Consequently, using eq. (251) for the white noise $S_{V_G,wh}$ in linear (ohmic) regime above threshold, $V_D<<(V_G-V_T)$, the number fluctuation Δn for the flicker noise S_{V_G} results in FOM_{fc}, given by

$$FOM_{f_c}\Big|_{lin}^{\Delta n} = f_c = \frac{1}{4}\frac{q^2\lambda N_t}{C_{ox}}\mu\frac{V_D^2}{L^2(V_G - V_T)} \propto \begin{cases} V_D^2 \propto I_D^2 \\ 1/(V_G - V_T) \propto 1/I_D \end{cases}, \tag{258}$$

the correlated mobility fluctuation $\Delta\mu>\Delta n$ at high gate bias, when $\theta(V_G-V_T)>1$, results in

$$FOM_{f_c}\Big|_{lin}^{\Delta\mu > \Delta n} = f_c = \frac{1}{4}\frac{q^2\lambda N_t}{C_{ox}}\theta^2\mu\frac{(V_G - V_T)V_D^2}{L^2} \propto \begin{cases} V_D^2 \propto I_D^2 \\ (V_G - V_T) \propto I_D \end{cases}, \text{ if } \theta\approx\text{constant,} \tag{259}$$

and the uncorrelated mobility fluctuation (represented by Hooge parameter α_H), or the Coulomb scattering noise represented with its "apparent" Hooge parameter α_{HC} at $\theta_C\sqrt{V_G - V_T} > 1$, see eq. (246), results in

$$FOM_{f_c}\Big|_{lin}^{\text{Hooge, or Coulomb at } \theta_C\sqrt{(V_G - V_T)} > 1} = f_c = \frac{1}{4}\frac{\alpha_H}{kT/q}\mu\frac{V_D^2}{L^2} \propto V_D^2 \propto I_D^2\,. \tag{260}$$

Evidently from the above eqs. (255), (256), (257), (258), (259) and (260), the corner frequency f_c between flicker noise and white noise, taken as FOM_{fc}, has different behaviors in different regimes of operation and by different dominating noise mechanisms in MOS transistors. These are illustrated in Figure 38. Nevertheless, there are scaling rules for FOM_{fc}, which can be deduced from the equations. These rules are $FOM_{fc}\propto\lambda N_t/C_{ox}$; $FOM_{fc}\propto\alpha_H$; $FOM_{fc}\propto\theta^2$; $FOM_{fc}\propto\mu$; $FOM_{fc}\propto 1/L^2$, but FOM_{fc} is independent of channel width W. The first and second rules imply that FOM_{fc} scales with the parameters for flicker noise. The third and fourth rules imply that FOM_{fc} is sensitive to scattering and mobility. The last rule $FOM_{fc}\propto 1/L^2$, is very important for the practice, because it implies that the corner frequency increases in short channel transistors used in RF and high speed applications, and in this way limits the noise performance of these applications, e.g. phase noise in oscillators.
While the above eqs. (255), (256), (257), (258), (259) and (260) provide insights for the scaling rules, they are not in a very compact form for rapid evaluation of FOM_{fc}. The more general form for FOM_{fc} is

$$FOM_{f_c} = m\, f\, S_{V_G}(f)\frac{g_m}{4kT}\frac{g_m}{g_{m,sat}} \approx (1...2)\frac{FOM_{S_{VG}}}{WL}\frac{g_m}{4kT}\frac{A}{A_{sat}}, \tag{261}$$

where $S_{V_G}(f)$ is the gate referred flicker noise at frequency f, $FOM_{S_{VG}}= WL\, f\, S_{V_G}(f)$ is the mostly accepted figure of merit for flicker noise, see before eq. (247) in the beginning of this section, the parameter m depends on the regime of operation of MOS transistor, m=1 in linear mode, m=1.5 in saturation regime above threshold, and m=2 in saturation regime below threshold, $g_{m,sat}$ is transconductance in saturation regime (both above or below

threshold), and the ratio $g_m/g_{m,sat} \approx A/A_{sat}$ is a factor which shows what portion of the available gain in saturation regime is used when the transistor is operating in linear mode, and $g_m/g_{m,sat} \approx V_D/(V_G - V_T)$ in linear mode. Normally, one uses saturation regime of operation in order to obtain maximum gain from the transistor, thus, $g_m = g_{m,sat}$ and $A \approx A_{sat}$, and the last term is taken equal to one. Interestingly, since $4kT/g_m$ can be regarded as "thermal noise voltage in transconductance", then FOM_{fc} evaluates the flicker noise at 1Hz in ratio to this equivalent "thermal noise in g_m", which is difficult to justify physically.

Note for the above eq. (261) that $fS_{V_G}(f) = FOM_{S_{VG}}/WL$ is constant only for the case of pure number fluctuation, a very rare case nowadays, while for cases of mobility fluctuations, both correlated and uncorrelated, S_{V_G} depends also on the bias, and FOM_{fc} will capture only the order of magnitude for the corner frequency f_c, but not the exact value of f_c. Again, FOM_{fc} is a figure of merit, but not a model. Also, eq. (261) implies that lower corner frequency between flicker noise and thermal noise can be achieved in larger-area MOS transistors when keeping the aspect ratio W/L unchanged, and thus $g_m \approx$ constant, but this trades off with the high frequency performance of the transistor, e.g. lowering the transit frequency according to eq. (264) below, since the gate capacitance $C_g \approx C_{ox}WL$ increases with the channel area WL.

Figure of merit $FOM_{f_c/f_T} \equiv f_c/f_T$

An important observation can be made in eq. (261): the LF noise corner frequency f_c between flicker noise and white noise increases with the transistor gain represented by g_m in the equation. Since g_m is also participating in the expressions for high frequency performance of the MOS transistor, one can relate the noise corner frequency f_c to cut-off high frequencies, such as the transit frequency f_T, at which the current gain is decreasing to unity. The ratio f_c/f_T is a useful figure of merit $FOM_{f_c/f_T} \equiv f_c/f_T$ for comparison of flicker noise performance to the RF performance and was introduced for BJTs [210] by

$$FOM_{f_c/f_T}\big|_{BJT} = \frac{f_c}{f_T} = \frac{\pi}{q}\frac{A_E \times K_F}{\beta}\left(\tau_t J_C + \varphi_t \frac{C_B}{A_E}\right), \tag{262}$$

where, as discussed by eq. (15) in the beginning of section III. "**Noise in BJT**", A_E is the emitter area, K_F is the 1/f noise coefficient of the transistor in the SPICE flicker noise model, $A_E \times K_F \approx$ constant is a figure of merit $FOM_{S_{IB}} \approx FOM_{S_{IC}}$ for 1/f noise in BJT from a given technology. The other parameters in the equation are collector current density $J_C = I_C/A_E$, current gain of BJT at low frequencies $\beta = I_C/I_B$, transit time τ_t for the carriers injected from emitter junction to reach the collector junction, and total capacitance of the base C_B with diffusion, barrier and other components in C_B. The above eq. (262) is for the case of base-referred current noise, when the 1/f noise S_{I_B} and shot noise $2qI_B$ in the base are considered, and the shot noise of the collector current is neglected. Since the shot noise in the collector current is a sum of collector current shot noise and the coupled shot noise from the base current, then the white noise $S_{I_C,wh}$ in the collector current given by

$$S_{I_C,wh} = 2qI_C + \beta^2 2qI_B = 2qI_C(1+\beta), \tag{263}$$

and $\beta \rightarrow \beta+1$ in eq. (262), when collector shot noise is added. Evidently, the white noise increases with current gain β of BJT.

There are some issues with the above FOM when the flicker noise is not quadratic function of collector current [210], or the flicker noise has frequency slope different from 1/f, and that the improvement of FOM_{f_c/f_T} at higher

β is at the price of higher white noise in the collector current, but not due to lower level of flicker noise. Nevertheless, eq. (262) suggests that the rise of the bias level to J_C corresponding to maximum transit frequency f_T is in tradeoff with low-frequency noise performance, since the corner frequency f_c between flicker noise and white noise increases faster and f_c becomes closer to f_T, which is observed experimentally for BJT, as shown in the bottom half of Figure 39 and reported previously in [29], with data from [211, 212, 213].

Now we construct the corresponding figure of merit $FOM_{f_c/f_T}\equiv f_c/f_T$ for the ratio of the corner frequency f_c between flicker noise and white noise to transit frequency f_T in MOS transistors. At a given AC gate voltage v_g, the input AC gate current is $i_{in}=v_g 2\pi f C_g$ and the output (short circuit) AC drain current is $i_{out}=v_g g_m$, where C_g is the total capacitance seen in the gate terminal and we assume that the frequency f is not very high, so that the current through gate-drain capacitance C_{gd} is negligible, that is, $v_g g_m >> v_g 2\pi f C_{gd}$. So, the current gain is $i_{out}/i_{in}=g_m/(2\pi f C_g)$ and it decreases with increasing the frequency. The extrapolation of this frequency dependence to $i_{out}/i_{in}=1$ gives the transit frequency f_T, given by [p.501 in 169]

$$f_T = \frac{g_m}{2\pi C_g} \sim \frac{g_m}{2\pi C_{ox} WL}, \tag{264}$$

where a good estimate for the gate capacitance is $C_g \approx C_{ox} WL$, as discussed in [p.501 in 169]. Considering also the saturation in the carrier velocity, $v_{sat}\sim 10^7$cm/s in silicon [p.280 in 169], we get

$$\frac{1}{f_T} \approx \frac{2\pi C_{ox} WL}{g_m} + \frac{2\pi L}{v_{sat}} = \frac{2\pi C_{ox} WL}{g_m}\left(1+\frac{g_m}{C_{ox} W v_{sat}}\right) = \frac{2\pi C_{ox} WL}{g_m}\left[1+\frac{\mu F_V(V_G, V_D)}{v_{sat} L}\right], \tag{265}$$

where the function F_V represents the bias voltages, $F_V<2\varphi_t$ in weak inversion, $F_V\sim V_D$ in linear regime and $F_V=\max\approx(V_G-V_T)$ in saturation regime. Provided that $\mu<200$cm²/Vs and $(V_G-V_T)<1$V in technology nodes below 100nm, then $\mu F_V/v_{sat}\sim 200$nm.

In the next step, we multiply eqs. (261) and (265), and obtain $FOM_{f_c/f_T}\equiv f_c/f_T$ for the ratio of the corner frequency f_c between flicker noise and white noise to transit frequency f_T in MOS transistors, as

$$\begin{aligned} FOM_{f_c/f_T}\big|_{MOS} = \frac{f_c}{f_T} &= \left[2\pi m\left(1+\frac{g_m}{C_{ox} W v_{sat}}\right)\right]\frac{WL\, f\, S_{V_G}(f)}{4kT} C_{ox}\frac{g_m}{g_{m,sat}} \\ &\le \left[2\pi m\left(1+\frac{200\text{nm}}{L}\right)\right]\frac{WL\, f\, S_{V_G}(f)}{4kT} C_{ox}\frac{g_m}{g_{m,sat}} \\ &= m_1 \frac{FOM_{S_{VG}}}{4kT} C_{ox}\frac{g_m}{g_{m,sat}} \end{aligned}, \tag{266}$$

where m, $g_m/g_{m,sat}$ and $FOM_{S_{VG}}= WL\, f\, S_{V_G}(f)$ are following from $FOM_{f_c}\equiv f_c$, as explained by eq. (261), and the parameter m_1 depends on biasing and channel length of the MOS transistor, as following.

For operation in linear regime, $g_m/g_{m,sat}\approx V_D/(V_G-V_T)$, g_m is low, m=1, and $m_1\approx 2\pi\sim 6$. Therefore

$$\left.FOM_{f_c/f_T}\right|_{lin} = \frac{f_c}{f_T} = 2\pi\frac{WL\,f\,S_{V_G}(f)}{4kT}C_{ox}\frac{g_m}{g_{m,sat}} \approx 6\frac{FOM_{S_{VG}}}{4kT}C_{ox}\frac{V_D}{(V_G - V_T)} \quad (267)$$

For operation in saturation regime, $g_m \approx g_{m,sat} \approx (W/L)\mu C_{ox}(V_G - V_T)$, g_m is high in short channel transistors, m=1.5, and $m_1 \approx 3\pi(1+200nm/L)$. Therefore

$$\left.FOM_{f_c/f_T}\right|_{sat} = \frac{f_c}{f_T} = 3\pi\left[1+\frac{\mu(V_G - V_T)}{\nu_{sat}L}\right]\frac{WL\,f\,S_{V_G}(f)}{4kT}C_{ox} \le 9\left(1+\frac{200nm}{L}\right)\frac{FOM_{S_{VG}}}{4kT}C_{ox} \quad (268)$$

For operation in sub-threshold regime, $g_m \approx g_{m,sat}$ is low, m=2, and $m_1 \approx 4\pi \sim 12$. Therefore

$$\left.FOM_{f_c/f_T}\right|_{V_G < V_T} = \frac{f_c}{f_T} = 4\pi\frac{WL\,f\,S_{V_G}(f)}{4kT}C_{ox} \approx 12\frac{FOM_{S_{VG}}}{4kT}C_{ox} \quad (269)$$

From the above equations, in contrast to BJT, it is clear that there no unique form for $FOM_{f_c/f_T} \equiv f_c/f_T$ that is valid for all regions of operation of MOS transistors, and that the bias dependence enters FOM via gate referred flicker noise, and thus, it is dependent on the dominant 1/f noise mechanism. Nevertheless, taking only saturation regimes both above and below threshold, which are of practical interest, and denoting with $g_{m,max}$ the maximum available transconductance at high $(V_G - V_T) \sim 1V$, one can use for comparisons the following expressions

$$\begin{aligned}\left.FOM_{f_c/f_T}\right|_{V_D > (V_G - V_T)} = \frac{f_c}{f_T} &\approx 3\pi\left(1+\frac{g_m}{C_{ox}W\nu_{sat}}\right)\frac{WL\,f\,S_{V_G}(f,V_G)}{4kT}C_{ox} \\ &\approx 12\left(1+\frac{200nm}{L}\frac{g_m}{g_{m,max}}\right)\frac{FOM_{S_{VG}}}{4kT}C_{ox}, \text{ with } g_{m,max} \approx \frac{W}{L}\mu C_{ox}1V, \\ &\approx 12\frac{FOM_{S_{VG}}}{4kT}C_{ox}, \text{ for long channel, } L > 500nm\end{aligned} \quad (270)$$

The last line of eq. (270) suggests that there is a proportionality between and $f_c \propto f_T \propto g_m \propto I_D$, since f_c/f_T=constant, if $FOM_{S_{VG}}$ is constant, which is deduced in [136], but it was not explicitly supported with measured data. However, note that $fS_{V_G}(f,V_G) \sim FOM_{S_{VG}}/WL$ is constant only for the case of pure number fluctuation, a very rare case nowadays. Also, as shown by the curve labeled with Δn in Figure 39, in short channel transistors FOM_{f_c/f_T} increases at higher biasing even for the pure number fluctuation, owing to saturation in the carrier velocity. Furthermore, the correlated mobility fluctuation degrades strongly the ratio f_c/f_T, as shown by curves labeled with $\Delta\mu$ and $\Delta\mu_C$ in Figure 39 for phonon and screened Coulomb scattering, respectively, at rates $\propto (V_G - V_T)^2$ for $\Delta\mu$ and $\propto (V_G - V_T)$ for $\Delta\mu_C$, when increasing the gate bias, and these rates are reinforced by 1/L dependence caused by saturation in carrier velocity.

The use of low-frequency model with realistic values for scattering coefficients given in the caption of Figure 39 implies that the flicker noise enters the region of high frequency and high speed operation of the MOS transistors, since the combination of all components given by the curve labeled with $\Delta n-\Delta\mu_C-\Delta\mu$ in Figure 39 suggests f_c/f_T=1%−30% when the MOS transistor is biased above threshold voltage V_T in a regime of strong inversion. This occurrence of 1/f noise close to f_T can cause many difficulties in the RF applications of MOS transistors. Unfortunately, and in contrast to the research on BJT, there is no publication for MOS transistors, which reports f_c and f_T simultaneously from measurements, which is needed to verify whether the extrapolation of low-frequency noise over many decades is accurate for the practice. The recalculation of the partially available data [86, 102, 105, 136, 214] is shown with symbols in Figure 39 and the recalculation uses the equations deduced in this section. The recalculated data suggest the same range for f_c/f_T predicted by using eq. (270) for FOM_{f_c/f_T} and taking typical values for the parameters related to low-frequency noise in MOS transistor. However, different trend in the bias dependence of FOM_{f_c/f_T} is observed from the reported data, as compared to the bias dependence deduced from (270). The trend is shown in Figure 39 with a thick gray line denoted by a question mark. This is an issue for the understanding of FOM_{f_c/f_T}. Interestingly, since $4kT/(WL2\pi fC_{ox})$ can be regarded as "thermal noise voltage in gate capacitance conductance", then FOM_{f_c/f_T} evaluates the flicker noise at 1Hz in ratio to this equivalent "thermal noise in gate capacitance WLC_{ox}", a relation which is also difficult to justify physically.

Look again at Figure 39 below the diagonal-patterned area, which separates MOS from BJTs. While it is clear that the ratio f_c/f_T in BJT and HBT increases with the bias level, f_c/f_T is much lower than FOM_{f_c/f_T} in MOS transistors. Nevertheless, FOM_{f_c/f_T} is independent of the device size both for MOS transistors and BJT.

V. Noise in advanced Si-based transistors

Silicon takes the major place in semiconductor technologies, owing to low leakage in silicon-based devices at room temperatures and excellent lattice quality when silicon monocrystals are grown, accompanied with mechanically and chemically stable silicon oxides for low leakage surface passivation and with low trap density for gate insulators. However, the downscaling of microelectronic devices for larger integration and the requirements to speedup the devices reached the limits for mobility of carriers in silicon. The body bias became an option to boost the performance of the MOS transistors. Also, the mobility is related to band structure of the semiconductor and the so-called "band engineering" is employed to enhance the mobility in silicon based layers. The low-frequency noise in these devices is discussed next.

V.1. Noise in SiGe, stacked gate and strained transistors – higher performance, but higher noise too

The band engineering takes two approaches simultaneously. One is to introduce strain or compress the silicon lattice, which modifies the electron and hole mobility. The second is to use silicon alloys with materials, which have higher mobility. Germanium (Ge) has been re-introduced in silicon devices [215], because Ge has higher electron and hole mobility, as compared to Si, the methods for addition of Ge in Si are compatible with silicon processing, Ge and Si have similar covalent bonding in monocrystal lattices, and Ge also induces strain when alloyed with silicon, which can be relaxed by using carbon, if necessary. Currently, SiGe heterojunction bipolar transistors (HBTs) are in the main stream of high-speed bipolar semiconductor technologies, SiGe HBTs are introduced in some CMOS technologies for RF applications [216, 217, 218], and SiGe MOS devices are also in a phase of intensive development [219, 220, 221, 222, 223].

V.1.1. Observations in SiGe HBTs

While it is well established that SiGe devices are faster than Si counterparts, the low-frequency noise performance of these devices is addressed widely and results from [33, 34, 35, 36, 37, 60, 81, 212, 224, 225] for 1/f noise in SiGe HBTs are shown in Figure 40 in terms of the SPICE parameter $K_F=f\times S_{I_B}/I_B^2$, defined by eq. (15). The data are stored in numerical form in [226]. The larger symbols in Figure 40 correspond to SiGe HBTs, and for the purpose of comparison, the smaller symbols and trends are for Si BJTs, as given earlier by Figure 2 and Figure 14. As seen in Figure 40, the 1/f noise in SiGe HBTs remains in the range observed in Si BJTs, and therefore, we did not discriminate SiGe HBTs from Si BJT in section III. "**Noise in BJT**".

The addition of Ge in the base of bipolar transistors has some processing issues [225] and difficulties to obtain uniform current flow in the base of the transistor [35], especially in proximity of trench isolation and small polysilicon emitter overlap [37], which may cause increased 1/f and RTS noise in small-area BJT [227]. This is illustrated in Figure 41. In the left-hand plot, as shown with thick lines by data from [224, 225] for npn SiGe HBTs, when changing the Ge content, but keeping other processing, bias and layout factors similar, the 1/f noise coefficient K_F normalized with the emitter area, see eq. (15), does not change significantly, and, it seems, that Ge improves the low-frequency performance of npn HBTs by approximately 20%-30% (~1-2dB), but the noise remains in the range observed in Si BJTs, as seen in the previous Figure 40. On the other hand, changing the epitaxial growth temperature or annealing temperatures (vertical lines in left-hand plot of Figure 41 with data from [35, 225]), the noise changes 1–2 decades. In the right-hand plot Figure 41, aggregated data are given for SiGe HBTs with different sizes and biasing levels from [33, 34, 36, 37, 60, 81, 212, 224], showing that the 1/f noise does not scale with the maximum transit frequency f_T of the technology nodes in a regular manner. Rather, other factors, such as aforementioned bias, layout (especially proximity of trench isolation and polysilicon emitter overlap), SOI, annealing, etc., impact the 1/f noise in HBTs.

While overall Ge does not impact the noise in SiGe HBTs, the addition of carbon (C) to the SiGe semiconductor does improves the $K_F=f\times S_{I_B}/I_B^2$ in SiGeC HBTs, with $(A_E\times K_F)_{avg}= 1.25\times10^{-10}$µm², 16.5dB below the range for npn SiGe HBTs (data shown with (■) in Figure 40 and taken from [40, 64, 65, 228, 229, 230]). Also, as observed in [64], SiGeC HBTs have a much lower noise response compared with the HBTs from other technologies.

On the other side, the addition of Ge in the channel increases the flicker noise in SiGe MOS transistors. This increase was illustrated by the triangles in the bottom-left plot of Figure 22 earlier in section IV.1. "**Models and predictability**", since the triangles are in the upper half of the noise distribution, above the trend for the noise in MOS transistors. In the same figure, we observed that the noise is not increased in MOS transistors with channel of silicon strained on SiGe layers.

V.1.2. Observations in SiGe MOS transistors

To get insight on how Ge affects the noise in SiGe MOS transistors, we arrange the data from [48, 52, 57, 79, 109, 139, 155, 207, 208, 231] in several data series (a), (b) … (h) shown in Figure 42, for the oxide trap density N_t , which is proportional to the level of the input referred 1/f noise voltage S_{V_G}, according to eqs. (104) and (241) based on the number fluctuation model for 1/f noise. Data (a) are for pMOS transistors with channel of ternary SiGeC alloy, in which N_t increases with increasing the content of carbon in the alloy, and it is higher than N_t in MOSFET made of pure silicon. This implies that the crystallinity in compound IV semiconductors is not as good as that of pure silicon or the quality of the cap layer degrades when it is grown on the top of the

ternary SiGeC alloy.

Data (b) and (d) are for silicon nMOS and pMOS transistors, in which an increase of noise and N_t is observed when using nitridation of silicon oxide dielectric in order to increase the gate breakdown voltage. Data (c) are from the attempts to increase the oxide capacitance using high-k dielectrics in Si nMOS transistors, which are accompanied also with increase of noise and the corresponding values for N_t. Data (b), (c) and (d) imply that a departure from SiO_2 in gate dielectrics causes increase of 1/f noise.

Data (e) are from several publications, which reported variation of 1/f noise when changing the MOS channel alloy from Si to SiGe, then using high-k dielectrics in the gate stack in transistors with SiGe alloy in the channel, followed by semiconductor-on-insulator (SOI) structures based on Si and on Ge. The reported values for N_t increased in this order of increasing complexity of the MOS structures.

Data (f) are from a research, in which annealing after complete fabrication of SiGe MOS transistor was carried out. The post-annealing with inert Argon gas reduced N_t, while the post-annealing with water vapor increased N_t, owing to removal or adding of oxide traps, respectively.

Overall, data (a) to (f) imply that there is an increase of the 1/f noise when using composite materials either for semiconductor or insulator, or adding non-uniformity in the transistor structure, which, perhaps, degrades the lattice quality in the regions related to the current transport in MOS transistor, e.g. conductive region of the channel and semiconductor-dielectric interface. These observations are cumulatively established, since the introduction of interfacial layer (~0.8−1.2nm) of SiO_2 between semiconductor and high-k dielectric [49, 50, 92, 99, 100], as well as the insertion of Si cap layers (~2−8nm) on the top of SiGe conduction channel [231, 232] usually result in lower noise, perhaps due to smoother transition between different materials and reduced density of point defects at the interfaces. This is further confirmed by using the strained Si cap layer as the conductive channel rather than the SiGe layer.

Since the Si lattice on the top of a SiGe layer is strained, then the electron mobility in Si can be increased, but the lattice quality of strained Si and the uniformity of SiO_2 grown on the top of Si layer are well preserved. Consequently, as shown with data (h) in Figure 42, the noise in nMOS transistors with strained Si channel is only slightly higher than the noise in “regular, control” MOS transistors with unstrained silicon, given by data (g), respectively.

It is interesting to mention that the way of introducing the strain in Si transistors is important for the 1/f noise. If the strain is applied from a region which does not participate in the current transport, then there is virtually no change in the 1/f noise. This has been demonstrated in [126], where a Si_3N_4 layer on the top of the gate for compression of Si layer below the gate (and enhancement of the hole mobility in pMOS transistors) did not increase the noise, while the use of SiGe for drain and source, which also compress the silicon in the channel, resulted in increase of the noise, because the SiGe regions participate in the path of the current flow and the out-diffusion of Ge toward the channel is possible during device processing. For locally strained devices like the above pMOS transistors [126] or small area MOS transistors, the strained Si layer does not introduce additional noise, since the point defects are few in the small devices. In larger-sized MOS transistors, however, threading dislocations in the strained layer are statistically occurring, and the noise can increase. The areal dependence of this increase was modeled in [82] using Poisson distribution for the extra traps in the dislocations and geometric averaging. This one more time confirms that the increase of the noise in SiGe and strained-Si MOS transistors is due to material defects in the path of the current transport, but it is not a fundamental material property of SiGe

alloys and strained Si lattices, which is somehow in contrast to the explanations in [127] for the effect of the mechanical stress from shallow trench isolation in pMOS transistors.

V.1.3. Modeling the 1/f noise in SiGe MOS transistors

Although the above observations are well established experimentally, the modeling and the understanding of the physics behind the 1/f noise in SiGe MOS transistors are still not very certain. In most cases, the 1/f noise models for silicon transistors are used, c.f. portion or the whole eq. (241), with some enhancements for Hooge noise from drain and source contact regions [208] and modifications in the correlated mobility term, splitting it into two parts for screened Coulomb and phonon scattering [52, 57]. For the case of SiGe channel and Si cap layer on the top of it, it is suggested in [232] to use current partitioning between the potential well in SiGe at the Si interface and the well in Si at gate oxide. This partitioning can be lumped formally into two transistors, one in SiGe layer and another in the Si cap layer, which are driven by the same flat-band noise voltage S_{FB}. The resulting equations are

$$S_{I_D,\Delta n-\Delta\mu} = S_{FB}\left[g_{m,Si}\left(1+\alpha_s C_{ox}\mu_{Si}\frac{I_{D,Si}}{g_{m,Si}}\right)+g_{m,SiGe}\left(1+R\alpha_s C_{ox}\mu_{SiGe}\frac{I_{D,SiGe}}{g_{m,SiGe}}\right)\right]^2 \quad (271)$$

$$S_{I_D,Hooge} = \frac{1}{f}\left[\frac{\alpha_{H,Si}}{n_{Si}}I_{D,Si}^2+\frac{\alpha_{H,SiGe}}{n_{SiGe}}I_{D,SiGe}^2\right] \quad (272)$$

$$S_{I_D} = S_{I_D,\Delta n-\Delta\mu}+S_{I_D,Hooge} \quad (273)$$

where S_{FB} is the same as for Si MOS transistors and given by eq. (109), the total noise S_{I_D} in the drain current is a sum of correlated number-mobility fluctuation $S_{I_D,\Delta n-\Delta\mu}$ and Hooge noise $S_{I_D,Hooge}$, α_s is the scattering parameter given by eqs. (127) and (136), C_{ox} is the gate oxide capacitance per unit area, and the parameter R~0.1–0.2 is taking into account the remote interaction between SiGe layer and oxide traps, separated by the Si cap layer. The other parameters are respectively for Si cap layer and SiGe channel, and they are transconductances $g_{m,Si}$ and $g_{m,SiGe}$, carrier mobilities μ_{Si} and μ_{SiGe}, Hooge parameters $\alpha_{H,Si}$ and $\alpha_{H,SiGe}$, DC currents $I_{D,Si}$ and $I_{D,SiGe}$, and total number of carriers n_{Si} and n_{SiGe} in the corresponding layer. It has been shown in [232] that this model can produce non-monotonic dependences for the noise levels as function of bias, thickness of cap layer and Ge content in the SiGe layer.

However, the above equations assume that the Si and SiGe parts of the transistor conduct independently by neglecting the charge transfer between the two potential wells, which occurs when the transistor is operating in saturation regime well above the threshold in strong inversion, and the charge from Si layer at the source side is transferred into SiGe well along the channel length in the region close to pinch off point. Furthermore, the model requires a precise knowledge for the structure, e.g. layers' thicknesses, concentrations, mobility, etc., which are not always available, and the correct partitioning can be done only after numerical simulation of the structure by using a semiconductor simulator. The issue with this noise model for SiGe MOS transistors, in fact, is that there is no mature characterization technique which can obtain the model parameters with unique values.

V.2. Forward body bias in MOS transistors – not a panacea, but it helps

In many circuit applications, especially in low-voltage circuits, the potential of the body and source of MOS transistors are different. Examples are dynamic threshold configurations [233, 234, 235], in which the body and

gate of the MOS transistor are tied together in order to obtain higher transconductance, differential amplifiers, when the transistor pair and biasing current source are in one well or in the substrate, or the body of the MOS transistor is used as input [236, 237, 238], cross-coupled pairs for oscillators [239, 240], in which the body is used for tuning of the oscillation. Different approaches for power management in digital circuits [241, 242, 243, 244] also use the body bias as a control to bring inactive portions of the circuit in sleep mode, applying reverse body–source voltage V_{BS}, or applying forward V_{BS} to accelerate the circuit in active mode and at low supply voltages.

The low-frequency noise in MOS transistors, however, is sensitive to the body biasing, as illustrated in Figure 43 with data from [245] for a pMOS transistor from 0.18μm technology. In the left-hand plots, the data for several quantities, which represent the 1/f noise at 1Hz, are shown at different body–source bias voltages V_{BS}, varying also gate bias voltage V_G, and at unchanged drain bias voltage V_D=0.6V. The different V_{BS} are from -0.6V (reverse) to +0.6V (forward). Reverse or forward body bias is in respect to the direction for conduction of the body–source p–n junction. The quantities DC drain current I_D, small signal transconductance g_m and their ratio in the horizontal axes are obtained from I–V curves of the transfer characteristics. The insets in the figures show that the these quantities depend on the gate overdrive (V_G–V_T), thus on carrier concentration $n' \propto (V_G - V_T)$ in the transistor channel, since the values for I_D and g_m (and their ratio, consequently) overlap when V_{BS} was varied, but (V_G–V_T) is kept constant.

In similar manner, the effect of body biasing on 1/f noise in one nMOS transistor was attempted to be explained in terms of carrier density n' in [51], but measured data at only one constant gate bias were reported in this publication. However, the case is different for the low-frequency noise, as one can see from the data series for the noise in Figure 43. In particular, the 1/f noise is not a unique function of areal carrier density n' in the channel, because at given values for the quantities in the horizontal axes, and therefore constant n' as follows from above discussion, the noise decreases by a transition from high gate bias and reverse body bias toward low gate bias and forward body bias. This transition indicates a decrease of noise when the current flow is moved from the semiconductor-dielectric interface (surface channel at reverse body bias and high gate bias) toward the bulk of the semiconductor (buried channel conduction at forward body bias and low gate bias). Such transition in more pronounced form (from Δn noise at surface MOS channel to Hooge noise in the bulk JFET channel) is observed in [246] for a MOS–JFET SOI structure with gates around the conduction body. Further discussion on the crossover from surface to bulk noise is given section VI. "**Noise in advanced transistor structures**".

As follows from the above observations, there is an issue related with the body biasing of MOS transistors. The problem is that all the flicker noise models are derived at the assumption for surface conduction and charge sheet approximation for channel, and the noise models allow for bias variation only of the areal concentration n' of charge carriers in the MOS transistor channel. Consequently, the noise parameters, either in number fluctuation or mobility fluctuation models, have to be varied with the body bias V_{BS}. Therefore, the compact noise models used for computer simulations are unable to reproduce effects related to the depth of the channel and the simulations of some specific connections of MOS transistor can be inaccurate. For example, applying reverse bias to MOS transistors in voltage controlled oscillator (VCO) will underestimate the phase noise, and using the dynamic threshold configuration (with gate and body tied together) will require another set of values for the coefficients in the noise model.

Qualitatively, the channel depth in pMOS transistor is depicted in the right-hand plots of Figure 43 for reverse,

zero and forward body bias (plots from top to bottom) at similar charge sheet concentrations (the dotted shapes for the charge carrier layers denote same area). Quantitative examples for the charge concentrations in Si and SiGe MOS transistors can be found in [247]. To the best of our knowledge, there is no 1/f noise model for MOS transistors that considers the volume distribution of the carriers in the depth of the channel, which can possibly explain the body bias dependence with variation of current density in terms of Hooge noise in its integral form of eq. (13), neither 1/f model, which considers variable distance from oxide traps to carriers in the channel, a distance which is usually taken zero for surface channels or constant in SiGe channels capped with Si layer, and which can possibly explain the body bias dependence with variation of Coulomb interaction between oxide traps and channel carriers in terms of number fluctuation models. Overall, the charge sheet approximation works well for DC and AC modeling of MOS transistor [233, 248], but it is not accurate for the noise.

Nevertheless, the practical rule is that the forward body bias V_{BS} reduces the noise in relative units, while the reverse V_{BS} increases the noise, as illustrated in Figure 44 with data from [52, 245] for the gate referred 1/f noise S_{V_G} in Si and SiGe pMOS transistors in saturation and linear regimes of operation, respectively. A similar observation for the dependence of gate referred noise voltage on body bias can be found in [247] for pMOS Si and SiGe, in [249] for nMOS and pMOS transistors from a 130nm technology in weak inversion, in [250, 251] for SOI MOS transistors operating in dynamic threshold (body tied to gate) and normal (body tied to source) connections. The data in [249, 250, 251] have been explained in terms of number fluctuation model for the 1/f noise in MOS transistors, see eqs. (104), (109), (120) and (126), and sheet approximation for channel charge carriers coupled to trapped oxide charge via oxide C_{ox} and depletion C_d capacitances (per unit area). From the analyses in these publications, it can be shown that the gate referred noise voltage is a result of coupling to the total capacitance (C_{ox}+C_d) seen at semiconductor-dielectric interface. That is,

$$S_{V_G} = \frac{S_{I_D}}{g_m^2} \approx \frac{1}{f}\left(\frac{q}{C_{ox}+C_d}\right)^2 \frac{kT\lambda N_t}{WL}\left[1+\alpha_s\left(C_{ox}+C_d\right)\mu_{eff}\frac{I_D}{g_m}\right]^2. \tag{274}$$

This equation reproduces the behavior of the 1/f noise as function of the body bias, because the depletion capacitance C_d increases with forward V_{BS}, resulting in decrease of S_{V_G}, and C_d decreases with reverse V_{BS}, resulting in increase of S_{V_G}. Also, the equation is in agreement with the general equations (4) and (5) for noise coupled from charge fluctuation. Eq. (274) was derived in [251] for dynamic threshold (DT) configuration (body tied to gate) in the form of

$$S_{I_D} = g_m^2\left(\frac{1}{1+\frac{C_d}{C_{ox}}} \pm \alpha_s C_{ox}\mu_{eff}\frac{I_D}{g_m}\right)^2 S_{FB}, \text{ with } S_{FB} = \frac{1}{f}\left(\frac{q}{C_{ox}}\right)^2 \frac{kT\lambda N_t}{WL} \tag{275}$$

where the oxide charge trapping is referred to the gate and attributed to noise in flat-band voltage V_{FB}, see eq. (109), and it was experimentally validated by the similar values for oxide charge density N_t obtained from DT connection after correction $(1+C_d/C_{ox})^2$ and normal connection (body tied to source) without correction. The attenuation of the noise by $(1+C_d/C_{ox})^2$ was also confirmed in [249] for weak inversion, where it was also observed that the body bias has no effect in strong inversion due to screening effect that the inversion layer charge has on C_d. There are also other issues with the physical consistence of eq. (274), such as the question why

C_d is omitted when body and source are tied together, and the explanations on these issues are mostly qualitative in the publications cited above. Overall, eq. (274) is useful for the practice, because $(1+C_d/C_{ox})^2<2$, which is within the experimental inaccuracy for characterization of N_t from noise measurements at low gate overdrive voltage $(V_G-V_T)<0.2V$, while at higher overdrive, when the term for correlated mobility fluctuation dominates, the capacitances cancel in (274), and the values for the scattering parameter α_s are unaffected by the choice $C=C_{ox}$ or $C=(C_{ox}+C_d)$. Thus, the addition of C_d in the number fluctuation model does not compromise with the convergence of the model, and gives a straightforward way to introduce the body bias dependence in this model.

VI. Noise in advanced transistor structures

There are several reasons to focus on advanced silicon transistor structures and noise in them [3]. The semiconductor industry has been based for more than 40 years on scaling of transistor dimensions to achieve performance gains, utilizing tremendous investment in infrastructure to the highest possible degree. While the role of nanoscale devices in meeting future computing and communications applications is not clear at present, there are significant limitations that arise with nanoscale devices and will impact their usefulness. In particular, the near-term applications require nanoscale devices to be functionally and technologically compatible with silicon transistors, at least for using semiconductor manufacturing and design infrastructures, and for interfacing to scales accessible by human.

On the other hand, there are many difficulties by downscaling silicon devices, even CMOS transistors, for which the advances are the most, because fundamental limitations for charge transport, electrostatic control and accuracy of device fabrication are reached in planar technologies, in particular bulk CMOS, and binary logic (state variables) based on electric charge are vulnerable to random errors due to low signal to noise ratio, and thermal and material instability.

In order to improve the electrostatic control, the advanced silicon transistors utilize the vertical dimension of the structures, by developing approaches originally introduced for SOI. In this section, we review the results obtained for several 3-D transistor devices, Fin FET, Two-, Three- and All-Around-Gate FETs, along with complications for the low-frequency noise in SOI MOS transistors. At the other end, the enhanced 1-D current transport in carbon nanotubes and semiconductor nanowires attracted the attention recently, and we also review in this section results presently available for the low-frequency noise in devices based on 1-D charge transport.

VI.1. From SOI toward gate all around

In attempt to enhance the electrostatic control of the surface charge transport in MOS transistors, the body of SOI MOS transistors is made thinner and the semiconductor bulk is removed and replaced with insulator. The removal of the semiconductor bulk resolves problems related to drain induced barrier lowering (DIBL) caused by the high electric field in the bulk of MOS transistors owing to drain bias and uncontrolled by the gate bias [252]. The corresponding structures are known as partially depleted (PD) and fully depleted (FD) SOI MOS transistors, which are depicted in Figure 45.

VI.1.1. Partially depleted SOI MOS transistors

The transition from bulk MOS (Figure 45a) to partially depleted SOI MOS (Figure 45b) is accompanied with reduced control of body potential V_B. In bulk MOS, the bias voltage of the body terminal sets reliably the level of V_B, because any excess charge generated or entered in the body (e.g. due to impact ionization or valence band tunneling through thin gate oxides) is sunk-out through the low impedance of the conductive bulk, as illustrated

by arrows in the figure. Thus, the depletion region under the MOS channel (dashed line in Figure 45a) is fixed by the gate bias accordingly, and it does not vary randomly, once the body terminal (normally tied to source terminal) is connected to low impedance bias. Consequently, the low-frequency noise in the channel current is coupled only from the gate oxide trapping and charge transport in the inversion layer, as discussed in previous sections.

However, the control of the body potential V_B is weak in PD SOI MOS (Figure 45b), because the thin bulk has high impedance to the body terminal, or the body terminal is removed in order to reduce the size of the transistor. Therefore, the excess charge generated or entered in the body increases the body potential V_B so that the junction body-source becomes forward biased and the excess charge flows toward the source terminal overcoming the impedance of the body-source junction. In this way, the body potential V_B becomes dependent on the body current I_B and the effects related to this dependence are depicted in Figure 46 with data from several publications [250, 252, 253, 254, 255].

As shown in Figure 46a, when the drain V_D and gate V_G bias voltages increase, then the electric fields at the drain side of the channel and in the gate of the MOS transistor increase, causing leakage, impact ionization and valence band tunneling currents I_B, which flow into the body [250, 251, 252, 253, 254, 255, 256, 257, 258, 259]. The body current I_B has to be readily sunk by the body-source junction, since all other interfaces surrounding body in SOI MOS transistor are either insulators or reverse biased junctions. Once I_B becomes larger than the (reverse) saturation current I_{B0} of the body-source junction, then the body voltage V_B increases, as shown in Figure 46b, and the variation in I_B with bias causes changes in V_B. The increase of V_B reduces the threshold voltage V_T of the transistor and undesirable kink in the transistor DC characteristics occurs, as illustrated in Figure 46c. Consequently, the noise S_{I_B} in I_B causes voltage noise $S_{V_B}=Z_B^2 S_{I_B}$ in the body potential, where Z_B is the AC impedance of the body to ground, which is further coupled to the inversion layer as excess voltage noise $S_{V_{G'},ex}$ by the body coefficient $\partial V_T/\partial V_B$.

Since I_B is low, usually less than a nanoampere, then the 1/f noise in I_B is negligible, and the shot noise in I_B dominates, because the charge carriers related to I_B cross the gate and drain potential barriers to enter the body, and body-source barrier to exit from the body. Assuming uncorrelated processes, then shot noises for currents entering from gate and drain into body (G&D→B) and exiting from the body toward source (B→S) are summed, and the noise S_{I_B} in I_B is

$$S_{I_B} \approx 2qI_B\big|_{G\&D\to B} + 2qI_B\big|_{B\to S} = 4qI_B\,, \tag{276}$$

where $q=1.6\times10^{-19}$ C is the electron charge, S_{I_B} is frequency independent (white noise) conservatively for the range of low frequencies, and S_{I_B} is proportional to the body current I_B.

The body impedance $Z_B=1/(G_B+j2\pi fC_B)$ can be lumped into a simple RC equivalent circuit [97, 251, 258, 255] of parallel connection of body-source junction dynamic conductance $G_B\approx I_B/\varphi_t$ and body capacitance C_B. The later has several components, e.g. depletion capacitance under the MOS channel, junction capacitances of the drain and source junctions, bottom oxide (back-gate) capacitance and other, depending of the SOI MOS transistor layout, but C_B varies less with bias, as compared to G_B, which is proportional to I_B, and I_B is close to exponential function of V_G and V_D. Thus, Z_B acts as a low-pass filter for S_{I_B}, and the voltage noise S_{V_B} in the body potential has a Lorentzian spectrum, given by

$$S_{V_B} \approx Z_B^2 S_{I_B} \approx \frac{4qI_B}{\left|G_B + j2\pi f C_B\right|^2} = \frac{S_{V_B}(0)}{1+(f/f_0)^2}, \tag{277}$$

where with φ_t=kT/q being the thermal voltage, the low-frequency plateau $S_{V_B}(0)=4qI_B/(G_B)^2 \approx 4q\varphi_t^2/I_B$ of the Lorentzian noise is inversely proportional to I_B, while the corner frequency $f_0=G_B/(2\pi C_B) \approx I_B/(2\pi C_B\varphi_t)$ is proportional to I_B, and S_{V_B} is reproduced by the body coefficient $\partial V_T/\partial V_B$ as gate referred excess voltage noise $S_{V_{G,ex}}$ given by

$$S_{V_{G,ex}} \approx S_{V_B}\left(\frac{\partial V_T}{\partial V_B}\right)^2 = \frac{S_{V_{G,ex}}(0)}{1+(f/f_0)^2}, \tag{278}$$

where $(\partial V_T/\partial V_B) \approx$constant is a weak function of the bias of the transistor. Thus, since I_B increases with gate and drain biases of the PD SOI MOS transistor, then the excess Lorentzian noise evolves with the bias of the PB SOI MOS transistor. The evolution of the excess Lorentzian noise is as shown in Figure 46e as function of the drain bias at constant gate overdrive (V_G-V_T)=constant. Apparently, at given low frequency and varying the drain bias, one will observes noise "overshot" due to the evolution of the $S_{V_{G,ex}}$ with V_D. The overshot is correlated to a biasing condition at the onset of the kink in DC characteristics, as illustrated in Figure 46d, but note that the position of maximum of the overshot is dependent on the selected frequency and it is not unique function of the biasing condition [250]. In fact, one can mistakenly conclude that that the noise power has a maximum at particular bias. This issue is discussed in [88, 259], where is shown that the total "power" $P_{V_{G,ex}}$ of the gate referred excess noise voltage in a wide frequency band is a weak function of the bias, when considering that

$$P_{V_{G,ex}} = \int_0^{f>>f_0} S_{V_{G,ex}} df \approx \frac{\pi}{2} f_0 S_{V_{G,ex}}(0) = \frac{q\varphi_t}{C_B}\left(\frac{\partial V_T}{\partial V_B}\right)^2 = \frac{kT}{C_B}\left(\frac{\partial V_T}{\partial V_B}\right)^2, \tag{279}$$

which follows from above analysis, because the body capacitance C_B and the body coefficient $(\partial V_T/\partial V_B)$ do not vary much with gate and drain biases. The experimental data in [88] imply that $P_{V_{G,ex}}$ increases with gate and drain biases mostly due to faster increase of f_0, as compared to the decrease of $S_{V_{G,ex}}(0)$, which can be attributed to the change of the body capacitance with bias.

Detailed analyses in [255, 259] demonstrated that the thermal noise in S_{I_B} can be also included, as well as the matching to experimental data is very well, when using the accurate models for the capacitances and resistance associated to the body of the MOS transistor. The above characterization approach was also confirmed for many SOI structures, in which the impact ionization [250, 251, 254, 255, 259] or the gate valence-band tunneling [88, 250, 253, 257] prevails, as well as for twin-gate MOS transistors [253, 257] and at different depletion levels of the body of PD SOI MOS transistor, achieved by varying the bias of the back gate of the SOI MOS transistor [88, 97, 256].

To summarize, for MOS transistors with floating partially depleted or undepleted bodies, an excess noise is coupled from the fluctuation S_{V_B} of the body potential into drain current noise $S_{I_D}=(g_{mb})^2 S_{V_B}$ by the body-drain transconductance g_{mb}. S_{V_B} can have several components, but the major contribution is from shot noise in body currents, owing to drain and gate leakages at high electric fields, and S_{V_B} has a Lorentzian spectrum, since the

shot noise is with white spectrum and it is filtered by the body impedance $Z_B=1/(G_B+j2\pi fC_B)$, where G_B is larger or equal to the conductance of body-source junction and C_B depends on the layout of the structure, and it is larger than the depletion capacitance of the MOS transistor. Any additional connection to the body reduces Z_B, and thus, reduces the excess noise. The excess noise is in close connection with the general equation (1) for coupling of noise, and it does not require "superficial" explanations or identification of extra generation-recombination noise sources, as one can find in old publications, e.g. [254]. Also, if the excess noise is due to body currents, then the corner frequency of the excess Lorentzian noise is the same as the corner frequency of the output AC conductance g_d of the drain [255, 258], and both corner frequencies are associated with Z_B and kink effect due to frequency dependent variation of body potential when the body is floating.

The above discussion provides a coherent picture, when extrapolating back from partially depleted SOI MOS transistors to bulk MOS transistors. Consider again Figure 45. The body capacitance is large in bulk MOS, and the body impedance is reduced, when the body terminal is connected to the source terminal. Thus, the excess noise in bulk MOS is low, although this noise was observed [260].

VI.1.2. Fully depleted SOI MOS transistors

The body is missing in fully depleted (FD) SOI MOS transistors, Figure 45c, and one cannot attribute noise to fluctuation of body potential. The experiments [255] and computer simulations [259] also showed that the kink effect and excess Lorentzian noise are suppressed when the body is fully depleted, but the FD SOI MOS transistors are not completely free from excess Lorentzian noise. This noise is just with smaller magnitude and high corner frequency (>100kHz), indicating high conductance G_B of the "body"-source junction [255].

The reduction of G_B is explained with reduction of "body"-source junction barrier, since the "body" potential V_B at source junction is modified by the front and back gate biasing in FD SOI MOS transistors, and it is different from the Fermi potential V_F set by doping of the body in partially depleted and bulk MOS transistors. Thus, the reduction of the barrier potential is $\Delta V_{rb}=V_B-V_F$ and the corresponding (reverse) saturation current $I_{B0,FD}$ of the "body"-source junction in FD SOI MOS transistor is increased

$$I_{B0,FD} = I_{B0}\exp\left(\frac{\Delta V_{rb}}{\varphi_t}\right), \text{ with } I_{B0,FD} >> I_B >> I_{B0} \tag{280}$$

as compared to body current I_B and junction saturation current I_{B0} in partially depleted MOS transistors. Consequently, at a given body current I_B, the "body"-source junction conductance $G_{B,FD}$ in FD SOI MOS transistor increases significantly, according to

$$G_{B,FD} = \frac{I_B + I_{B0,FD}}{\varphi_t} \approx \frac{I_{B0,FD}}{\varphi_t} >> G_B = \frac{I_B + I_{B0}}{\varphi_t} \approx \frac{I_B}{\varphi_t}, \tag{281}$$

and according to eq. (277), the excess Lorentzian noise has reduced magnitude and higher corner frequency, which are weakly dependent on the body current I_B in FD SOI MOS transistors.

An alternative explanation for the reduction of the Lorentzian noise FD SOI MOS transistors is given in [259], referring to the strong coupling between the front and back gates, which reduces the "body" transconductance g_{mb}, and prevents the shot noise from source junction being transferred to the drain. The explanation is qualitative and does not provide insight for the shift of the Lorentzian corner frequency. Nevertheless, both explanations in [255] and [259] consider bipolar effects and modulation of potentials at body-source junction,

which is viewed by the dashed lines in Figure 45c. The issue is that such modulation is generally neglected in MOS transistor models, and the source-channel interface is not considered in low-frequency noise models, to the best of our knowledge.

VI.1.3. Effects of oxide traps in the back gate insulator

As the silicon film becomes thin in fully depleted SOI MOS transistors, the oxide traps in back gate insulator (usually called buried oxide, BOX) also contribute to the low-frequency noise, and the most common approach is to divide the conduction in MOS transistors in front and back channels. Then, one applies the noise models separately for the two channels, and the noise in the drain terminal is the sum of the contributions of the two channels plus the noise from the semiconductor bulk between them in depletion mode MOS transistors. This approach was taken long time ago, for example in [93], and results in a simple equation, such as

$$S_{I_D} = S_{FrontGate} + S_{BackGate} + S_{BulkVolume}, \tag{282}$$

which is correct only if the three conduction paths are independent and isolated each from other, so that the noise source related to one of them does not affect the other. Obviously, this is not true, because the capacitances of the insulators and of the semiconductor couple the potentials from the front gate all along to the back gate, and the coupling is significant in FD SOI MOS structures, since the semiconductor film is thin. Partial solutions based on the above eq. (282) are possible and used in device characterization when the dominant conduction path is one, i.e. only front channel [88, 93, 246], only back channel [93], or only the bulk volume [93, 246], the later in the case of depletion mode MOS transistors. At crossover biasing regimes [93, 246], one always observes excess noise mostly with Lorentzian spectrum, and since the Lorentzian spectrum varies with the bias, then one attributes this variation to "superficial" traps [93, 246], while the excess noise is probably due to frequency limited coupling, similarly to the case for the excess Lorentzian noise in partially depleted SOI MOS transistors, as discussed just above.

The partial solution of eq. (282) is proposed in [88] for the front channel of FD SOI MOS transistors, following the approach for noise coupling to the bottom of the channel [251]. The proposal assumes two capacitances, $C_F \approx C_{ox}$ of the front gate oxide above the conduction channel and below the conduction channel $C_S \approx (1/C_d + 1/C_{BOX})^{-1} \approx C_{BOX}$, where the back oxide capacitance C_{BOX} is much smaller than either C_{ox} or the capacitance C_d of the depleted silicon film in FD SOI MOS transistor. Then, the coupling (attenuation) coefficient $\gamma = \partial V_T / \partial V_B$ of the front gate threshold voltage V_T to the back gate bias voltage V_B is taken as from the capacitive divider C_F–C_S, given by

$$\gamma = \frac{C_S}{C_F + C_S} \approx \frac{C_S}{C_F} \approx \frac{C_{BOX}}{C_{ox}} < 1, \tag{283}$$

where the highly conductive inversion charge layer is assumed disconnected from the source terminal, thus from ground, an assumption not stated, but seen from equations in [88]. Finally, the flat band voltage noise S_{FBB} from back oxide is referred to the front gate as $\gamma^2 S_{FBB}$ and added to the flat band voltage noise S_{FB} of the front oxide, resulting in gate referred voltage, given by

$$S_{VG} = S_{FB} + \gamma^2 S_{FBB}, \tag{284}$$

where S_{FB} and S_{FBB} are expressed also according to eq. (101) for the number fluctuation model, and the noise in the drain current becomes

$$S_{I_D} = g_m^2 S_{VG} = g_m^2 S_{FB} + g_m^2 \gamma^2 S_{FBB} \equiv g_m^2 S_{FB} + g_{mb}^2 S_{FBB}, \tag{285}$$

since the coupling coefficient $\gamma=\partial V_T/\partial V_B=g_{mb}/g_m$ also is the ratio between the front gate transconductance g_m and back gate (or body) transconductance g_{mb} in the MOS transistor (see page 369 in [169], for example). The latter result can be derived easily from the general eq. (1) for coupling of two independent voltage noise sources into fluctuation of one current.

It is worth mentioning that a guide line [88] can be provided by the substitution of the oxide trap number fluctuation model from eq. (101) in eq. (284), which gives

$$S_{VG}^{FD\,SOI} = S_{FB}\left(1+\gamma^2\frac{S_{FBB}}{S_{FB}}\right) \approx S_{VG}^{Bulk\,MOS}\left[1+\frac{\lambda_{BOX}N_{t,BOX}}{\lambda N_t\left(1+\frac{C_{BOX}}{C_{ox}}+\frac{C_{BOX}}{C_d}\right)^2}\right]. \tag{286}$$

Assuming that the same type of oxides are used for front and back gates, $\lambda_{BOX}=\lambda$ and $N_{t,BOX}=N_t$, the last equation suggests that the 1/f noise can increase twice in fully depleted SOI MOS transistors with thick back gate dielectrics and ultra-thin semiconductor films, when only the front gate is used as input. On the other hand, there should be a reduction of noise in double gate and Fin FETs, since the two gates are tied together, the gate area is doubled, thus S_{FB} is a half, and the denominator in the last expression is always larger than four when $C_{BOX}=C_{ox}$. Evidently, the back-front coupling in advanced structures with increased electrostatic control should not increase the noise, provided that the quality of gate dielectrics and gate area are not reduced. Such reduction of the noise is demonstrated for Fin FET [94] and cylindrical gate-all-around MOS transistor [137], reporting low values for oxide trap density $1.5\times10^{17}eV^{-1}cm^{-3}$ and less than $0.5\times10^{17}eV^{-1}cm^{-3}$, respectively.

Also, the analysis of the 1/f noise from the back gate in FD SOI MOS transistors indicates one more time a convergence between noise models for different structures in a close relation with the general principle of noise coupling, expressed by eqs. (1), (2), (3), (4) and (5). We demonstrate below that the capacitive coupling in the semiconductor body can reduce the noise in multiple gate MOS transistors, by using eq. (4) with the DC current cancelled.

VI.1.4. <u>Capacitive coupling in the semiconductor body</u>

Bulk MOS transistor

Consider first the obvious bulk MOS transistor with gate capacitance WLC_{ox} and small depletion capacitance WLC_d, as depicted on the top in Figure 47. As a statistical variance, the oxide charge fluctuation $S_Q=WLS_{Qo}$ is proportional to the gate area WL, where S_{Qo} is the charge fluctuation per unit area, and it is a constant at an assumption for uniform spatial distribution of oxide traps. At given biasing condition, the transconductance eq. (4) is $g=g_m=g_oW/L$, where g_o is the transconductance of square-shaped gate. The capacitance in eq. (4) is the capacitance seen by the charges at semiconductor-dielectric interface and $C=WL(C_{ox}+C_d)\approx WLC_{ox}$. The substitution in eq. (4) gives the equation for the drain current noise $S_{I_D}|_{N=1,\,one\,gate}$, given by

$$S_{I_D}\Big|_{N=1,\,one\,gate} = g_m^2\frac{S_Q}{C^2} = g_o^2\left(\frac{W}{L}\right)^2\frac{WLS_{Qo}}{[WL(C_{ox}+C_d)]^2} \approx \frac{g_m^2 S_{Qo}}{WLC_{ox}^2} \tag{287}$$

Two-gate MOS transistor

Consider now the two-gate MOS transistor with thin semiconductor film, such as FD SOI MOS transistors in a configuration of two gate "Fin FET" [261, 262 263], which is depicted in the middle of Figure 47. The channel width is W/2, since there are two identical conduction channels (top and bottom) and the gates are tied together, and from AC point of view, the gates are grounded. The capacitance seen at semiconductor-dielectric interface is the oxide capacitance plus the capacitance of the path in the bulk, which is series connection of the depletion capacitance and the capacitance of the other gate dielectric. Therefore, by summing the noise of the two channels, one gets

$$S_{I_D}\Big|_{N=2,\text{ two gates}} = 2\left\{ g_o^2\left(\frac{W}{2L}\right)^2 \frac{\frac{W}{2}LS_{Qo}}{\left[\frac{W}{2}L\left(C_{ox}+\frac{C_dC_{ox}}{C_d+C_{ox}}\right)\right]^2}\right\} \approx \frac{g_m^2 S_{Qo}}{WLC_{ox}^2}\frac{1}{\left(1+\frac{C_d}{C_d+C_{ox}}\right)^2} \tag{288}$$

Four-gate structure

Considering the four-gate structure in the bottom of Figure 47, one sees three paths of series connections of C_d and C_{ox}. Following the procedure for the two-gate MOS transistors, one gets for the four-gate MOS transistor that

$$S_{I_D}\Big|_{N=4,\text{ four gates}} = 4\left\{ g_o^2\left(\frac{W}{4L}\right)^2 \frac{\frac{W}{4}LS_{Qo}}{\left[\frac{W}{4}L\left(C_{ox}+3\frac{C_dC_{ox}}{C_d+C_{ox}}\right)\right]^2}\right\} \approx \frac{g_m^2 S_{Qo}}{WLC_{ox}^2}\frac{1}{\left(1+3\frac{C_d}{C_d+C_{ox}}\right)^2} \tag{289}$$

By comparing eqs. (287), (288) and (289), we write

$$S_{I_D}\Big|_{N\text{ gates}} \approx \frac{S_{I_D}\Big|_{\text{one gate}}}{\left[1+(N-1)\frac{C_d}{C_d+C_{ox}}\right]^2} = \frac{S_{I_D}\Big|_{\text{one gate}}}{\left(1+\gamma\frac{C_d}{C_d+C_{ox}}\right)^2}\ , \tag{290}$$

where N=1, 2, 3 or 4 is the number of gates. The last equation shows that the coupling under the channel reduces the noise. The parameter $\gamma\sim(N-1)$ depends of the channel geometry, in particular, how the channel is surrounded with conductive layers and what is the coupling $C_d/(C_d+C_{ox})$ between the channel and these layers. Nowadays, $C_d/(C_d+C_{ox})\sim0.3$, and it is expected that it will increase to 0.5 in ultra-thin body SOI, Fin FETs and nanowire-based MOS structures. Provided that three-four gate structures will be employed in these transistors, then the noise in them will be $(1+3\times0.5)^2\sim6$ times less at the same trap density in the gate insulators, as compared to the prediction of the model for one-gate transistors, which is currently only available. The reduction could be even larger in cylindrical-channel structures, as observed in [137], since $\gamma>3$ is possible for this geometry, along with $C_d>C_{ox}$, as calculated in [264] for nanowire based gate-all-around transistors.

Note that eq. (290) is qualitative. Unfortunately, the impact of the coupling on noise in advanced structures is not addressed at present accordingly, except for the excess noise in SOI MOS transistors due to filtered shot noise,

which was discussed in the beginning of this section. Due to surface channel conduction at dielectric interface, the analyses of the noise in MOS Fin FETs operating in inversion mode [94, 137] is carried out in terms of number fluctuation model for single gate MOS transistors, while for depletion mode Fin FETs the Hooge model of eq. (6) is usually taken, when the conduction is in the bulk [246].

VI.1.5. Two-MOS two-junction gate transistor

It is interesting to observe the crossover between surface and bulk noise reported in [246] for the four-gate structure with two MOS and two junction gates. The structure is depletion mode field effect transistor with doped n-type channel and n^+ drain and source, and it is shown in Figure 48a. At positive bias voltage applied to the top MOS gate G1 and negative bias voltage applied to junction gates JG1 and JG2, the carriers are accumulated under the MOS gate, as shown in the upper-right insert of Figure 48b, resulting in surface channel. At this condition, the noise is high, as shown with dashed line in the figure. By reducing the biasing of G1, the surface becomes depleted, and channel moves into the bulk of the silicon bar, as shown in the bottom-left insert of Figure 48b. Consequently, the noise also decreases, as shown with dotted line in the figure. Further change in the biasing of the MOS gates to negative voltages causes inversion at channel surface, and the inversion layer screens the charge fluctuation at the gate dielectric. Consequently, the surface noise is effectively eliminated, and the noise in the drain current is reduced to the value of the bulk noise, as shown with solid line in Figure 48b. The bulk noise shows nearly $1/I_D$ dependence, which is typical for intrinsic noise, and therefore the authors of [246] analyzed the data in terms of Hooge eq. (6), estimating for JFET operation mode that the Hooge parameter is $\alpha_H \sim 2\times10^{-5}$ for the bulk noise, when the surface is inverted, with a slight increase to $\alpha_H \sim 4\times10^{-5}$, when the surface is depleted. These values are obtained from data corresponding to $I_D=6\mu A$ with estimated number of carriers $n=4.5\times10^5$ in the transistor channel. For comparison, in MOS operation mode, the effective value for the Hooge parameter is one order of magnitude higher, $\alpha_H \sim 2\times10^{-4}$, and also, the normalized noise does not follow very close the $1/I_D$ dependence, having a plateau at low currents, which is typical for coupled number fluctuation due to charge trapping in gate oxides. In addition, in either mode, Lorentzian noise is reported at biasing around the threshold. Therefore, the Hooge equation was used in a modified form for the noise in a frequency band from $f_{min}=1Hz$ to $f_{max}=500Hz$, given by

$$\frac{i_D^2}{I_D^2} = \int_{f_{min}}^{f_{max}} \frac{\alpha_H}{n}\frac{df}{f} = \frac{\alpha_H}{n}\ln\left(\frac{f_{max}}{f_{min}}\right). \tag{291}$$

Thus, along with the clearly demonstrated crossover between surface and bulk noise in [246], one also observes that there are additional effects at crossover points, and the application of simple models, such as that in eq. (11) may meet with difficulties in the practice of characterization of advanced structures. For instance, in later publications [265, 266] there were found operating regions in this transistor where the cutoff frequency of the Lorentzian spectra, originated by volume deep centers, changed with gate voltage. This result contradicts the classical assumption that no variation of the cutoff frequency occurs in Lorentzians related to a given trap level in the depletion layer of an MOS transistor. This is explained mainly by the fact that a bipolar structure (a p-n junction) is added to a unipolar device (an MOS transistor).

VI.2. Nanotubes and nanowires – 1D seems too noisy

The difficulties in the practice of characterization of advanced structures are even more pronounced, when the semiconductor nanowires (SNW) and carbon nanotubes (CNT) are used as the basis for the device structure. The

initial study on noise [267] in two terminal CNT structures (thick, thin films and single CNT between electrodes) indicated that the these devices behave as resistors with a resistance R and the 1/f noise is anomalously large, as compared to other resistors, such as metal film and carbon composite resistors. Otherwise, carbon nanotube resistor behaves "normally". That is, the conductance 1/R is proportional to the number of parallel CNTs, the white noise is the thermal noise with voltage spectrum density 4kTR, and the magnitude of 1/f noise is proportional to the square of the DC bias and inversely proportional to the resistance, which suggests that the 1/f noise is intrinsic noise of the current transport, and according to Hooge eq.(6), the 1/f noise is given by

$$S_{norm} = \frac{S_I}{{I_{DC}}^2} = \frac{S_V}{{V_{DC}}^2} = \frac{S_R}{R^2} = \frac{\alpha_H}{nf} = \frac{A}{f} \equiv \frac{K_F}{f} = \frac{BR}{f}, \tag{292}$$

where the parameter $A \equiv K_F = \alpha_H/n = BR$ is proportional to a nearly constant parameter $B \sim 10^{-11}\Omega^{-1}$. Note that the parameter A is the same as the SPICE parameter K_F in eq. (15). The nearly constant value B=A/R is empirically observed in [267], the proportionality between A and R is speculatively explained with a general statement that $R \propto 1/n$, since n is total number of carriers, while R also depends on the ratio of the cross-section to length of the conductive bar, and the estimated value for $\alpha_H \approx 0.2$ is large and in quantitative disagreement with values for conductive materials, in which $\alpha_H \approx 0.002$. Consequently, S_{norm} in the CNT resistors is 4 to 8 orders of magnitude larger, as compared to the values for carbon composite and metal film resistors, respectively. All these quantitative details for the noise in CNT are stated in [267]. Nevertheless, the ratio

$$\frac{A}{R} = \frac{fS_{norm}}{R} \equiv \frac{K_F}{R} \sim 10^{-11}\Omega^{-1} \equiv FOM_{CNT} \tag{293}$$

is often taken as a reference figure of merit for comparisons when the 1/f noise in CNT and SNW structures is investigated, although the physical significance of this ratio is not clear at present.

Further insight on the low-frequency noise in two terminal CNT devices is provided in [268], by analyzing the shape of the spectra and temperature dependences. The temperature measurement confirmed that $FOM_{CNT}=A/R$ ranges between $(1.3\text{-}7.6)\times10^{-11}\Omega^{-1}$ for the 1/f noise over a range from cryogenic temperatures of 77K to room temperature of 300K. Also Lorentzian components and change of the frequency slope is observed in these measurements. From the evolution of the corner frequency in the Lorentzian spectra with the temperature, several activation energies were found at 0.080eV, (0.2–0.21)eV, (0.27–0.31)eV, 0.42eV and 0.51eV, or roughly, at every 100meV, and it was suggested the energy difference between subbands in CNT may cause these activation energies. The crosses in Figure 49 illustrate the data for the activation energies, and the lines without symbols illustrate the density of states (DOS) in CNTs with different diameters [264]. The variations of the frequency slope of the 1/f noise was also analyzed in terms of Dutta-Horn equation, see eqs. (167) and (169), and it was observed [268] that the density of traps, if such are assumed as origin of the 1/f noise, decreases from low to high activation energy, similar to what one observes for a tail trap distribution in forbidden energy band gap of semiconductors. The lines with symbols in Figure 49 illustrate the distribution of trap density D(E). Assuming hole conduction in CNT (we will show shortly results that confirm p-type field-effect behavior of CNT), it was suggested that the distribution of hole activation energy could be associated with a tail defect density distribution referenced to the top of the nanotube valence band, similarly to the defect distributions observed on silicon and other semiconductor surfaces. The nanotubes used in the experiments [268] have a diameter of about 1 nm and thus an energy band-gap of about 1 eV [269]. A complementary experiment in

[268], however, showed that there is no measurable difference between samples with CNT placed on SiO_2 and other, suspended samples, in which the oxide was removed by etching, ruling out the oxide-CNT interface as significant contributor to the noise in CNT devices. The FOM_{CNT} was $A/R=4.3\times10^{-11}\Omega^{-1}$ for the suspended sample, and the activation energies were virtually at the same energy positions. The density of states in CNT, as illustrated in Figure 49, is also non-uniformly distributed. Therefore, the approximation of valence band edge with a step function is rough, and transitions between subbands in CNT is possible [270]. These details are not elaborated for the noise in CNT. The authors of [268] concluded that the interpretation of excess noise is complicated by the fact that carbon nanotubes operate as mesoscopic quantum devices, where traditional methods of noise measurements and interpretation may not readily apply. For example, the four-point probe measurement method to eliminate contact noise from the measurements will not be conclusive for a nanotube device. The source and drain contact reservoirs are an integral part of a nanotube device and the addition of two more contacts may completely change the operation of the device itself.

From the above introduction to the noise in CNT one can readily see that there are difficulties to understand the noise in nanowire and nanotube devices. One is the constant value $FOM_{CNT}=A/R$, which does not scale with area of the device, second is the large value for normalized noise S_I/I_{DC}. There are also other problems, such as nanowire and nanotube conduction network, weak contact to probing terminals, dependences on environment, giant RTS and temperature dependences. In fact, the noise depends on everything in these devices and no dominant factor for the low-frequency noise can be identified in nanowire and nanotube devices. Also, the extrapolation of mesoscopic noise models for these devices is uncertain. Some published results related to these difficulties are discussed below.

Mesoscopic noise models. Contact effects.

The extrapolation of mesoscopic noise models to individual CNT or SNW assumes superposition of noise sources

$$S_{tot} = \sum S_{NW/NT} + \sum S_C \ , \tag{294}$$

where S_{tot} is the measured total noise, $S_{NW/NT}$ are the noise contributions from individual nanowire or nanotube and S_C are the noise contributions from contacts to the nanowire or nanotube. In the case of single CNT devices, the summation symbols are omitted. In the case of film-like devices, the individual contributions are lumped in one noise source. In intermediate cases, when the network of nanowires is known, coupling coefficients, $K_{NW/NT}$ and K_C, are suggested for the contributions from individual noise sources to the total noise, the values of the coupling coefficients are determined and eq. (294) is rewritten in terms of normalized noise currents as

$$\frac{S_{tot}}{I_{tot}^2} = \sum\left(\frac{I_i^2}{I_{tot}^2} K_{NW/NT} \frac{S_{NW/NT}}{I_i^2}\right) + \sum\left(\frac{I_i^2}{I_{tot}^2} K_C \frac{S_C}{I_i^2}\right) , \tag{295}$$

where I_i are branch currents flowing through individual CNTs or SNWs and contacts, and I_{tot} is the total current flowing through the two-terminal sample.

The determination of the coupling coefficients is conditional and depends on the structure and assumptions for this structure. For example, SEM (scanning electron microscope) images of devices with silicon nanowires bridging between electrode bars lead to assumptions for a network of parallel connected nanowires, and the coupling coefficients are evaluated from measurements of DC resistance and diameters and lengths of nanowires

observed in SEM images [271]. Having variation between sizes and number of nanowires in the samples and unknown resistances of the nanowires and contacts, one can see that there were many difficulties to resolve the network, and additional assumptions are taken in [271] that the carrier mobility μ and specific resistivity ρ in the semiconductor nanowires are constants and are the same as in bulk silicon, the contact resistances scale inversely with nanowire cross-sectional area, estimating a constant prefactor in this dependence $\rho_c \sim 1.7\times10^{-5}$ Ωcm². All these assumption are valid for mesoscopic devices, but it is difficult to prove for nanowire networks, although the worst case estimate for depletion due to surface charge trap with density $\sim 2\times10^{12}$ cm^{-2} indicated that the difference between physical and electrical cross-sectional areas can be neglected for the nanowires with diameters ~100nm in [271]. In this way, the resistance of individual nanowire branch (i) becomes

$$R_i = R_{NW,i} + R_{C,i} \ , \tag{296}$$

with $R_{NW,i}=\rho L_i/(\pi r_i^2)$ being the nanowire resistance and $R_{C,i}=\rho_c/(\pi r_i^2)$ being the contact resistance, where L_i is the length of the nanowire and r_i is the radius of the nanowire. Then, the spectral densities S_i of the noise currents associated with the individual nanowire branch are

$$S_i = S_{NW,i}\left(\frac{R_{NW,i}}{R_i}\right)^2 + S_{C,i}\left(\frac{R_{C,i}}{R_i}\right)^2 \ , \tag{297}$$

where $S_{NW,i}$ are the spectral densities of noise currents from nanowires and $S_{C,i}$ are the spectral densities of noise currents from contacts.

Next, one takes the simplest SPICE model for 1/f noise from eq. (15) , so that

$$S_i = \frac{K_{F,i}}{f} I_i^2 ; S_{NW,i} = \frac{K_{FNW,i}}{f} I_i^2 ; S_{C,i} = \frac{K_{FC,i}}{f} I_i^2 \ , \tag{298}$$

where I_i is the DC current flowing in the nanowire branch (i) and the SPICE parameters $K_{F,i}$, $K_{FNW,i}$ and $K_{FC,i}$ are associated with the branch, nanowire and contacts, respectively. Combining all parallel branches, one gets for the total normalized noise in eq. (295) that

$$\frac{S_{tot}}{I_{tot}^2} f = K_{F,tot} = \sum_i K_{FNW,i}\left(\frac{I_i}{I_{tot}}\right)^2\left(\frac{R_{NW,i}}{R_i}\right)^2 + \sum_i K_{FC,i}\left(\frac{I_i}{I_{tot}}\right)^2\left(\frac{R_{C,i}}{R_i}\right)^2 \tag{299}$$

In the last equation, S_{tot} and I_{tot} are the quantities, which can be obtained during noise measurements of a sample, and the SPICE parameter for 1/f noise of the sample can be determined as $K_{F,tot}=fS_{tot}/(I_{tot})^2$. The remaining quantities require additional assumptions and measurements. One obvious and useful for the practice note is that in resistor networks, the individual noise sources have “attenuated” contribution at sample terminals, since the coupling coefficients $K_{NW}=(R_{NW}/R_i)^2$ and $K_C=(R_C/R_i)^2$ are less than one.

Unfortunately, even when the branch resistances R_i and currents I_i, and the resistances of the nanowires and contacts are known, and even for single nanowire (i≡1), eq. (299) does not have unique solution for the two SPICE parameters K_{FNW},i and K_{FC},i of the nanowire and contact to it, because the equation is only one. The equation becomes highly undetermined when the number of parallel branches in the sample increases. Therefore, in addition to the assumptions for scalable and reproducible resistances, more assumptions for scalable and reproducible noise between different nanowires in different samples is required and made in [271], using again

extrapolation of a mesoscopic noise model for the intrinsic (Hooge) 1/f noise from eq. (6). So, having j=1,2… samples with different number i(j) of nanowires in different samples, and assuming single-valued parameters for specific bulk resistance ρ [Ωcm], mobility μ [cm²/Vs] and Hooge parameter α_H in nanowires, as well as characteristic resistance ρ_c [Ωcm²] and unit-area noise α_c [cm²] for the contact to the nanowires, one obtains two systems of equations – one for resistances and second for noise.

The system of equations for resistances is in the form of

$$\left[\frac{1}{R_{tot}}\right]_j = \left[\sum_{i(j)} \frac{1}{R_{NW,i} + R_{C,i}}\right]_j = \left[\sum_{i(j)} \frac{\pi r_i^2}{\rho L_i + \rho_c}\right]_j, \tag{300}$$

in which the resistances of the samples, $R_{tot}|_{j=1,2...}$ are electrically measurable, while the number i(j=1,2…), the lengths $L_i|_{j=1,2...}$ and the radiuses $r_i|_{j=1,2...}$ of the nanowires in these samples has to be obtained from microscope images, e.g. using SEM, as in [271]. Ideally, two samples are needed to determine the specific bulk resistance ρ and characteristic resistance ρ_c, but inaccuracy and non-repeatability between samples, as well as the cylindrical approximation and non-linear nature of the equation, require more samples and a use of guided optimization-discrimination procedure, in order to obtain reliable values for ρ and ρ_c in each sample. Such procedure was carried out in [271], and one can see that it was not a simple task, as well as in fig. 4 in this publication, the accuracy and convergence between measured and calculated resistances is far from ideal and repeatable between different samples. Also, having aspect ratio L_i/r_i~8μm/50nm=160 for typical nanowire in these samples, it is somehow unrealistic the have ratio R_{NW}/R_C~2 by assumption of nearly uniform doping and mobility in the sample, unless there is a really difficult processing problems to make electrical contact to electrode sidewall at the impinging end of the nanowire, as stated in [271].

Nevertheless, once having the specific bulk resistance ρ in nanowires and characteristic resistance ρ_c for the contact, one can estimate the mobility μ and carrier concentration n' in nanowires from graphs for bulk silicon available in many textbooks for semiconductors and write the second system of equations for the noise. For the purpose, at given voltage bias V across the sample, and since the number of carriers in the nanowire is $n=n'L_i\pi r_i^2$, then the different quantities in eq. (299) are expressed as

$$\frac{R_{NW,i}}{R_i} = \frac{\rho L_i}{\rho L_i + \rho_c}; \quad \frac{R_{C,i}}{R_i} = \frac{\rho_c}{\rho L_i + \rho_c}, \tag{301}$$

the ratio between individual nanowire branch current I_i and total current I_{tot} in the sample is

$$\frac{I_i}{I_{tot}} = \frac{\frac{V}{R_i}}{V\sum \frac{1}{R_i}} = \frac{\frac{r_i^2}{\rho L_i + \rho_c}}{\sum \frac{r_i^2}{\rho L_i + \rho_c}}, \tag{302}$$

and from Hooge 1/f noise model in eq. (6) and SPICE 1/f model in eq.(15)

$$K_{FNW,i} = \frac{\alpha_H}{n} = \frac{\alpha_H}{\pi r_i^2 L_i n'}; \quad K_{FC,i} = \frac{\alpha_c}{\pi r_i^2}. \tag{303}$$

Thus, by substituting in eq. (299), the second system of equations for the noise is

$$\left[\frac{S_{tot}}{I_{tot}^2}f\right]_j=\left[K_{F,tot}\right]_j=\left[\frac{1}{\pi}\sum_{i(j)}\left\{\frac{r_i^2}{\left(\sum\frac{r_i^2}{\rho L_i+\rho_c}\right)^2}\left[\frac{\frac{L_i}{n'}\rho^2\alpha_H+\rho_c^2\alpha_c}{(\rho L_i+\rho_c)^4}\right]\right\}\right]_j \tag{304}$$

Similarly to eq. (300) for the resistances, the system of equations in eq. (304) has two unknowns, Hooge parameter α_H in nanowires and characteristic unit-area noise α_c [cm²] for the contact to the nanowires, and ideally two samples are needed to determine α_H and α_c. And also similarly, in practice one needs to carry out a guided optimization-discrimination procedure over several samples, in order to obtain reasonable values for α_H and α_c in each sample. Note that even after careful characterization in [271], the values of the parameters related to the noise in nanowires and contacts scatter over several decades in fig.6 in this publication, while the values for the resistance scatter within less than a half decade in worse case in fig.4 for the same samples. Evidently, the results in [271] symptomatically imply that the issues in nanowire and nanotube devices are related mostly to the access to these devices. If there is a contact problem for the DC performance of these devices, increasing the device resistance with 50%, then this problem actually dictates the low-frequency noise, pushing the noise 1-3 decades above the noise levels in the nano-conductor. The solutions of the contact problems were also recognized in ITRS predictions [3] as enabling factor to have access to advanced nano-devices.

The problem with the contact is even more pronounced in single carbon nanotube devices. In [272], multiwall carbon nanotubes (MWCNT) were placed to cross the gap between gold contact pads, using atomic force microscope (AFM), and the resistance and low-frequency noise of these samples were measured at different low bias currents (I_{DC}) from room temperature to cryogenic temperatures 77K and 4.2 K, down to 1.5K. It is observed in [272] that the low-frequency noise is 1/f at room temperature and at 77K, but at 4.2K and 1.5K, the noise spectra showed large Lorentzian components, as illustrated in the left-hand plot of Figure 50. Interesting observations for the behavior of the samples are made in [272] at 4.2K and 1.5K, when changing the direction of the bias current, as illustrated in the right-hand plots of Figure 50. The top plot shows that the time constants of the Lorentzian spectra are bias dependent, they are different at different bias polarities, but the time constants do not depend on the temperature. The middle plot shows that the prefactor $S_0/\tau=4(\Delta I/I_{DC})^2\tau/(\tau_1+\tau_2)$ also varies with bias, while from eq. (179) one expects $(\Delta I/I_{DC})$=constant for a particular trap and $\tau/(\tau_1+\tau_2)$ =F(1-F)=constant in conductive materials, in which the Fermi level, and thus the trap occupancy, are weak function of the bias. In the bottom plot, the contact resistance is larger than the resistance of the nanotube. The differential resistance is bias dependent, nearly $\propto 1/I_{DC}$ for $|I_{DC}|<1\mu A$, and also not fully symmetrical in respect to the direction of the bias current. In addition, the resistance is proportional to ~7mV/I_{DC}, rather than to 0.13mV/I_{DC}, which one would expect from the ratio φ_t/I_{DC} with thermal voltage $\varphi_t\approx 0.13$mV at temperature T=1.5K. This indicates tunneling junctions at the contacts, accompanied with Coulomb blockade, charge trapping and strong bias dependent Random Telegraph Signal (RTS) noise, which resulted in Lorentzian noise spectra at temperature 4.2K and below.

Provided that 4.2K is very low temperature, then Shockley–Read–Hall (SRH) statistics for the process of trapping and de-trapping is not realistic, because the thermal velocity v_{th} is low in eq. (63) and the experiments in

[272] did not show temperature dependence between 1.5K and 4.2K. A variation of the prefactor S_0/τ with the bias is not expected from SRH statistics, since the Fermi level is almost independent of bias in conductive materials and the trap occupancy, c.f. F(1-F) in eq. (179), should not vary with bias in these materials, unless there is a junction interface and potential bending in it. Therefore, RTS is explained in [272] with trapping and de-trapping in terms of tunneling to charge trap at the contact interface between nanotube and metal. So, instead of using of SRH relations from eq. (63), the capture and emission time constants are expressed in terms of tunneling, given by

$$\tau = \frac{1}{\omega_o}\exp\left[\frac{d}{\lambda}(1-\gamma V)\right] = \tau_o \exp\left(-\frac{I_{DC}}{I_o}\right). \tag{305}$$

Here, ω_o is attempt frequency, d is effective tunneling distance, λ is tunneling attenuation distance, γ is parameter related to the shape of tunneling barrier, $\tau_o=\exp(d/\lambda)/\omega_o$ is tunneling time constant at equilibrium (no bias), and the characteristic tunneling current $I_o=\lambda/(d\gamma R_{eff})$ is a fitting parameter that accounts for the effective resistance $R_{eff}=V/I_{DC}$ of the region, where the tunneling occurs. While the slopes for τ as function of the bias current in the right-hand plot on top of Figure 50 confirm the exponential dependence of eq. (305), an interesting observation in [272] is that the fitting parameters τ_o and I_o in eq. (305) are different for different time constants and they are not the same for capture and emission time constants τ_1 and τ_2, since the prefactor $S_0/\tau=4(\Delta I/I_{DC})^2\tau/(\tau_1+\tau_2)$ also varies with the bias, as shown in the right-hand plot of Figure 50 in the middle. This implies that tunneling distances, the attenuation distances and the barrier shapes vary, which is possible by having junctions at the weak contacts between nanotube and metal pads.

Nanotube FETs

The low-frequency noise in field-effect transistor configuration of nanotubes was recently also addressed. In these devices, one or several nanotubes, or thin film of random nanotube network bridges between metal contacts, and the gate is usually a conductive substrate, on the top of which the gate oxide and the metal pads are placed. These devices appear to be very noisy, affecting even DC measurements, as illustrated with several "I–V" curves in Figure 51.

In Figure 51(a) and (b), transfer I–V characteristics of field-effect transistors (FETs) based on single CNT measured at room temperature are shown. The values for the currents scatter between 10% and 20% around a trend. The trend is similar to the I–V curve of pMOS transistors. Since the scattering is large in Figure 51(a), the current and the threshold voltage (cross point of two lines) are analyzed in terms of stochastic resonance in [273], demonstrating that CNT FET can be used as a detector for signals below threshold voltage. Figure 51(b) illustrates that the current and its scattering are larger in ambient atmosphere, and they are reduced in vacuum [274]. Figure 51(c) depicts the cracked "I–V" curves of single CNT FET at cryogenic temperatures [171], owing to giant and bias dependent RTS noise with amplitudes 30% to 60% of "DC" current. Note that the currents and the transitions between the segments in the plot are different at opposite directions of the current flow, which is similar to the observations for multiwall CNT devices, discussed just above. On the other hand, in contrast to single CNT devices, the I–V characteristics in Figure 51(d) are smooth when the FET is based on a random network of single-wall CNTs [275]. A hysteresis is evident when the device was operating in air, using the silicon wafer as solid-state gate. Interestingly, the hysteresis is reduced when a liquid solution is used to mediate between electrochemical gate comprised by a pair of reference (Ag/AgCl 4M KCl or a saturated calomel) and by

a working (Pt) electrodes. The improvement when using liquid gate was attributed in [275] to enhanced electrostatic control and suppression of charge trapping effects. It was observed that the threshold voltage of the liquid gate CNT FET is a function of pH and concentrations in the chemical solution. Referring to the discussion on the multiwall CNT given above, we note, however, that the problems in the single CNT FETs can be due to contacts, rather than due to traps around the CNT, since the contact of metal to a network of CNTs and addition of electrolyte at this contact, as it was in [275], would greatly improve the repeatability of the contact. Unfortunately, the noise from the contact was not addressed in [171, 273, 274, 275], perhaps, due to a lack of scalable model for noise from contacts when interfacing 1D to 3D current transports. Obviously following the style in the initial publication on noise for CNT films [267], the noise is first phenomenologically investigated, and then related to noise model for mesoscopic devices, e.g. for MOS transistor noise model, relying on the similarity that exists to some extend between CNT FET and MOS transistors. Note again in Figure 51 that all devices at all measurement conditions behave similarly to p-type MOS transistors – a fact, which is experimentally observed and reported many times in the literature, although not very well justified theoretically, since the band structure of semiconducting CNT is quite symmetrical above and below the band-gap – see fig. 3 in [270], for example.

To illustrate the situation when investigating the noise in CNT FET and MOS transistors, we present a typical outcome from noise experiments in Figure 52. The data are from [274] for measurement of single CNT FET in vacuum (18mTorr) at room temperature. It is stated in the publication that this CNT FET has 4μm gap between gold electrodes and it was made as described in [171]. That is, CNT diameter is d=1−3nm and the gate is silicon wafer with t_{ox}=500nm thermal SiO_2. Considering the information from the publication, the CNT FET has length L=4μm equal the gap between electrodes, and taking an average diameter d=2nm, the channel width of the CNT FET is W=πd≈6.3nm. Assume that there is no gap between wafer surface and CNT, and consider $d<<t_{ox}$, then the gate capacitance per unit area is C_{ox}=7nF/cm² for the SiO_2gate dielectric with thickness t_{ox}=500nm.

The experiments in [274] were carried out in linear mode of operation of CNT FET, at low drain bias voltage $|V_D|<(|V_G-V_T|-0.5V)$. Therefore, a rough estimation for the total number of carriers n can be made by assumption for uniform charge density in CNT, resulting in $n=WL|V_G-V_T|C_{ox}/q$, which is approximately 11 electron charges per one volt of gate overdrive voltage $|V_G-V_T|$.

Figure 52a presents results based on measurement, in which the gate bias was varied, while the low drain bias was constant $-V_D$=0.1V. For the set of biasing points $\{V_G, I_D\}$, both the input referred (gate) noise voltage S_{V_G} and output (drain) noise current S_{I_D} are reported in [274]. From these, we obtain the transconductance $g_m=(S_{I_D}/S_{V_G})^{0.5}$≈17nA/V (almost constant, as expected for linear mode of operation of FET transistors) and draw the evolution of the ratio I_D/g_m with bias current in the Figure 52a on top. For operation of FET in linear mode, we expect $I_D/g_m=|V_G-V_T|$ and $I_D\propto|V_G-V_T|$, and we observe linear dependences with slope 60MΩ for I_D/g_m and slope $40M\Omega=1/[(W/L)\mu C_{ox}V_D]$ for the relation between $|V_G-V_T|$ and I_D as function of the bias current I_D. From the latter dependence, $|V_G-V_T|$ vs. I_D, shown just under the plot for I_D/g_m, we have estimated mobility μ≈24000 cm²/Vs. Comparing to crystal semiconductors, the value for mobility is impressive, but it is somehow in the middle of the range 4000-120000 cm²/Vs reported in [270], thus it is reasonable.

Having the above information for the sample handy, we pursue analysis of noise in terms of mesoscopic noise models for MOS transistors.

For the number fluctuation model with correlated mobility fluctuation, as shown in the middle of Figure 52a, we plot the square root $\sqrt{S_{V_G}}$ of the power spectrum density of the input (gate) noise voltage S_{V_G} at 1Hz, referring the reported data for S_{V_G} from 40Hz to 1Hz, by $S_{V_G}(1\text{Hz})=40\text{Hz}\times S_{V_G}(40\text{Hz})$. The constant in the linear fit to $\sqrt{S_{V_G}}$ yields flat-band noise voltage, and from eqs.(101), (109) for noise from tunneling and trapping in gate oxide, we get

$$\begin{aligned} S_{FB}(1\text{Hz}) &= 36\times10^{-6}\,\text{V}^2/\text{Hz} \\ &= \left(\frac{q}{C_{ox}}\right)^2 \frac{kT\lambda N_t}{1\text{Hz}\times WL} \\ &= N_t[\text{eV}^{-1}\text{cm}^{-3}]\times5.4\times10^{-22}\,\text{eVcm}^3\text{V}^2/\text{Hz} \end{aligned} \quad , \tag{306}$$

when substituting the values for the parameters of sample and using kT=0.026eV for thermal energy at room temperature and tunneling attenuation distance λ=0.1nm. Therefore, we obtain $N_t=(36\times10^{-6}\text{V}^2/\text{Hz})/(5.4\times10^{-22}\text{eVcm}^3\text{V}^2/\text{Hz})=6.7\times10^{16}\text{eV}^{-1}\text{cm}^{-3}$, which is a reasonable value for the trap density is SiO_2, see Figure 42, despite that S_{FB} is high.

To evaluate the parameters for the correlated mobility fluctuation, see the discussion after eq. (121), we use the bias dependent term in the linear fit of $\sqrt{S_{V_G}}$, which is I_D0.4MΩ. Since $|V_G-V_T|=I_D$40MΩ, as seen from the transfer $|V_G-V_T|$ vs. I_D curve in Figure 52a, then

$$\frac{\sqrt{S_{V_G}}}{\sqrt{S_{FB}}} = 1+\left|V_G-V_T\right|\frac{0.4\text{M}\Omega/\sqrt{\text{Hz}}/40\text{M}\Omega}{6\text{mV}/\sqrt{\text{Hz}}} = 1+\theta\left|V_G-V_T\right| \ , \tag{307}$$

or $\theta=0.01/0.006=1.67\text{V}^{-1}$, which is in the range observed for MOS transistors (see Figure 37). There are two definitions for scattering parameter α_s, as shown in eq. (243). Using one of them, given by eqs.(127) and (137), then $\alpha_s=q\theta/(\mu C_{ox})=1.6\times10^{-15}$Vs. The other definition gives $\alpha_s=\theta/(\mu C_{ox})=10^4$Vs/C. Both values are within ranges deduced from Si MOS transistors – see eq.(243). Thus, both by N_t and α_s, the number fluctuation can be justified for CNT FET.

The 1/f noise in CNT FET can be justified also in terms of intrinsic (Hooge) noise. We calculate the SPICE parameter $K_F=f\times S_{I_D}/I_D^2$ and plot it in the bottom of Figure 52a. Obviously, from eqs.(6) and (15), $K_F=\alpha_H/n$ decreases, when the total number of charges $n=11\times|V_G-V_T|$ increases with gate bias, as mentioned above. The calculated values for the Hooge parameter α_H are shown above the plot for K_F in Figure 52a. The values scatter, but the average $\alpha_H\sim2\times10^{-3}$ is a reasonable value for conductors, and by this value, the 1/f noise in CNT FET can be also justified as mobility noise.

The above discussion implies that 1/f noise in CNT FET is easily explained in terms of downscaled mesoscopic models for Δn and Δμ fluctuation, when data as function of gate bias are used. However, in the published analyses, which are similar to the above, there are details which are neglected. We show these details in Figure 52b. In this figure, from top to bottom, although $V_D<<V_G$, the relation between gate bias and drain current is not linear, the relation between drain bias and drain current has a step at low voltage, the normalized noise in terms of K_F is function of drain bias at low drain bias levels, and K_F obtained from experiment with variable gate bias is different from K_F obtained from experiment with variation of drain bias, even for the bias point {$-V_G$=2V,

$-V_D$=0.1V}, which was common in the two experiments. For this bias point, even the DC currents were different, as shown with arrows in the figure. All these indicate that the access contact to CNT, trapping and barrier at it, might be significant noise sources, as deduced for CNT films, networks and multiwall CNT [268, 272]. To the best of our knowledge, there is no publication that explains these "small" details in the behavior and noise related to them in CNT FET. The explanations for the contact effects and functionalized surface of CNT by adsorption or by changing of chemical environment (air or other atmosphere [274], or pH of liquid [275, 276]) on noise in CNT FET are still qualitative.

Analysis of the ratio $K_F/R \equiv A/R$

We address now the consequences from the empirical observation made in [267] that ratio $K_F/R \equiv A/R$ is approximately constant in CNT devices – see again eqs. (15), (292) and (293), where $A \equiv K_F$ is the SPICE parameter as defined in eq. (11). The CNT devices are typically arranged in thin-film structures, as shown in Figure 53a, b and c. The CNT networks have conductive branches, which are shown with arrows in these figures. Since the transport in CNT is 1-D, then the conduction branches are independent each from other. Having on average L carbon nanotubes in each conduction branch and W conduction branches in the device, then we can represent the CNT percolation network by and idealized resistor network, as shown in Figure 53d. For simplicity, we will assume that the resistance R_O and the noise v_o^2 of each CNT have the same values, that the number L of serially connected CNTs in each conduction branch is the same in the network, that the number of identical parallel conduction branches is W, and that the noise v_c^2, which may originate to contact between nanotube and metal electrode, is the same for every single conduction branch. By these idealizations, when applying external bias voltage V (or current I), the DC and noise currents in each branch are $I_O=I/W=V/(LR_O)$ and $i_o^2=(v_c^2+Lv_o^2)/R_O^2=i^2/W$, respectively, and one can easily find the total resistance $R=V/I=LR_O/W$ and voltage noise $v^2=R^2i^2$ of the circuit for the case of current biasing I. In this way, the normalized noise of the CNT network is

$$S_{norm} = \frac{v^2}{V^2} = \frac{i^2}{I^2} = \frac{r^2}{R^2} = \ldots = \frac{1}{W}\frac{v_c^2}{V^2} + \frac{L}{W}\frac{v_o^2}{V^2} = \frac{R}{R_O}\left(\frac{v_c^2/L}{V^2} + \frac{v_o^2}{V^2}\right), \tag{308}$$

and in terms of the empirically observation in [267] that $K_F/R \equiv A/R$ is approximately constant, we get

$$f\frac{S_{norm}}{R} = \frac{K_F}{R} \propto \frac{1}{R_O}\left(\frac{v_c^2/L}{V^2} + \frac{v_o^2}{V^2}\right) \approx \text{constant}, \tag{309}$$

where we assume 1/f noise, and the parameters resistance R_O=constant of single CNT and number L=constant of nanotubes in a conductive branch are constants for given sample (and given gate bias, if the CNT device is a TFT transistor). Note that S_{norm}/R and K_F/R do not depend on the number W of parallel conductive branches.

It is evident from eq. (309) that the empirical observation in [267] requires two conditions for the noise sources in CNT networks. The first condition is that the noise from individual CNT has to scale with the bias, but not with the number L of serially connected CNTs. The second condition is that the noise from contacts also has to scale with bias and it has to decrease with the number L of serially connected CNTs.

The bias dependence is easily reproducible by mesoscopic models for noise. The dependence on the number of CNTs serially connected in conductive branch is, however, not. Let us take the first condition, for example, and try to analyze in terms of Hooge eq.(15) at given bias voltage, assuming number of carriers n_o in a single

nanotube.

For branch with one nanotube we have $v_{o1}^2/V^2=\alpha_H/n_o$. For a branch with L nanotubes, $Lv_{oL}^2/V^2=\alpha_H/Ln_o$, according to the equivalent circuit of the network. In contrast to the expectation from eq. (309) for $v_{oL}^2/V^2= v_{o1}^2/V^2$, we get $v_{oL}^2/V^2=\alpha_H/L^2n_o\neq\alpha_H/n_o=v_{o1}^2/V^2$. The condition will be satisfied, if one assumes that the total number Ln_o of carriers in the branch decreases when increasing the number L of serially connected CNTs, that is $n_o\propto 1/L^2$. We did not find a way or publication to justify physically such dependence for the number of carriers, although many publications use the empirical observation in [267] for comparisons. The issue is that $n_o\propto 1/L^2$ dependence cannot be derived from any mesoscopic model for noise.

Nevertheless, it seems that the noise in nanowire [271, 277, 278, 279, 280, 281] and CNT [268, 272, 274, 275, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292] devices scales according to the rules for mesoscopic devices. This is illustrated in Figure 54. The publications discussed above argue that the dominant sources of 1/f noise in CNT and nanowire devices are number fluctuation and from contacts. Therefore, we calculate the surface area and cross-section area of the CNT and nanowires, and plot versus these areas and versus data for MOS and bipolar transistors in Figure 54a and b, respectively. When the information for the size and number of CNTs in the devices was not stated, we assumed that CNT diameter is 3nm and the resistance due to one CNT is about 100kΩ. Interestingly, the data in Figure 54 show that the geometrical scaling rule for 1/f noise, the smaller is the area – the noisier the device is, also applies for nanowires and nanotube devices.

The comparison of surface area dependence of the 1/f noise in Figure 54a to data for MOS transistors implies that the noise in nanowire (NW) and carbon nanotube (CNT) devices can be explained in a manner similar to the models for the 1/f noise in MOS transistors, because the noise in NW and CNT devices is less than and in the range of the noise in MOS transistors, although the scattering is large (σ_{dB}=9.7dB). Thus, one can assume surface origin for the noise in NW and CNT devices.

The comparison of cross-section area dependence of the 1/f noise in Figure 54b to data for bipolar transistors implies that the contact noise in NW and CNT is less than the noise in bipolar transistors, the scattering of the data is less ($\sigma_{dB}=5.6_{dB}$), but the noise rapidly increases in single-wall CNT FETs with single or small count of CNTs. Thus, one can deduce a crossover between dominant noise sources, since the normalized noise (K_F) increases steeper than 1/area in small-area CNT devices, as compared to the lower noise in NW samples with several parallel nanowires. Qualitatively, the crossover is from bulk noise in NW devices, to surface noise in multiple CNT, toward injection noise (either tunneling or thermionic) in single CNT devices. The published data scatter over several decades, and a reliable estimate for the crossover points is not possible. Therefore, one can find a variety of models and explanations for the noise in semiconductor nanowire (NW) and carbon nanotube (CNT) devices, which causes difficulties when comparing devices from different publications.

Again looking at Figure 54, one can see that the CNT devices are noisy, but in relative units, not noisier than MOS and BJT, if the latter are scaled down to the sizes of carbon nanotubes. This demonstrates one more time that there is convergence of noise models and behaviors from very large down to very small devices. The issue is that the normalized 1/f noise (at 1Hz) in nano-devices is larger than $K_F>10^{-4}$. Therefore, these devices will be difficult to use deterministically, since for an application that requires 4-5 frequency decades, the peak-to-peak noise becomes more than 20%, according to

$$\frac{V_{pk-pk}}{V_{DC}} \approx 6 \sqrt{K_F \ln\left(\frac{f_{max}}{f_{min}}\right)}. \quad (310)$$

In three sentences, although the physical origin is not very well determined, the 1/f noise in semiconductor nanowire (NW) and carbon nanotube (CNT) devices scales according the rules for mesoscopic devices. Since the area (either surface, or cross-section of contacts) of NW and CNT is very small, then the 1/f noise in these devices is a limiting factor for the use in the practice. The single nanotube devices seem are not anymore deterministic, that is, they are behind the down-scaling barrier set by the 1/f noise.

VI.3. Between 3D and 1D – the graphene and transition metal dichalcogenide 2D transistors

The reduction of the mobility in ultra-thin silicon body SOI and FINFETs (transistors with 3D charge transport), and the difficulties in the mass-production of nanowire and nanotube transistors with 1D charge transport, brought the interest in exploring graphene and transition metal dichalcogenide transistors, which have semiconducting "body" of single to few atomic layers and 2D charge transport. The 2D transistors attempt to utilize the better electrostatic control in field-effect transistors with thinner body, the high intrinsic mobility of the graphene and the apparent advantage of atomic layer growth of metal dichalcogenides, along with the compatibility with the lithography for planar devices in the microelectronic manufacturing.

However, the properties of the 2D semiconducting layers deviate from the properties of well-understood crystalline layers in the 3D silicon transistors, inheriting also from the quantum effects in the 1D nanowires and nanotubes. Typical issues in the 2D transistors are the poor contact with the metal electrodes of the device terminals and the non-covalent (van der Waals) bonding between the atomic layers. The latter, basically, implies that the 2D semiconductor is a stack of several atomic layers with increased spacing and energy barriers between the atomic layers, but not a homogeneous layer as in the crystalline 3D semiconductors. Below, we illustrate the consequences for the low-frequency noise in 2D transistors with an example from [293] for a MoS_2 (a metal dichalcogenide) transistor.

Figure 55 (a) shows the barriers ϕ_v in the energy diagram and the spacing d_c between the MoS_2 monolayers in the spatial cross-section schematic diagram. The barriers and the spacing are due to van der Waals bonding between the MoS_2 monolayers, which is weaker than the covalent bonding in the MoS_2 monolayers and in the crystalline 3D semiconductors. According to these diagrams, the authors of [293] consider the following physics and relations. The noise is due to Δn fluctuation of the trapping in the gate oxide, combining several processes that affect the time constants of the trapping and the noise measurement. One process is the Shockley–Read–Hall (SRH) recombination at semiconductor-dielectric interfaces with time constant $\tau_o \propto 1/n$ inversely proportional to the carrier density $n \propto (V_{GS}-V_T)$ in the semiconducting layers and the gate overdrive voltage $(V_{GS}-V_T)$, thereof. A second process is the charge tunneling to/from traps at different distances in the gate dielectric, which randomizes τ_o in a range of larger values, resulting in band-limited 1/f noise. A third process is the additional increase of the time constant values for charges from semiconducting monolayers non-adjacent with the gate dielectric, owing to the energy barriers and spacing due to van der Waals bonding of the semiconductor monolayers. The fourth consideration is that 1/f noise is band limited and measurable only when the frequency band of the spectrometer (2Hz – 1000Hz) and the band limited 1/f noise overlap. This fourth consideration is essential for the explanation of the non-monotonic dependence of the normalized noise S_o as function of the gate bias V_{GS} shown in Figure 55 (b).

Figure 55 (b) shows the normalized noise S_o referred to 1Hz, S_o=average($f \times S_I(f)/I_{DS}^2$), averaged over logarithmically spaced frequencies f in the range from 2Hz to 1000Hz, vs. the gate bias voltage V_{GS}. From left, S_o reduces with the gate overdrive voltage (V_{GS}-V_T), because the time constant $\tau_o \propto 1/n$ of the (first) Shockley–Read–Hall (SRH) process is inversely proportional to the carrier density $n' \propto (V_{GS}-V_T)$ in the monolayer adjacent with the gate dielectric. The measured noise is 1/f, owing to the (second) process of the charge tunneling to/from traps at different distances in the gate dielectric, which randomizes τ_o in a range, resulting in band-limited 1/f noise. One can deduce mathematically the reduction of S_o and higher n' by using fraction of $\tau \propto 1/n'$ in the numerator of the integrand in eq. (73).

The (third) process of additional increase of the time constant values for charges from semiconducting monolayers non-adjacent with the gate dielectric brings the upper boundary $1/(2\pi\tau''_{min})$ of frequency range of the band-limited 1/f noise from the non-adjacent semiconducting monolayer below the lower boundary of 2Hz of the spectrometer (the fourth consideration above), when the carrier concentration $n'' \propto (V_{GS}-V_T)$ in the non-adjacent monolayer is low at low gate overdrive voltage (V_{GS}-V_T). Therefore, the noise associated with the non-adjacent semiconducting monolayer was not measured and missing in the left-hand side of the plot of S_o in Figure 55 (b). However, increasing the overdrive voltage (V_{GS}-V_T), the carrier concentration $n'' \propto (V_{GS}-V_T)$ in the non-adjacent monolayer increases, $\tau''_{min} \propto 1/n''$, the upper boundary $1/(2\pi\tau''_{min})$ of the frequency range of the band-limited 1/f noise increases, reaching the spectrometer range 2Hz-1000Hz at $V_{GS} \approx 5V$, and the overlap of the 1/f noise and spectrometer frequency ranges increases, resulting in increase of S_o to a peak value at V_{GS}=20V. At higher V_{GS}>20V, the overlap of the frequency ranges of the 1/f noise from the non-adjacent semiconducting monolayer and the spectrometer is full, but S_o reduces at increasing V_{GS}, because the time constants $\tau'' \propto 1/n''$ of the noise, owing to the (first) Shockley–Read–Hall (SRH) process (as above for the noise from the adjacent monolayer in the left-hand side of the plot of S_o in Figure 55 (b))

In summary, the low-frequency noise in 2D transistors follows the noise behavior in 3D transistors, but energy barriers and spatial spacing between monolayers bring the noise parts from different monolayers outside the ranges for noise spectrum measurement, which may cause apparently spurious non-monotonic data series for the noise levels, e.g., as function of bias, as shown in Figure 55 (b). The main difference from the noise in the 3D transistors is that each monolayer in the semiconductor film of the 2D transistor is likely contributing by band-limited 1/f noise in different frequency ranges, and the measurements can miss band-limited 1/f noise at very low frequency, e.g., below 1Hz. Thus, an extrapolation of 1/f noise spectra measured at higher frequency toward lower frequency is uncertain for 2D transistors. The band-limited 1/f noise in 2D transistors is actually a predicator to the "peculiarities" observed in 1D nanowire and carbon nanotube transistors, discussed in the preceding Sec. VI.2. **Nanotubes and nanowires – 1D seems too noisy**.

VII. Impact of LFN in circuits

The impact of low-frequency noise (LFN) depends on the purpose of the electronic circuit. Since the variety of electronic circuits is large, then it is generally impossible to look at every single case of application. We have selected two of them: radio-frequency (RF) circuits and sensors. Current efforts in RF circuits are to reduce the supply and power of the electronic circuits and to increase their speed. In the first part of this section, we shall discuss the impact of the low-frequency noise on the performance of low-voltage and low-power circuits and the up-conversion of LFN in RF circuits. Finally, the implications of the noise in sensors are briefly addressed in

sub-section VII.3. **Noise in sensors**.

VII.1. Trading power for noise

The advances in CMOS technologies put them as the preferable choice for building low-voltage and low-power electronic circuits. This is because the threshold voltages (~0.2V-0.4V) of modern MOS transistors is a fraction of the turn-on voltage (~0.6V) of silicon bipolar transistors (BJT), the static current consumption of CMOS pairs is negligible especially in digital circuits and the input gate leakage current of MOS transistors is much lower than the base current of BJT at given input bias. Also, the diversity of functions integrated in CMOS circuits is larger than that in BJT circuits at higher density of integration.

However, the low-frequency noise emerges as a problem in low-voltage and low-power electronic circuits. Some manufacturers of integrated circuits provide in datasheets, for example in [267], that the product $I_Q \times S_{V_{IN}}$ of the quiescent current I_Q and input referred noise $S_{V_{IN}}$ is a good figure of merit for the noise in amplifiers, but it is not possible to minimize the product below certain limit. Here, we shall study details that are related to the product $I_Q \times S_{V_{IN}}$ in BJT and MOS amplifiers by using characteristic values deduced in previous sections for the parameters of the transistors.

The circuit in Figure 56 is a typical topology of low-frequency amplifier with voltage feedback. The transistors, which mostly determine the noise performance, are the amplification transistor T_A and the loading transistor T_L in the first differential stage. These transistors are surrounded by a dashed line and can be MOS or BJT in BiCMOS technologies, as depicted in the figure. For simplicity, assume that the common node in the differential pair T_A–T'_A is grounded for AC signals and the voltage V_i and the resistance R_i are ½ of the actual voltage magnitude and impedance of signal source connected between input nodes IN–IN'. The differential amplifier DA in the second stage usually is with low impedance R_L, and DA suppresses the noise from biasing current source I_1. The noise from reference circuit can be filtered out by the capacitor C connected to node REF. When the loading transistors are identical, the noise from node REF also results in in-phase signal, which is suppressed by DA.

The noise contribution of the amplification transistor T_A at the input terminal IN of the circuit has voltage $S_{V_{IN}}$ and current $S_{I_{IN}}$ components. The input referred voltage noise S_{V,T_A} of the amplification transistor T_A contributes directly to $S_{V_{IN}}$. To the first order of approximation, the current component is $S_{I_{IN}}=S_{V,T_A}/(g_{IN})^2$, where is g_{IN} is the input conductance of the transistor T_A. Since the loading transistor T_L does not have a connection to the input, then T_L does not contribute to input noise current, and the noise from the T_L is referred to the input node IN of the circuit in Figure 56 as a voltage noise by $S_{I,T_L}/(g_{m,T_A})^2$, where S_{I,T_L} is the output noise current of T_L.

Therefore, the total input referred voltage of the amplifier noise is

$$S_{V_{IN}} = S_{V,T_A} + \frac{S_{I,T_L}}{g_{m,T_A}^2} = S_{V,T_A} + S_{V,T_L}\left(\frac{g_{m,T_L}}{g_{m,T_A}}\right)^2 = S_{V,T_A}\left[1+\frac{S_{V,T_L}}{S_{V,T_A}}\left(\frac{g_{m,T_L}}{g_{m,T_A}}\right)^2\right], \qquad (311)$$

where g_{m,T_A} and g_{m,T_L} are transconductances and S_{V,T_A} and S_{V,T_L} are input referred noise voltages of transistors T_A and T_L, respectively. From this equation is clear that the relative contribution of T_L to the input referred noise of the amplifier is proportional to the ratio of noise levels in T_L to T_A and it is a quadratic function of the ratio of the transistors' transconductances. Note that the ratio $g_{m,T_A}/g_{m,T_L}$ cannot be varied freely, because the same DC

current flows through T_A and T_L, since T_A and T_L are connected in series, as seen in Figure 56, and $I_{T_A}=I_{T_L}=I_{DC}=I_I/2$. Depending on whether T_A and T_L are MOS or BJT, S_{V,T_A} and S_{V,T_L} are given later by eqs. (314) and (315) for 1/f noise, and by eqs. (317), (319) and (320) for white noise.

The current noise at the input of the circuit is

$$S_{I_{IN}} = S_{V,T_A} g_{IN}^2 + S_{I_{LEAK}}, \text{ with } g_{IN} \approx \begin{cases} \dfrac{I_B}{\varphi_t} = \dfrac{I_C}{\beta\varphi_t}, \text{ for BJT} \\ 2\pi f WLC_{ox}, \text{ for MOS} \end{cases}, \tag{312}$$

We have discussed this conversion for BJT by eqs. (19) and (20), and will illustrate again with several examples below, when the conversion holds.

The additional noise from gate leakage or protection diode leakage is

$$S_{I_{LEAK}} = \frac{K_{F_{LEAK}}}{f} I_{LEAK}^2 + 2qI_{LEAK}, \tag{313}$$

which is the sum of 1/f and shot noise components, since the noise is due to overcoming of junction or insulator barrier – see eqs. (215) and (220) for gate leakage. For simplicity, we will neglect the noise from leakage, although it can be significant in MOS transistors with very thin gate insulators.

The characteristic relations and values for the parameters of the transistors, as deduced from the previous sections, are now summarized.

Form the trends in Figure 15, discussed by eqs. (213) and (214), the input referred 1/f noise voltage of a transistor is

$$S_{V_G,1/f} = \frac{1\text{Hz}}{f} \frac{FOM_{S_{VG}}}{WL} = \frac{1\text{Hz}}{f} \frac{1.3\times10^{-9}\,\mu\text{m}^2\text{V}^2/\text{Hz}}{WL}, \text{ for a MOS transistor,} \tag{314}$$

or

$$S_{V_B,1/f} = \frac{1\text{Hz}}{f} \frac{FOM_{S_{VB}}}{A_E} = \frac{1\text{Hz}}{f} \frac{3.8\times10^{-12}\,\mu\text{m}^2\text{V}^2/\text{Hz}}{A_E}, \text{ for a BJT.} \tag{315}$$

Here, WL is the gate area of MOS transistor and A_E is the emitter area of BJT. The term 1Hz is added to match the dimensions. One expects that the input referred 1/f noise voltage of the amplifier will be higher, if replacing the BJTs of emitter area A_E with MOS transistors of same gate area $WL \sim A_E$, according to

$$\left.\frac{S_{V_G}}{S_{V_B}}\right|_{1/f} = \frac{1.3\times10^{-9}\,\mu\text{m}^2\text{V}^2/\text{Hz}}{3.8\times10^{-12}\,\mu\text{m}^2\text{V}^2/\text{Hz}} \frac{A_E}{WL} \sim 300\frac{A_E}{WL}. \tag{316}$$

When using the data from ITRS [3] shown in Figure 1a, the ratio in the last equation is between $100A_E/(WL)$ and $300A_E/(WL)$, which implies that one should use large MOS transistor in order to achieve the low-noise performance of smaller-area BJT in terms of input referred 1/f noise voltage.

As shown by eq. (263), the white noise in the collector current is a sum of the collector current shot noise and the coupled shot noise from the base current. When referred to the input base terminal by the transconductance of BJT, the input referred white noise voltage of BJT is

$$S_{V_B,\text{white}} = \frac{S_{I_C}}{g^2_{m,BJT}} = \frac{2qI_C(1+\beta)}{\left(I_C/\varphi_t\right)^2} \approx \frac{2q\varphi_t\beta}{I_C/\varphi_t} = \frac{2q\varphi_t\beta}{g_{m,BJT}}, \text{ since } \beta>>1. \quad (317)$$

By multiplying with the square of the input conductance of BJT, $g_{IN}= I_B/\varphi_t=I_C/(\beta\varphi_t)$, we obtain (as expected) the white shot noise in the base current

$$S_{I_B,\text{white}} = S_{V_B,\text{white}}\left(\frac{I_C}{\beta\varphi_t}\right)^2 = 2q\frac{I_C}{\beta}\frac{(1+\beta)}{\beta} \approx 2qI_B, \text{ since } \beta>>1, \quad (318)$$

which constitutes the white noise component in the input noise current $S_{I_{IN}}$ of the amplifier, when the BJT is used as the amplification transistor T_A.

As shown by eqs. (249) and (250), and since the MOS transistor transconductance is $g_{m,MOS}=\mu C_{ox}(V_G-V_T)W/L$ in saturation regime of operation and $g_{m,MOS}=I_D/\varphi_t$ in sub-threshold regime, then the white noise in the drain current is referred to the input gate terminal by the transconductance of MOS transistor, and the input referred white noise voltage of the MOS transistor is

$$S_{V_G,\text{white}} = \frac{S_{I_D}}{g^2_{m,MOS}} = \frac{2qI_D}{\left(I_D/\varphi_t\right)^2} = \frac{2q\varphi_t}{I_D/\varphi_t} = \frac{2q\varphi_t}{g_{m,MOS}}, \text{ if } V_G<V_T \text{ (sub-threshold regime)}, \quad (319)$$

and

$$S_{V_G,\text{white}} = \frac{4kT\frac{2}{3}g_{m,MOS}}{g^2_{m,MOS}} = \frac{\frac{4}{3}2q\varphi_t}{g_{m,MOS}} \approx \frac{2q\varphi_t}{g_{m,MOS}}, \text{ if } V_D>V_G-V_T>2\varphi_t>0 \text{ (saturation regime)}. \quad (320)$$

To obtain higher gain from the input stage in the circuit, see again Figure 56, a high transconductance in the amplification transistor T_A and high impedance of loading transistor T_L at circuit node OUT are desired. Therefore, when using MOS transistors for T_A or T_L, the MOS transistors are chosen to operate in regimes either of saturation or weak inversion (sub-threshold), but not in linear (ohmic) regime. Therefore, we omit the case of linear regime.

Note the similarity between eqs. (319) and (320). By multiplying these equations with the square of the input conductance of MOS, $g_{IN}\approx 2\pi fWLC_{ox}$, we obtain the "white" noise in the gate current as

$$S_{I_G,\text{white}} = S_{V_G,\text{white}}g^2_{IN,MOS} \approx \frac{2q\varphi_t}{g_{m,MOS}}(2\pi fWLC_{ox})^2. \quad (321)$$

which is known as gate-induced noise in MOS transistors, it is frequency dependent and its power spectrum density increases with frequency, that is, $S_{I_G,\text{white}}$ actually is not white. The precise analytical expression for gate-induced noise is given in [174], it implies 4.22 times smaller value as compared to eq. (321), since the gate capacitance in saturation regime is $2WLC_{ox}/3$, the charge fluctuation is distributed along the transistor channel non-uniformly and the gate current fluctuation at given position x along the channel length coordinate is the difference between the corresponding drain-side current fluctuation and the source-side current fluctuation. The calculation is complex, it uses Bessel function. Nevertheless, the final analytical expression for gate-induced

noise in [174] has the same form as eq. (321), and since $S_{I_{G,white}}$ is low at low frequencies, then we will assume that eqs. (312) and (321) hold when estimating conservative values for gate-induced noise at low frequency.

By comparing eq. (317) for BJT to eqs. (319) and (320) for MOS transistor, one observes that the input referred white noise voltage in BJT is higher than in MOS transistors, since

$$\left.\frac{S_{V_G}}{S_{V_B}}\right|_{white} = \frac{1}{(1+\beta)}\frac{g_{m,BJT}}{g_{m,MOS}} \sim \frac{1}{100\text{ to }300}\frac{g_{m,BJT}}{g_{m,MOS}}, \tag{322}$$

by a factor equal to the current gain $\beta=I_C/I_B$ in BJT, assuming similar transconductances $g_{m,MOS}\approx g_{m,MOS}$ at given DC current. Note that this observation for input referred white noise voltage in eq. (322) is the opposite to the observation for 1/f noise in eq. (316), and the observation is valid for the input current noise of BJT, converted to voltage by $S_{V_B}=S_{I_B}z_B^2$, see below for details, which is the case of high resistance R_i of signal source in Figure 56.

Note also that the quantity

$$2q\varphi_t = 2kT = 8.3\times10^{-21}\ \text{AV/Hz} \qquad \text{at room temperature} \tag{323}$$

is twice the thermal energy and the unit AV/Hz≡Joule.

When T_A and T_L are BJTs, one uses eqs. (315) and (317) . Then, from eq. (312) and (318), the input current noise is

$$\left.S_{I_{IN}}\right|_{T_A=BJT,\,T_L=BJT} = \left(\frac{1Hz}{f}\frac{FOM_{S_{VB}}}{A_{E,T_A}}\left(\frac{I_C}{\beta\varphi_t}\right)^2 + \frac{2q\varphi_t}{\beta}\frac{I_C}{\varphi_t}\right). \tag{324}$$

and from eq. (311), by adding the contribution of loading transistor, the input referred voltage noise of the amplifier is

$$\left.S_{V_{IN}}\right|_{\substack{R_i >> z_B \\ T_A=BJT,\,T_L=BJT}} = S_{I_{IN}}z_B^2 = \left(\frac{1Hz}{f}\frac{FOM_{S_{VB}}}{A_{E,T_A}} + \frac{2q\varphi_t\beta}{I_C/\varphi_t}\right)\left(1+\begin{cases}\dfrac{A_{E,T_A}}{A_{E,T_L}}, \text{for 1/f noise}\\ 1, \text{for white noise}\end{cases}\right), \tag{325}$$

where $z_B=r_b+\varphi_t/I_B+(\beta+1)r_e\approx\varphi_t/I_B\approx\beta\varphi_t/I_C$ is the input impedance of the BJT, r_b and r_e are the resistances of the base and emitter passive regions, including resistances connected intentionally in the circuit in series with base and emitter terminals – see the discussion between eqs. (16) and (20).

Note that eq. (325) for $S_{V_{IN}}$ in BJT amplifiers is for the case when the signal source impedance R_i is large, $R_i>>z_B$, and, thus, eq. (325) represents the input current noise, which is the dominant noise source in BJT. For the case of low impedance signal source, $R_i<<z_B$, $S_{V_{IN}}$ is much lower, because $S_{V_{IN}}\approx S_{I_{IN}}(R_i+r_b+r_e)^2$ and $z_B/(R_i+r_b+r_e)$ ~10 to 100, see again the discussion between eqs. (16) and (20). We use the high impedance case when comparing to MOS amplifiers. In the application practice, however, $(r_b+r_e)<R_i<\beta\varphi_t/I_C$ and one should consider values for input voltage noise 100 to 1000 times lower than that from eq. (325) for high impedance case.

Several observations for the noise in a BJT amplifier can be made from eqs. (324) and (325) for the high-impedance case $R_i>>z_B$. Larger transistor areas reduce the 1/f noise, but they do not have effect on white noise.

The area of the loading transistor T_L has to be several times larger than the area of the amplification transistor in order to reduce the effect of 1/f noise from T_L. The input referred 1/f voltage noise is bias independent, while the input referred 1/f current noise is a strong quadratic function of the bias current. Therefore, a reduction of the bias current in micropower circuits will reduce the input current noise, but it will increase the input referred white voltage noise. The transistor current gain β has no effect on 1/f voltage noise, a higher β will decrease the white noise in the input current, but it will increase the input referred white voltage noise. The input current noise and voltage noise have similar frequency shape and corner frequencies. The loading transistor T_L always doubles the white voltage noise of the amplification transistor T_A, and this effect cannot be remedied. The critical source impedance $R_i=R_{eq}$, at which the input noise current and input noise voltage have equal contribution, is lower than the input resistance $z_B \approx \varphi_t/I_B \approx \beta\varphi_t/I_C$ of the base terminal. In approximate calculations, a reasonable choice is $R_{eq} \sim r_b \sim z_B/30$, where r_b is the resistance of the base passive region. From the discussion between eqs. (238) and (239), the noise figure has a local minimum at given frequency, when $R_i=R_{eq}$.

To optimize the performance in low power amplifiers, one can minimize the product of bias current and noise level. For BJT amplifier, assuming equally sized amplification and loading transistors in eq. (325), the product is

$$\begin{aligned} I_C S_{I_{IN}} z_B^2 = I_C S_{V_{IN}} &= \frac{1\text{Hz}}{f} FOM_{S_{VB}} \frac{I_C}{A_{E,T_A}} \left(1 + \frac{A_{E,T_A}}{A_{E,T_L}} \right) + 2q\varphi_t^2 (\beta+1)(1+1) \\ &\approx 2\left[\frac{1\text{Hz}}{f} FOM_{S_{VB}} J_C + 2q\varphi_t^2 (\beta+1) \right], \text{with } J_C = \frac{I_C}{A_E},\ A_E = A_{E,T_A} = A_{E,T_L} \text{ and } \beta = \frac{I_C}{I_B}. \end{aligned} \tag{326}$$

To preserve the contribution of shot noise from the collector current, we have put back the term (β+1), which was reduced to β in eq. (325).

Eq. (326) shows that the product $I_C S_{V_{IN}}$ is bias independent for white noise, and for 1/f noise, the product is proportional to the current density J_C, which is typically between 30μA/μm² and 3mA/μm² - see Figure 39, and higher for high-speed and RF applications, according to ITRS [3]. At given bias current I_C, in order to increase the frequency, one has to increase J_C by reducing the emitter area of T_A, which in turn, increases the 1/f noise. Thus, BJT are not very suitable for micropower amplifiers, and since $I_C S_{V_{IN}}$ is bias and size independent (at given J_C), then the tradeoff between low bias currents and high noise in micropower BJT amplifiers is evident. Using typical values for $FOM_{S_{VB}}$ from eq. (315) and range of values for the current gain β, we plot in Figure 57 the product $I_C S_{V_{IN}} = I_C S_{I_{IN}} z_B^2$ separately for 1/f noise (middle-left in the figure) and white noise (bottom-left). In the plots for BJT in this figure, the diamonds, squares and triangles correspond to $\beta_{min}=50$, $\beta_{typ}=150$ and $\beta_{max}=500$, respectively. The corner frequency f_c between 1/f noise and white noise is then plotted at the top-left in the figure, according to

$$f_c = 1\text{Hz} \frac{FOM_{S_{VB}} J_C}{2q\varphi_t^2 \beta} \approx 1\text{Hz} \frac{FOM_{S_{VB}}}{A_E 2q\varphi_t z_B} = \frac{4.6\times10^8 \mu\text{m}^2 \Omega\text{Hz}}{A_E z_B}, \tag{327}$$

as follows from eq. (326) for the condition S(1/f)=S(white). Note that the plots for BJT amplifier in Figure 57 are for the high-impedance case, $R_i >> z_B$, which represent the input current noise, while the actual input voltage noise at $R_i << z_B$ will be about 2-3 decades lower, since $S_{V_{IN}} \approx S_{I_{IN}}(R_i+R_{eq})^2 \approx S_{I_{IN}}(R_{eq})^2$ and $R_{eq} \sim r_b \sim z_B/30$, as mentioned above, but never below the limit $2q\varphi_t^2 = 2.1\times10^{-22}$ V²A/Hz, which is set by the shot noise of the

collector current, since $2q\varphi_t^2=I_C\times 2qI_C\times(\varphi_t/I_C)^2=I_C\times S_{I_C shot}/g_m^2=I_C S_{V_{IN}}$.

When T_A and T_L are MOS transistors operating in weak inversion (sub-threshold) regime, $V_G<V_T$, one uses eqs. (314) and (319) . Then, from eq. (311), the input referred voltage noise of the amplifier is

$$S_{V_{IN}}\Big|_{\substack{V_G<V_T-2\varphi_t\\ T_A=\mathrm{MOS},T_L=\mathrm{MOS}}}=\left(\frac{1\mathrm{Hz}}{f}\frac{FOM_{S_{VG}}}{W_{T_A}L_{T_A}}+\frac{2q\varphi_t}{I_D/\varphi_t}\right)\left(1+\begin{cases}\dfrac{W_{T_A}L_{T_A}}{W_{T_L}L_{T_L}}, \text{for 1/f noise}\\ 1, \text{for white noise}\end{cases}\right), \tag{328}$$

and from eq. (312), the input current noise is

$$S_{I_{IN}}\Big|_{\substack{V_G<V_T-2\varphi_t\\ T_A=\mathrm{MOS},T_L=\mathrm{MOS}}}=S_{V,T_A}g_{IN}^2=$$
$$\approx(2\pi C_{ox})^2\times\begin{bmatrix}(1\mathrm{Hz}\,FOM_{S_{VG}}W_{T_A}L_{T_A}f), \text{from 1/f noise}\\ +\dfrac{2q\varphi_t}{I_D/\varphi_t}(W_{T_A}L_{T_A}f)^2, \text{from white noise}\end{bmatrix}. \tag{329}$$

The corresponding equations, when T_A and T_L are MOS transistors operating in strong inversion (saturation) regime, $V_D>V_G-V_T>0$, by using eq. (320) instead of eq. (319), are: for the input referred voltage noise of the amplifier

$$S_{V_{IN}}\Big|_{\substack{V_D>V_G-V_T>2\varphi_t\\ T_A=\mathrm{MOS},T_L=\mathrm{MOS}}}=$$
$$=\left(\frac{1\mathrm{Hz}}{f}\frac{FOM_{S_{VG}}}{W_{T_A}L_{T_A}}+\frac{2q\varphi_t}{I_D/\varphi_t}\frac{V_G-V_T}{2\varphi_t}\right)\left(1+\begin{cases}\left(\dfrac{L_{T_A}}{L_{T_L}}\right)^2, \text{for 1/f noise}\\ \sqrt{\dfrac{W_{T_L}/L_{T_L}}{W_{T_A}/L_{T_A}}}, \text{for white noise}\end{cases}\right), \tag{330}$$

and for the input current noise

$$S_{I_{IN}}\Big|_{\substack{V_D>V_G-V_T>2\varphi_t\\ T_A=\mathrm{MOS},T_L=\mathrm{MOS}}}=S_{V,T_A}g_{IN}^2=$$
$$\approx(2\pi C_{ox})^2\times\begin{bmatrix}(1\mathrm{Hz}\,FOM_{S_{VG}}W_{T_A}L_{T_A}f), \text{from 1/f noise}\\ +\dfrac{2q\varphi_t}{I_D/\varphi_t}\dfrac{V_G-V_T}{2\varphi_t}(W_{T_A}L_{T_A}f)^2, \text{from white noise}\end{bmatrix}. \tag{331}$$

As compared to eqs. (324) and (325) for the noise in a BJT amplifier, several similarities and differences for the noise in MOS amplifier can be made. In sub-threshold regime, the input referred voltage noise has the same

expression as for BJT, with differences that the white noise component is not multiplied by current gain β, see again eq. (322), and $FOM_{S_{VG}} >> FOM_{S_{VB}}$, see eq. (316). In contrast to BJT, when MOS transistors are operating in saturation regime, the amplifier input referred 1/f voltage noise component, $S_{V_{IN}}$ in eq. (330), depends on the ratio between gate lengths of T_A and T_L, rather than on ratio of gate areas, and the amplifier input referred white voltage noise component is decreasing "slowly" as $I_D^{-0.5} \propto (V_G - V_T)/I_D$, rather than as $1/I_D$. Also, the relative contribution of the loading transistor depends on the square root of aspect ratio W/L of T_L and T_A, rather than being equal to 1. Note that eqs. (330) and (331) are derived at condition of equal DC currents flowing through T_A and T_L, as mentioned earlier for the circuit in Figure 56, and the current noise $S_{I_{IN}}$ in the case of MOS amplifier increases with the frequency, while in the case of BJT amplifier, $S_{I_{IN}}$ was decreasing as 1/f at low frequency or constant in the frequency range of white noise.

Other important difference between MOS and BJT amplifiers, is that the critical source impedance $R_i = R_{eq}$, at which the input noise current and input noise voltage have equal contributions, is frequency dependent in the case of MOS amplifiers. For sub-threshold regime, R_{eq} is

$$R_{eq}\Big|_{\substack{V_G < V_T - 2\varphi_t \\ T_A = \text{MOS}, T_L = \text{MOS}}} = \sqrt{\frac{S_{V_{IN}}}{S_{I_{IN}}}} = \frac{1}{2\pi f W_{T_A} L_{T_A} C_{ox}} \times \left\{ \begin{array}{l} \sqrt{1 + \dfrac{W_{T_A} L_{T_A}}{W_{T_L} L_{T_L}}}, \text{for 1/f noise} \\ \sqrt{2}, \text{for white noise} \end{array} \right\}, \quad (332)$$

and for saturation regime, R_{eq} is

$$R_{eq}\Big|_{\substack{V_D > V_G - V_T > 2\varphi_t \\ T_A = \text{MOS}, T_L = \text{MOS}}} = \sqrt{\frac{S_{V_{IN}}}{S_{I_{IN}}}} = \frac{1}{2\pi f W_{T_A} L_{T_A} C_{ox}} \times \left\{ \begin{array}{l} \sqrt{1 + \left(\dfrac{L_{T_A}}{L_{T_L}} \right)^2}, \text{for 1/f noise} \\ \sqrt[4]{1 + \dfrac{W_{T_L} / L_{T_L}}{W_{T_A} / L_{T_A}}}, \text{for white noise} \end{array} \right\}. \quad (333)$$

Evidently from eqs. (332) and (333), the critical source impedance $R_{eq} \propto 1/g_{IN}$, R_{eq} is inversely proportional to frequency f and to the gate area ($W_{T_A} L_{T_A}$) and gate capacitance ($W_{T_A} L_{T_A} C_{ox}$) of the amplification transistor T_A in MOS amplifier. Note that in a MOS amplifier, R_{eq} does not depend on the bias at given geometry of MOS transistors.

The geometry of MOS transistors, however, depends on the range of bias currents and biasing conditions, while the geometry of BJT depends mostly on the desired current density I_C/A_E, as discussed above. For a given MOS process, there is a characteristic current I_{DO}, given by

$$I_{DO} = 2\mu C_{ox} \varphi_t^2, \quad (334)$$

which can be regarded as the drain current of square-shaped MOS transistor (W=L) at $V_G = V_T$ [169], see after eq. (253). The characteristic current I_{DO} depends on the MOS technology, and it ranges between 100–1000nA, since μC_{ox} ~0.07–0.7mA/V² for mobility μ~150–300 cm²/Vs, EOT~1.4–7nm and $\varphi_t \approx 26$mV at room temperature. Typically, I_{DO}~330nA, corresponding to μC_{ox} ~0.25mA/V². At a gate bias V_G apart about $\pm 2\varphi t \approx \pm 50$mV from threshold voltage V_T, the drain current in the square-shaped MOS transistor is

$$I_{Dsq} = I_D \frac{L}{W} \approx I_{DO} \ln^2\left[1+\exp\left(\frac{V_G - V_T}{2\varphi_t}\right)\right]$$
$$\approx I_{DO} \times \begin{cases} \exp\left(\frac{V_G - V_T}{\varphi_t}\right), \text{for } (V_G - V_T) < -2\varphi_t \text{ (in sub - threshold regime)} \\ \left(\frac{V_G - V_T}{2\varphi_t}\right)^2, \text{for } (V_G - V_T) > 2\varphi_t \text{ (in saturation regime)} \end{cases} \quad (335)$$

Here, $W=W_{T_A}=W_{T_L}$ and $L=L_{T_A}=L_{T_L}$ are the gate widths and lengths, and the drain current in the MOS transistors is $I_D=I_{Dsq}W/L$. Substituting in eqs. (328) and (330), by also assuming nearly equally sized amplification and loading transistors, and following the long-channel MOS transistor model (page 427 in [169]), the product of bias current and noise level in a MOS amplifier is

$$I_D S_{V_{IN}} \approx 2\left[\frac{1Hz}{f} FOM_{S_{VG}} \frac{I_{DO}}{L^2}\left(\frac{I_{Dsq}}{I_{DO}}\right) + 2q\varphi_t^2\left(0.5+\sqrt{0.25+\frac{I_{Dsq}}{I_{DO}}}\right)\right]$$
$$\approx 2\left[\frac{1Hz}{f} FOM_{S_{VG}} \frac{I_{DO}}{L^2}\left(\frac{I_{Dsq}}{I_{DO}}\right) + 2q\varphi_t^2 \times \begin{cases} 1, & \text{for } (V_G - V_T) < -2\varphi_t, \text{ and } V_D > 3\varphi_t \\ \sqrt{\frac{I_{Dsq}}{I_{DO}}} \approx \frac{V_G - V_T}{2\varphi_t}, & \text{for } V_D > (V_G - V_T) > 2\varphi_t \end{cases}\right], \quad (336)$$

where the term $\left(0.5+\sqrt{0.25+\frac{I_{Dsq}}{I_{DO}}}\right)$ is denoted as f(u) on page 427 in [169], it changes from 1 in subthreshold regime to $(V_G-V_T)/(2\varphi_t)$ in strong inversion regime above threshold, and it provides for the transconductance g_m a smooth transition from exponential to linear dependence as function of gate bias below and above threshold voltage, respectively. Other soothing functions for g_m can be found in [294].

Eq. (336) shows for MOS amplifiers that the product $I_D\times$(white input voltage noise) is bias independent in subthreshold regime, and the product increases linearly as $(V_G-V_T)/(2\varphi_t)$ at gate bias above threshold. For the 1/f noise, the product increases with the level of channel inversion (I_{Dsq}/I_{DO}) and with "surface current density" (I_{Dsq}/L^2), which is similar to the dependence on current density in BJT. There is no dependence on gate width W in eq. (336), but W links I_D, L and I_{Dsq} by eq. (335), and W<L, if $I_{Dsq}>I_D$, which is possible at high gate bias and low currents in micropower circuits, e.g., $(V_G-V_T)/(2\varphi_t)>3$ for $(V_G-V_T)>0.15V$, I_{DO}~0.33µA for EOT=2.8nm and µ=200cm²/Vs, and $I_D<3\mu A$, since $I_{Dsq}\sim 3^2\times 0.33\mu A\sim 3\mu A$.

Similarly to BJT amplifiers, the product $I_D S_{V_{IN}}$ implies a tradeoff between low bias currents and high noise in micropower MOS amplifiers, at given current density (I_{Dsq}/L^2) and level of channel inversion (I_{Dsq}/I_{DO}), the later corresponding to a fixed gate overdrive voltage (V_G-V_T). Using typical values for $FOM_{S_{VG}}$ from eq. (314), we plot in Figure 57 the product $I_D S_{V_{IN}}$ given by eq. (336) for MOS amplifiers separately for 1/f noise (middle-right in the figure) and white noise (bottom-right). We assume that the gate length of MOS transistors for analog circuits is about 3-5 times the minimum gate length L_{min} of particular CMOS technology. The diamonds are for minimum $I_{DO}=0.1\mu A$ and L=1.5µm, and correspond to pMOS of from nodes with $L_{min}=0.35\mu m$ to 0.5µm (EOT~7nm, C_{ox}~0.5µF/cm², µ~150cm²/Vs). The squares are for $I_{DO}=0.333\mu A$ and L=0.5µm, and correspond to

nMOS from nodes with L_{min}=90nm to 130nm (EOT~2.8nm, C_{ox}~1.2μF/cm², μ~200cm²/Vs). The triangles are for I_{DO}=1μA and L=0.2μm, and correspond to advanced MOS nodes with L_{min}<65nm (EOT~1.4nm, C_{ox}~2.5μF/cm², μ~300cm²/Vs), which is expected to be used also in analog applications in the near future. For convenience, and in order to show the gate overdrive (V_G–V_T) in the horizontal axis at the top of Figure 57, the plots for MOS amplifiers are given versus the level of channel inversion (I_{Dsq}/I_{DO}), which is a dimension-less quantity, rather than versus current density.

The corner frequency f_c between 1/f noise and white noise is then plotted at the top-right in Figure 57, for MOS amplifiers that use transistors with the abovementioned parameters, according to

$$f_c = 1\text{Hz}\frac{\text{FOM}_{S_{VG}}\dfrac{I_{DO}}{L^2}\left(\dfrac{I_{Dsq}}{I_{DO}}\right)}{2q\varphi_t^2\left(0.5+\sqrt{0.25+\dfrac{I_{Dsq}}{I_{DO}}}\right)} \approx 1\text{Hz}\frac{\text{FOM}_{S_{VG}}\dfrac{I_{DO}}{L^2}}{2q\varphi_t^2}\times\begin{cases}\left(\dfrac{I_{Dsq}}{I_{DO}}\right)\approx\exp\left(\dfrac{V_G-V_T}{\varphi_t}\right), & \text{for } (V_G-V_T)<-2\varphi_t, \text{ and } V_D>3\varphi_t \\ \sqrt{\left(\dfrac{I_{Dsq}}{I_{DO}}\right)}\approx\dfrac{V_G-V_T}{2\varphi_t}, & \text{for } V_D>(V_G-V_T)>2\varphi_t\end{cases} \tag{337}$$

as follows from eq. (336) for the condition S(1/f)=S(white). Note that the plots for MOS amplifier in Figure 57 are for low-impedance case, when the signal source resistance $R_i<<2\pi fWLC_{ox}$, and the plots represent the input referred voltage noise of MOS amplifier, whereas, the plots for BJT amplifier are for high-impedance case, when the signal source resistance $R_i>>z_B=\beta\varphi_t/I_C$, and the plots represent the input current noise of BJT amplifier via $S_{V_{IN}}=S_{I_{IN}}z_B^2$, as discussed earlier after eq. (327).

Several useful comparisons between MOS and BJT amplifiers can be made in Figure 57 for the product $I_{DC}S_{V_{IN}}$. For the white noise, the product is independent of transistor size, but increases with bias level (I_{Dsq}/I_{DO}) in MOS amplifiers, whereas, the product is independent of bias level $J_C=I_C/A_E$, but it depends on the current gain β of BJT used in the amplifier. For the 1/f noise, the product increases with the bias level in both MOS and BJT amplifiers, but the product is independent of the current gain β in BJT amplifiers, whereas, it increases as $I_{DO}/L^2\propto\mu C_{ox}/L^2$ in MOS amplifiers. The corner frequency f_c decreases with the current gain β in BJT amplifiers, owing to higher white noise, whereas, f_c increases as C_{ox}/L^2 in MOS amplifiers, owing to higher 1/f noise. Consequently, f_c is usually in the kHz range for BJT amplifiers, whereas, f_c can reach GHz range in MOS amplifiers. Thus, one can observe both 1/f noise and white noise in low-frequency BJT amplifiers, whereas, the 1/f noise dominates in the entire low-frequency range below 1MHz in MOS amplifiers with short gates, e.g. L<1μm. Note again that the comparison is between voltage noise in MOS amplifiers and current noise in BJT amplifiers, the later multiplied by the square of input resistance of BJT.

The relations for the product $I_{DC}S_{V_{IN}}$ in eq. (326) for BJT amplifiers and in eq. (336) for MOS amplifiers are generic, and we illustrate in Figure 58 how they are reflected in commercially available integrated amplifiers from different fabricators, Analog Devices, Linear Technology, and Texas Instruments and Burr-Brown. From the datasheets of the amplifiers, we have collected the values for 1/f and white noise, maximum quiescent

current I_Q and bandwidth. As shown in the left-hand plots of Figure 58, on the top for BJT and below for MOS amplifiers, increasing I_Q, the bandwidth increases, the white noise decreases, while the 1/f noise scatters generally.

Obviously, the datasheets provide performance parameters of the amplifiers and the details for the designs are omitted. Since the current I_{DC} and the sizes of the transistors in the input stage are not stated in datasheets, then we assume that I_{DC} is a portion of the quiescent current I_Q of the amplifier, and I_{DC} scales approximately linearly with I_Q, in order to preserve the bandwidth of the amplifier. Then, the current density in the input transistors of BJT amplifiers is estimated from the corner frequency f_c between 1/f and white noise, according to eq. (327), and using typical values for β=150 and $FOM_{S_{VB}}=3.8\times10^{-12}$ $\mu m^2V^2/Hz$ from eq. (315). Re-arranging the data against the current density, as illustrated at the top-right plot of Figure 58, the product $0.5I_Q\times S_{V_{IN}}$ for BJT amplifiers fits with the predictions of eq. (326) given earlier for the design window in Figure 57. In particular, the 1/f noise is in agreement with the general trend (solid squares) for $2J_C FOM_{S_{VB}}$, having a distribution illustrated with bell-shaped curve with standard deviation 5.5dB on right of the trend. The product $0.5I_Q\times S_{V_{IN}}$ for white noise (triangles) also became with trend independent of the current density, having a distribution with a peak within the limits given by the design window for β=50 (solid diamonds) and β=500 (solid triangles), and standard deviation also 5.5dB. The similar values for the standard deviations for 1/f noise and white noise, as well as the increase of the bandwidth as function of current density, imply that the scattering around the trends is due to differences in the design of the amplifiers, but in general, the predictions from of eqs. (326) and (327) hold for BJT amplifiers, by taking 50% of I_Q in the product $I_{DC}\times S_{V_{IN}}$. The high value of 50% is empirical, and it is justified by the fact that the input referred voltage noise of BJT is normally lower in low impedance circuits, see again the discussion after eq. (327), and one has to take 10-20 times the biasing current I_{DC} of the input stage of the amplifier in order to compare to eq. (326), which is derived for the dominant current noise in high-impedance circuits. Indeed, as given in the datasheet of the ultimately low noise amplifiers LT1028 and LT1128 from Linear Technology, the DC current is 1.8mA in the input stage, and it is a significant portion (~25%) of the quiescent current 7.5mA.

As for MOS amplifiers, it is not possible to calculate gate overdrive and gate length from the information in datasheets. Therefore, the portion of I_Q in the product $I_{DC}\times S_{V_{IN}}$ was varied until no data point left below the ultimate minimum of $2q\varphi t^2$ for the product, as follows from eq. (336) and assuming negligible noise contribution of the loading transistor. Using $0.02I_Q\times S_{V_{IN}}$, the results are shown in the bottom-right plot of Figure 58. In agreement with the prediction from eq. (336), the product with the white noise (triangles) is low at low I_Q, and it increases slowly at higher I_Q. There is bimodal distribution in the data, which corresponds to $I_{DO}(W/L)$=2µA and 150nA, and for these values, the predictions from the term $2\times2q\varphi t^2[0.5+(0.25+I_{Dsq}/I_{DO})^{0.5}]$ of eq. (336) is illustrated by solid lines through the triangles. Along with the increase of the bandwidth, this confirms that the current density (I_{Dsq}/L^2) increases with I_Q, and according to the prediction from eq. (336) for the 1/f noise, the product $0.02I_Q\times S_{V_{IN}}$ is nearly proportional to I_Q for 1/f noise, as shown in the bottom-right plot of Figure 58 with circles and a trend line through them. The distribution in the data is wide, having standard deviation about 7dB, with limits shown by solid lines. The upper limit corresponds to 5V MOS with EOT=10nm, L=3µm, WL=450µm², low input capacitance of 1.6pF, and $FOM_{S_{VG}}$=500 $\mu m^2\mu V^2/Hz$. The low limit corresponds to 30V MOS with EOT=30nm, L=40µm WL=16000µm², high input capacitance of 18pF , and $FOM_{S_{VG}}$=100

μm²μV²/Hz. The capacitances are within the range given in the datasheets, the oxide thicknesses correspond to the maximum supply voltages in the datasheets, and the values for $FOM_{S_{VG}}$ are within the range for analog MOS in ITRS [3].

Overall, the right-hand plots in Figure 58 demonstrate that the relations given in eq. (326) for BJT amplifiers and in eq. (336) for MOS amplifiers for the product $I_{DC}\times S_{V_{IN}}$ are valid, and the relations are observed in the commercially available integrated amplifiers as extension of the design window in Figure 57 toward low current densities, I_C/A_E for BJT or I_{DO}/L^2 for MOS. Therefore, the IC manufactures use $I_Q\times S_{V_{IN}}$ as the figure of merit for their low-power low-noise amplifiers – see [295] for a datasheet from Linear Technology, for example.

VII.2. Up- and cross-conversion in phase noise

VII.2.1. ***Definitions***

The noise in ideal linear circuits is additive, that is, the noise and the signal occur simultaneously and independently each from other. However, many circuits used for signal generation and processing are not linear, they are also time variant, and the low-frequency noise is "multiplied" in these circuits, even when the circuit operates at high frequency. This effect is known as up-conversion of low-frequency noise, in which the properties of the high frequency signal, e.g. amplitude, phase, delay, are affected by the low-frequency noise. The up-conversion of noise degrades the spectral purity of the signal, widening the spectrum of the signal, and in time domain it causes jitter in the transitions of the stationary cycling signals. There are many works devoted on up-conversion of low-frequency noise, and the up-conversion was reviewed in [296] with emphasis on BJT circuits.

The problem of up-conversion can be introduced when looking at the modifications of ideal signal, when passing through a circuit. The modifications are assumed as modulations, and given generally by

$$V(t) = \left[V_s + \Delta_a(t)\right]\sin\left[2\pi f_s t + \Delta_\varphi(t)\right], \tag{338}$$

where Δ_a is fluctuation in amplitude and Δ_φ is fluctuation in phase of the original signal $V_o=V_s\sin(2\pi f_s t)$, the latter with amplitude V_s and frequency f_s. The frequency f_s of RF signals is usually high as compared to the frequency range of the spectrum of the low-frequency fluctuations Δ_a and Δ_φ. Therefore, Δ_a and Δ_φ appear as amplitude and phase modulations of the original signal, and respectively, are regarded as amplitude and phase noise. (Other forms of signal modification can be also assumed, for example, as in perturbation phase noise theories, which we will discuss later.)

The noise from devices contributes to both amplitude and phase noise in eq. (338). Both contributions result in broadening of the spectrum of the signal around frequency f_s. For the amplitude noise, the product $\Delta_a(t)\sin(2\pi f_s t)$ results in convolution $\Delta_a(\Delta f)*\delta(f_s-\Delta f)$ in frequency domain between noise spectrum $\Delta_a(\Delta f)$ at low frequency $\Delta f<<f_s$ and spectral line $\delta(f_s-\Delta f)$ of the signal. The convolution creates side-lobes around signal frequency f_s at offset frequencies Δf, which can be further modified by the shape of the amplitude-frequency response of the load, e.g. LC tank in RF circuits. In similar manner, the phase noise also creates side-lobes around signal frequency f_s, since $\sin[2\pi f_s t+\Delta_\varphi(t)]=\sin(2\pi f_s t)\cos[\Delta_\varphi(t)]+\cos(2\pi f_s t)\sin[\Delta_\varphi(t)]$ in time domain, and when converted in frequency domain, it also results in convolution $\delta(f_s-\Delta f)*\Delta_\varphi(\Delta f)/\Delta f$. (The division on Δf occurs, because one can take the spectrum of phase noise at Δf as constant at $f_s>>\Delta f$, having spectrum inversely proportional to Δf after Fourier transformation. Precisely, the division is due to the fact that the frequency is time derivative of the

phase, as shown later.) The term $\Delta_\varphi(\Delta f)/\Delta f$ results in "addition" of $1/\Delta f^2$ slope to the frequency dependence of low-frequency noise, when addressing power spectrum densities (PSD) at high frequency. In other words, the PSD of phase noise decreases as $1/\Delta f^2$ steeper, when compared to PSD of the noise source that causes the phase fluctuation. The above explanations are heuristic, but they capture the main properties of amplitude and phase noise. The typical spectrum of an oscillator is shown in Figure 59a for a CMOS oscillator [297]. When taking one side of the spectrum, either below or above oscillation frequency f_s, one obtains the single side band (SSB) noise. The ratio of the power spectrum density of SSB noise to the magnitude of the carrier signal at the oscillation frequency defines the normalized spectrum of the phase noise, PhN, according to

$$\begin{aligned} \mathrm{PhN}(\Delta f) &= \frac{S_{SSB}(\Delta f)}{P_{carrier}}, \text{in unit } [1/\mathrm{Hz}] \\ \Leftrightarrow &= 10\mathrm{dB} \times \log_{10}(\mathrm{PhN}), \text{in}[\mathrm{dBc}/\mathrm{Hz}], \end{aligned} \tag{339}$$

where $P_{carrier}$ is the power of the signal at the output of the oscillator and S_{SSB} is the power spectrum density of SSB noise at $\Delta f=f-f_s$ frequency offset from oscillation frequency f_s. One usually reports the data for PhN in units [dBc/Hz], which are according the second line of the equation. Figure 59b illustrates the PhN spectrum obtained from the spectrum of the oscillator in Figure 59a, showing also the two typical slopes, $1/\Delta f^3$ and $1/\Delta f^2$, in PhN, which are result of up-conversion of 1/f and white noise, respectively, by "addition" of $1/\Delta f^2$ to the slopes of the low-frequency noise, as mentioned above and discussed later.

The amplitude noise is not a severe problem, since most practical circuits possess amplitude limiting mechanism [296], e.g. automatic gain control in oscillators and limiters for mixers, resulting in substantial suppression of the amplitude noise. However, the phase noise is serious problem, since a phase feedback is difficult, and the spectral broadening of the oscillation signal due to phase noise is perhaps the most critical issue for reference oscillators, in which the oscillator is free-running. Therefore, from here to the end of this sub-section, we focus on phase noise in oscillators.

VII.2.2. Phase-noise in oscillators

There are three approaches studying the phase noise in oscillators. These are time invariant, cyclostationary and perturbation approaches.

Time invariant approach

The time invariant approach was first suggested in [298], it is used widely, and a detailed review of the approach is given in [296]. It is assumed the obvious configuration of an electronic oscillator, in which an amplifier with noise figure NF, see eq. (228), has a frequency dependent LC–R_s feedback so that at frequency $(2\pi f_s)^2=1/(LC)$ the feedback is positive without phase shift, and the circuit generates sinusoidal signal at f_s. An assumption is made in [298] that power spectrum density (PSD) of the phase fluctuation is equal to the ratio of amplifier noise to oscillation power. This assumption is based on eq. (338) in the following manner.

$$\frac{dV(t)}{V_s} = d\left\{\sin\left[2\pi f_s t + \Delta_\varphi(t)\right]\right\} = \cos\left[2\pi f_s t + \Delta_\varphi(t)\right] d\Delta_\varphi(t), \tag{340}$$

which rewritten for noise becomes

$$\frac{S_V}{V_s^2} = \cos^2\left[2\pi f_s t + \Delta_\varphi(t)\right] \; S_\varphi \approx \frac{S_\varphi}{2}, \tag{341}$$

since the time invariant approach takes the average value of $\cos^2(x)=0.5$. Then, by taking into account that $V_{RMS}=V_s/\sqrt{2}$ for sinusoidal oscillation, then the voltage noise of the amplifier is expressed with noise figure, as

$$\frac{S_V}{V_s^2}=\frac{4kTR_sNF}{V_s^2}=\frac{4kT}{V_s^2/R_s}NF=\frac{2kT}{P_{carrier}}NF=\frac{2kT}{P_{carrier}}NF_{wh}\left(1+\frac{\Delta f_c}{\Delta f}\right), \quad (342)$$

where NF_{wh} is effective noise figure of for white noise in respect to the resistance R_s of the LC tank at resonance frequency f_s, and Δf_c is effective corner frequency between 1/f and white noise, as introduced in [298]. Thus, the PSD of the noise in the phase is

$$\frac{S_\varphi}{2}\approx\frac{2kT}{P_{carrier}}NF_{wh}\left(1+\frac{\Delta f_c}{\Delta f}\right), \quad (343)$$

where $S_\varphi/2$ is SSB component corresponding to one side-lobe of noise.

The fluctuation S_φ in the phase is applied to the LC tank, which converts it in frequency noise $S_{\Delta f}$, as follows. The LC tank has (-3dB) bandwidth

$$\pm\frac{BW}{2}=\frac{f_s}{Q}, \quad (344)$$

where Q is the quality factor of the resonator used in the feedback of the oscillator. When the frequency offset is small, $\Delta f<\pm BW/2$, then the relation between phase and frequency in the LC tank is approximately

$$\pm\frac{\partial\varphi}{\partial(\Delta f)}=\frac{2Q}{f_s}, \text{ at } f\approx f_s \text{ and } f-f_s=\Delta f<\pm BW/2. \quad (345)$$

Therefore, $d(\Delta f)=d\varphi\times f_s/(2Q)$, which rewritten for noise $S_\varphi/2$ in $\pm BW/2$ is

$$S_{\Delta f}=\left(\frac{f_s}{2Q}\right)^2\frac{S_\varphi}{2}, \text{ at } f\approx f_s \text{ and } f-f_s=\Delta f<\pm BW/2. \quad (346)$$

For frequency deviation larger than $\pm BW/2$, the phase of the LC tank does not change significantly with frequency. Therefore, the frequency and phase fluctuations are directly coupled each to other, without modification from LC tank. Owing to the general relation $2\pi(df)=\partial(d\varphi)/\partial t$ between phase φ and frequency f, which also holds for the output of the oscillator, then phase fluctuation can be equally rewritten as frequency fluctuation [298]

$$S_{\Delta f}(\Delta f)=(\Delta f)^2\frac{S_\varphi(\Delta f)}{2}, \text{ at } f\sim f_s, \text{ but } f-f_s=\Delta f>\pm BW/2. \quad (347)$$

Rewritten for the oscillator output, the last relation is

$$\frac{S_{\Delta f}(\Delta f)}{(\Delta f)^2}=\frac{S_\theta(\Delta f)}{2}=\frac{S_{SSB}}{P_{carrier}}=PhN(\Delta f), \text{ at } f\approx f_s \text{ and any } f-f_s=\Delta f, \quad (348)$$

where S_θ is noise in the phase and S_{SSB} is the single-side band PSD of the oscillator output signal. Combining eqs. (346) and (347), as suggested in [298], and substituting in eq. (348), one gets

$$\mathrm{PhN}(\Delta f)=\frac{S_{SSB}}{P_{carrier}}=\frac{S_{\Delta f}(\Delta f)}{(\Delta f)^2}=\frac{S_{\varphi}(\Delta f)}{2}\left[1+\left(\frac{f_s}{2Q\Delta f}\right)^2\right],\ \text{at } f\approx f_s \text{ and any } f-f_s=\Delta f. \tag{349}$$

When substituting with eq. (343), the expression for time invariant approach for phase noise in oscillators becomes as

$$\mathrm{PhN}(\Delta f)=\frac{S_{SSB}}{P_{carrier}}=\frac{2kT}{P_{carrier}}NF_{wh}\left(1+\frac{\Delta f_c}{\Delta f}\right)\left[1+\left(\frac{f_s}{2Q\Delta f}\right)^2\right],\ \text{at } f\approx f_s \text{ and any } f-f_s=\Delta f, \tag{350}$$

where the noise figure NF_{wh} is for white noise in respect to the resistance R_s of the LC tank at resonance frequency f_s, as mentioned above. In RF oscillators generating in the frequency range of GHz, one usually uses LC tanks with Q<100, and $f_s/(2Q)$>10MHz. Therefore, the last term in the square brackets dominates, and practically for Δf<1MHz

$$\mathrm{PhN}(\Delta f)\approx\frac{2kT}{P_{carrier}}NF_{wh}\left(1+\frac{\Delta f_c}{\Delta f}\right)\left(\frac{f_s}{2Q\Delta f}\right)^2\propto\begin{cases}(1/\Delta f)^2, \text{if } \Delta f>\Delta f_c, \text{white noise up-conversion}\\ (1/\Delta f)^3, \text{if } \Delta f<\Delta f_c, \text{1/f noise up-conversion}\end{cases} \tag{351}$$

Consequently, one observes two slopes, $1/\Delta f^3$ and $1/\Delta f^2$, in oscillator phase noise, as illustrated in Figure 59b. Eqs. (350) and (351) are known as Leeson formula for phase noise [298].

Certainly, the time invariant approach captures very essential features related with the phase noise. These are the abovementioned two frequency slopes, and the requirements for high quality factor Q of the resonator, low noise figure and low corner frequency between 1/f and white noise in the amplifier. However, there are issues, because Δf_c is different, usually lower, than the corner frequency f_c between 1/f and white noise in the amplifier, and also, there is no accurate expression that relates NF_{wh} to the noise figure NF of the amplifier, the latter given for low-frequency noise by eq. (228), for example.

Cyclostationary approaches

The above issues were found to originate to circuit asymmetry and high-order harmonics in oscillators, after applying cyclostationary analyses. Earlier experiments, such as that in [299], showed that improving the circuit symmetry, and thus reducing harmonics related to non-linearity, decrease the SSB lobes and phase noise. The cyclostationary analyses are area of active research [300, 301, 302], they are lengthily to be presented here in full, and there is no review available at present, which compares different approaches, in computer simulators. The approach is introduced in [303], it is sketched in [304], the derivation of the equations is presented at [305], and summarized in [212], as

$$\mathrm{PhN}(\Delta f)\approx\left|\frac{\partial f_s}{\partial i_0}\right|^2\frac{S_{1/f}}{(\Delta f)^2}+\sum_{m=0}^{N}\sum_{n=0}^{N}\left[\frac{\partial f_s}{\partial i_m}\right]S_{m,n}\left[\frac{\partial f_s}{\partial i_n}\right]^*\frac{1}{2(\Delta f)^2}, \tag{352}$$

where $\partial f_s/\partial i_m$ are sensitivities of the (oscillation) frequency f_s to changes in harmonic with number m=0,1…N, $[\partial f_s/\partial i_n]^*$ are complex conjugated values of the sensitivities of the (same) harmonics with number n=0,1…N, and the white noise power spectrum densities $S_{m,n}$ are from all noise sources weighted by magnitude of the harmonics (e.g. for shot noise in BJT, $S_{m,n}=2qI_{m-n}$, where I_{m-n} is the magnitude of the $(m-n)^{th}$ harmonic in the BJT current). Note that the 1/f noise, $S_{1/f}$, and any other low-frequency noise, contributes only by the DC

component i_0 of the signal, as shown with the term in front of the sum, whereas the high frequency white noise around all harmonics also contributes to the phase noise around the fundamental harmonic with frequency f_c. In other words, the up-conversion of low-frequency noise is only via "DC component", the 0^{th} harmonic, of the signal derivatives in oscillators, while the white noise around all harmonics adds on the top of the up-converted low-frequency white noise, according to harmonic balance method for cyclostationary analyses of phase noise. Thus, in oscillators, in which the signals are large and harmonics are present, the phase noise is not solely due to up-conversion of low-frequency noise; and the corner frequency Δf_c between $1/\Delta f^2$ and $1/\Delta f^3$ phase noise components is lower in oscillators as compared to the corner frequency f_c between low-frequency white and 1/f noise of the transistors biased at the same DC conditions as in the oscillators. This is also confirmed by perturbation methods for analysis of phase noise, which are summarized below. Owing to the importance of balance between DC, amplitude of oscillation and linearity for phase noise, we will discuss the tradeoff immediately after the paragraphs for the perturbation approach, as the third design consideration – see later and below eq. (398).

Perturbation approach

The perturbation approach for analysis of phase noise assumes that the noise perturbs the operation of the noiseless oscillator. As mentioned earlier, eq. (338) is not the only way to describe the modification of ideal signals. The perturbations can be assumed and introduced in various manners, e.g. in amplitude Δ_a and phase Δ_φ by eq. (338), or in other form, such as

$$V(t) = \Delta_a(t) + v_o[t + \Delta_t(t)] = \Delta_a(t) + \sum_{k=-\infty}^{+\infty} V_k \exp\{j2\pi k f_s[t + \Delta_t(t)]\}, \tag{353}$$

where $v_o(t)=\Sigma V_k \exp(j2\pi k f_s t)$ is the un-perturbed (ideal) signal of the oscillator with fundamental frequency of oscillation f_s, and also having harmonics with amplitudes V_k, which are the Fourier coefficients of v_o. The noise perturbation is usually assumed in Δ_a as an amplitude perturbation, and the response of the oscillator converts it into jitter Δ_t or, equivalently, in phase deviations $2\pi k\Delta_t$. Thus, the phase deviation $\Delta_\varphi=2\pi f_s\Delta_t$ around fundamental harmonic (k=1) is the phase noise modulation in eq. (338), and eq. (353) also implies that there is phase noise modulation around all harmonics of the oscillator, which, indeed, are k-times larger than the phase noise modulation Δ_φ of the fundamental harmonic, where k is the number of the harmonic. Among the several phase noise theories that use perturbation, we discuss two, since the mathematical derivations are lengthy, while the results from the different analyses appear to converge each to other. A list of publications that deal with phase noise theory is provided at the end of this sub-section.

First approach. One approach to perturbation phase noise analysis is to look at the differential equations that describe the oscillator, by adding perturbation in the equations, and obtain insights for the behavior of the oscillator signals in time and frequency domains. Such approach was taken in [306], the comprehensive derivations are published in [307] and [308] for white noise and colored noise (1/f noise and band-limited Lorentzian noise), respectively, and the approach was summarized in [309], as follows.

$$PhN(\Delta f, kf_s) \approx \frac{S_{SSB}}{V_k^2} = \frac{(kf_s)^2\left(c_w + \sum c_f S_f(\Delta f)\right)}{\left[\pi(kf_s)^2\left(c_w + \sum c_f S_f(\Delta f)\right)\right]^2 + (\Delta f)^2}, \tag{354}$$

where k(=1, 2, 3…) is the number of the harmonic of interest with frequency (kf_s), c_w is a parameter that reflects

the contribution of white noise sources in the oscillator, the sum is over noise sources in the oscillator, c_f are parameters that weight the contribution from colored noise sources, the latter having power spectrum densities S_f, and $c_f S_f$ is normalized noise, that is, if S_f is in unit A^2/Hz, then c_f is in unit $1/A^2$. The expression in the square brackets is small, e.g. less than 10Hz, and it causes the phase noise spectrum to level-off at small frequency deviations. The parameter c_w can be determined from the circuit analysis both in time and frequency domains [307], but the calculation involves finding solution of the equations based on the oscillator circuit equations, which are also differential and non-linear in principle. Nevertheless, at particular biasing and in particular circuit, c_w has a single value, and it is shown in [307] that c_w is also related to the variance of the jitter, or

$$c_w = \frac{\sigma^2_{\Delta t,w}}{t_m}, \tag{355}$$

where t_m is the duration of observation (measurement) of the jitter, and $\sigma_{\Delta t,w}$ is the standard deviation of the jitter Δ_t in eq. (353) during the observation and due to white noise. One can assume only the white noise causing the jitter, that is $\sigma_{\Delta t,w} \approx \sigma_{\Delta t}$, by neglecting other components in $\sigma_{\Delta t}$, resulting from 1/f or Lorentzian noise. Then, if the jitter measurement is at the n^{th} period of the clock of a digital signal (after slope triggering of an oscilloscope by the clock signal, for example), then $t_m = nT_s = n/f_s$ and from histogram of the time of the signal transitions around nT_s, one can calculate $\sigma_{\Delta t}$ and estimate c_w from eq. (355). Then, one can measure the phase noise spectrum and from the region with $1/(\Delta f)^2$ slope in the spectrum, to obtain other estimate for $c_w = PhN \times (\Delta f/fs)^2$, according to eq. (354), and verify whether the contribution of white noise is the dominant in the jitter, which would be true, if the values for c_w obtained by both measurements are close. One note should be made here, that Gaussian white noise was used in the deviations of eqs. (354) and (355), while in the practice, the jitter may have pattern dependent component, which might be not Gaussian. Thus, a simple pattern should be used in jitter measurement, but not pseudo random sequence. For the clock signals, one useful relation between cycle-to-cycle jitter and phase noise, which has been derived in a simple manner and verified experimentally in [310], is

$$\sigma^2_{\Delta t,w}\left(t_m = 1/f_s\right) = \frac{(\Delta f)^2}{f_s^3}\,\mathrm{PhN}(\Delta f, f_s), \text{ cycle-to-cycle jitter.} \tag{356}$$

The expression for cycle-to-cycle jitter follows directly from eqs. (354) and (355) when setting observation (measurement) interval $t_m = 1/f_s$ reciprocal to of the oscillation frequency f_s and consider the phase noise in the fundamental harmonic (k=1) at dominance of c_w in the nominator and dominance of $(\Delta f)^2$ in the denominator of eq. (354).

Apart from the complications to calculate the parameters c_f from differential equations of the oscillator, another problem in eq. (354) is a singularity that occurs at very small offsets $\Delta f \to 0$, if the colored noise $S_f = K/f$ is 1/f noise. In such case, eq. (354) reduces to $\Delta f/[(\pi k f s)^2 c_f K] \propto \Delta f \to 0$, and the phase noise reduces, instead to increase or level off, when the frequency offset Δf decreases. Such reduction was never observed for phase noise in the practice of free running oscillators. (The reduction is evident in fractional PLL with sigma-delta modulator.) To remedy, a low-frequency corner f_{cf} for the spectrum of 1/f noise was introduced in [308], below which frequency S_f is constant. Thus, the 1/f noise at very low-frequency $\Delta f \to 0$Hz and DC was limited to $S_f(0Hz) = 4/f_{cf}$ in perturbation theory, by terming the flicker noise spectrum as

$$(S_f =) S_{BL}(f) = 4\int_{f_{cf}}^{\infty} \frac{dx}{x^2 + (2\pi f)^2}$$

$$= \frac{1}{|f|} - \frac{4}{2\pi f}\,\mathrm{artg}\left(\frac{f_{cf}}{2\pi f}\right) \quad \text{, spectrum of 1/f noise limited to low frequency } f_{cf}. \quad (357)$$

$$\approx \frac{1}{f} - \frac{4}{2\pi f}\left(\frac{\pi}{2} + \frac{2\pi f}{f_{cf}} + \ldots\right) = \frac{4}{f_{cf}}, f_{cf} > f \to 0$$

The autocorrelation function R_f=$R_{1/f}$ and the variance σ_f =$\sigma_{1/f}$ (related to the jitter) of 1/f noise with low-frequency corner f_{cf}, respectively, are [308]

$$(R_f =) R_{1/f}(t) = 2\mathrm{EI}(f_{cf}|t|), \quad \text{autocorrelation function of 1/f noise,} \quad (358)$$

$$\sigma^2_{1/f}(t) = 2\frac{2f_{cf}t - 1 + \exp(-f_{cf}t)(1 - f_{cf}t) + (f_{cf}t)^2\,\mathrm{EI}(f_{cf}t)}{f_{cf}^2}, \text{ variance of 1/f noise,} \quad (359)$$

with the exponential integral EI(x) being

$$\mathrm{EI}(x) = \int_1^{\infty} \frac{\exp(-xz)}{z} dz\,. \quad (360)$$

For comparison, for RTS, burst and GR noise, which have low-frequency band-limited normalized Lorentzian spectrum of

$$(S_f =) S_{BL}(f) = \frac{1}{1 + \left(2\pi\dfrac{f}{f_{cf}}\right)^2}, \text{ Lorentzian spectrum} \quad (361)$$

the autocorrelation function and the variance are [308]

$$(R_f =) R_{BL}(t) = \frac{f_{cf}}{2}\exp(-f_{cf}|t|), \text{ autocorrelation function of noise with Lorentzian spectrum,} \quad (362)$$

$$\sigma^2_{BL}(t) = \frac{f_{cf}t - 1 + \exp(-f_{cf}t)}{f_{cf}}, \text{ variance of noise with Lorentzian spectrum,} \quad (363)$$

and the white noise has jitter variance $(\sigma_{\Delta t,wh})^2$=$c_w t$, as follows from eq. (355) given earlier. Note that variances are functions of the time, and the complexity of the functions increases when the type of the noise changes from white noise, through Lorentzian noise, to flicker (1/f) noise.

In addition, the observation of phase noise or jitter is never for infinite time, but for the time window t_m of the measurement, that is for the acquisition time by spectrum analyzers or oscilloscopes. The time window t_m modifies the results for variance, and at high time and frequency resolution, so that both frequency and time being assumed continuous, it was shown in [308] that the variance $\sigma_{\Delta t}$ for the jitter Δ_t, see again eq. (353), during the observation time t_m can be calculated by

$$\sigma_{\Delta t}^2(t_m) = c_w t_m + \sum 2c_f \int_0^{t_m} (t-\tau) R_f(\tau) d\tau , \tag{364}$$

using eqs. (358) and (362) for autocorrelation functions of 1/f noise and Lorentzian noise, respectively, by performing the calculation in time domain, or equivalently by calculation in frequency domain

$$\sigma_{\Delta t}^2(t_m) = c_w t_m + \sum 2c_f \int_{-\infty}^{+\infty} S_f(f) \frac{1-\exp(j2\pi f t_m)}{(2\pi f)^2} df , \tag{365}$$

using eqs. (357) and (361) for power spectrum densities of 1/f noise and Lorentzian noise. The summation in the last two equations is along noise sources in the oscillator circuit. To meet the assumptions for Gaussian random variables, which is used in the derivations, t_m has to be in large enough steps Δt_m, so that the results for $\sigma_{\Delta t,i}=\sigma_{\Delta t}(i\times\Delta t_m)$ and $\sigma_{\Delta t,k}=\sigma_{\Delta t}(k\times\Delta t_m)$ by any values of i and k must satisfy the condition

$$\left.\begin{aligned} &\exp\left[-2\pi^2 f_s^2 (i-k)^2 \sigma_{\Delta t,i}^2\right] \\ &\exp\left[-2\pi^2 f_s^2 (i-k)^2 \sigma_{\Delta t,k}^2\right] \end{aligned}\right\} \approx 0, \text{if } i \neq k \text{ , when using eqs. (364) and (365).} \tag{366}$$

Second approach. Noticeably, the use of the equations above meets with difficulties in the practice, they might be mapped in numerical methods, but then, they become vulnerable for quantization errors. Therefore, other approach for building of solvers for phase noise and jitter was used in [309]. The approach is based at macro model for the relation phase-frequency, in particular, the phase deviations are integrated frequency deviations. That is, the macro model is an ideal integrator in time domain. The circuit noise sources are weighted and summed into power spectrum density S_ω of a macro noise source at the input of the integrator, and the output of the integrator is the phase deviation Δ_φ or jitter $\Delta_t=\Delta_\varphi/(2\pi k f_s)$ of the oscillator, where k is the harmonic number and f_s is the frequency of oscillation. The weighting is with the parameters s_w and c_f are according to eq. (354), and the macro noise source at the input of the integrator is

$$S_\omega(\Delta f) = c_w + \sum c_f S_f(|\Delta f|). \tag{367}$$

Note that S_ω is a normalized power spectrum density in unit 1/Hz, as mentioned after eq. (354).

In frequency domain, the transfer function of the integrator is $H_I=1/(j2\pi\Delta f)$. The observation (or measurement) is practically for finite time t_m, and then, it is repeated, if desired. In other words, the integration captures the evolution of the increment of the phase and its variations (e.g. jitter or phase noise) only for time t_m. Then, the integrator is "zeroed" before the next observation. This is exactly what happens by triggering the measurement in real instruments.

Let us neglect the pause between single measurements. Every single measurement will give the increment of the phase and its variation between current and delayed with t_m integrations. In such situation, the delay-difference operator in frequency domain is $H_{t_m}=1-\exp(-j2\pi\Delta f\times t_m)$, and H_{t_m} is a transfer function that multiplies the transfer function H_I of the integrator. In this way, the observation time window is introduced in [309], and the power spectrum density S_φ of the phase at the output of the integrator for observation time t_m becomes, as

$$S_\varphi(\Delta f) = |H_I|^2 \left|H_{t_m}\right|^2 S_\omega(\Delta f) = \frac{\sin^2(\pi\,\Delta f\; t_m)}{(\pi\,\Delta f)^2} S_\omega(\Delta f). \tag{368}$$

Interestingly, the mathematical problem of singularity at $\Delta f \to 0$Hz is resolved, once $S_\omega(0\text{Hz})$ is finite, since $\lim_{x\to 0}\left[\frac{\sin(x t_m)}{x}\right] = t_m$. Thus, at finite time of measurement t_m, the power spectrum density S_φ of the phase at the origin $\Delta f \to 0$Hz will be observed as

$$S_\varphi(0\text{Hz}) = t_m^2 S_\omega(0\text{Hz}), \tag{369}$$

which is a version scaled by square of the duration of the observation of the power spectrum density of the macro noise source S_ω at the origin $\Delta f \to 0$Hz, where S_ω causes the phase noise in the oscillator, but S_φ is not an integrated value of S_ω, thus the integration function in the macro model is "lost" when the product $(\Delta f \times t_m)$ is small, although mathematically everything is perfect. Similarly, there is no mathematical problem to obtain the variance in the phase from

$$\begin{aligned}\sigma^2_{\Delta\varphi}(t_m) &= \int_{-\infty}^{+\infty} S_\varphi(\Delta f)\, d(\Delta f) \\ &= 2\int_0^{+\infty} \frac{\sin^2(\pi\,\Delta f\; t_m)}{(\pi\,\Delta f)^2} S_\omega(\Delta f) d(\Delta f), \text{since the noise spectra are even functions of frequency,} \\ &= 2t_m^2 S_\omega(0Hz)\Delta f_{\min} + 2\int_{\Delta f_{\min}}^{+\infty} \frac{\sin^2(\pi\, f\, t_m)}{(\pi\, f)^2} S_\omega(f) d(f)\end{aligned} \tag{370}$$

The last integral can be calculated to a finite value either by analytical or numerical method, by taking appropriate minimum offset frequency Δf_{min}, which in the practice is set by the frequency resolution of the spectrum analyzer or by t_m in the case of jitter measurement by oscilloscope, so that $[\sin(\pi\Delta f_{min} t_m)/(\pi\Delta f_{min} t_m)]^2 \geq (1-\text{error})$. However, note again that the phase-frequency relation (of integration) is lost at low frequency, and at $\Delta f=0$, it is replaced with a term that scales the "frequency" noise S_ω by the product $(t_m \times t_m \times \Delta f_{min})$ of measurement settings, instead of having integration of S_ω. It is worth mentioning, without expanding a discussion on measurement uncertainty and errors, that $\Delta f_{min} t_m \approx N$, where N is the number of spectral lines resolved by the spectrum analyzers that use discrete Fourier transformation, whereas in digital oscilloscopes, t_m is the time from triggering and $\Delta f_{min} = f_{samp}/N$, where N is the number of points acquired by the oscilloscope at the sampling rate f_{samp} of ADC, but Δf_{min} does not have direct relation with the time scale on the screen or with the time resolution or bandwidth of the oscilloscope. There are important issues related to the above consequences from finite time observation of phase noise, and one has to carefully take into account the consequences when analyzing experimental data and comparing to simulations. The simulators preserve the calculation consistence in circuits, trying to provide a result that is independent from the length of the data window, whereas the measurements are affected by the time window, because of limited accuracy and resolution that also depend on measurement settings. Simply, the spectrum analyzers and oscilloscopes do not "know" that they measure phase noise, and certainly, the instruments do not make corrections to remedy artifacts from measurements, unless the user programs the instrument to do post processing in specific way.

So, when using the integrator with delay-difference operation, as shown in [309], one obtains from eq. (370) the

following expressions for the variance of the jitter

$$\sigma_{\Delta t}^2(kf_s, t_m) = \frac{\sigma_{\Delta\varphi}^2}{(2\pi)^2} = \frac{\int_{-\infty}^{+\infty} S_\varphi(\Delta f)\, d(\Delta f)}{(2\pi)^2} = \frac{2\int_0^{+\infty} \frac{\sin^2(\pi \Delta f\, t_m)}{(\pi \Delta f)^2} S_\omega(\Delta f) d(\Delta f)}{(2\pi)^2}$$

$$= (kf_s)^2 c_w t_m, \text{ for white noise, and} \qquad (371)$$

$$\approx (kf_s)^2 c_f t_m^2 [0.8456 - 2\ln(f_{cf} t_m)], \text{ for } 1/f \text{ noise with low frequency corner } f_{cf},$$

where kf_s is the frequency of the k^{th} harmonic of the oscillator, f_{cf} is the low-frequency corner for the spectrum of 1/f noise, below which frequency $S_f(0Hz)=4/f_{cf}$ is constant, and 0.8456 is twice the difference between one and Euler's constant, the latter ≈0.5772. The expression in square brackets is truncated for high-order terms [309], it is a rough approximation, and for practical purposes one can take the expression in square brackets equal to one. Observe that the standard deviation $\sigma_{\Delta t}$ of the jitter Δ_t is linear function of the measurement time t_m, when the white noise sources dominate, while $\sigma_{\Delta t}$ becomes nearly quadratic function of the measurement time t_m, when the 1/f noise sources dominate. The parameters c_w and c_f are as defined by eq. (354), which describes the phase noise spectrum.

The variances in eq. (371) correspond to the so-called Allan variance, which is defined in [311] as frequency variance $(\sigma_{\Delta f})^2$ between delayed observations (measurements) with given durations t_m. A simple comparison between eq. (371) and eq. (10) in [311] establishes that

$$\frac{\sigma_{\Delta t}^2(kf_s, t_m)}{\sigma_{\Delta f}^2(kf_s, t_m)} \propto t_m^2. \qquad (372)$$

The analysis of the delayed observations in [311] has adopted the approach for incremental evaluation of variance by delay-difference of observations, as introduced in [312]. The approach in [312], in principle, is the same as the commented above, although the differences were taken algebraically in [312] for statistical variables of averages, rather than by converting the integration and delay operators to transfer functions in frequency domain, as explained above, using Fourier transformation. Other works that make use of the Allan variance and other variance definitions for analysis of the phase noise in oscillators or for phase noise with different slopes of the spectrum are [313] and [314].

One issue, which should be considered by measuring of the phase noise spectra of stable oscillators by spectrum analyzers with discrete Fourier transformation, is that the acquisition time t_m of the spectrum analyzer may become a multiple of oscillator period $1/f_s$. In such cases, the $\sin^2(x)$ term in eq. (368) causes modification of the original phase noise spectrum of eq. (354), and peaks and notches or slow variations in spectra can be observed. Proper windowing and averaging should be used in these cases. Worth mentioning, the parameters c_f in eq. (354) to (371) are related to 0^{th} harmonics V_o, that is, to DC components of the solution of the stochastic differential equations of the oscillator, $c_f=2|V_o|^2$, as explained in [308], which is the same as what was noted after eq. (352) earlier for the up-conversion of low-frequency noise when harmonic balance method for analysis of phase noise is used. Thus, the cyclostationary and perturbation approaches converge in their results, in principle, although some discrepancies in values and computational complexity and time can be found in practice. The convergence also allows to simplify the assumptions in perturbation approach, and to relax the computational complexity, by avoiding the computation of stochastic differential equations. One perturbation approach that illustrates the

avoidance of computation of stochastic differential equations is now summarized.

Third approach: Impulse sensitivity function. The perturbation approach, which avoids calculation of stochastic equations and not necessary requires a calculation of contribution of harmonics in frequency domain, is introduced in [315]. The approach is based on determination of the impulse response of the phase of the oscillator when a perturbation impulse with very short duration (that is, a Dirac function $\delta(t)$ disturbance) is applied at the place of noise source. The impulse response of the oscillator phase to the Dirac function-like disturbance $A_\delta\delta(t)$ is termed as impulse sensitivity function (ISF). It is noted in [315], that the disturbance $A_\delta\delta(t)$ results both in instantaneous amplitude Δ_a and excess phase Δ_φ of the oscillator, see again eq. (338), and the response of the oscillator, including ISF, is time variant and different for different phases within a period of oscillation. However, the response is nearly cyclostationary, that is periodic with the period $1/f_s$ of the oscillation, as long as the perturbation $A_\delta\delta(t)$ is not very strong, as compared to the signals in the oscillator, an assumption that is obvious for noise. (Mathematically, in other words, the derivatives of the of the differential equations that describe the oscillator can be linearized, which corresponds to a use of Jacobian matrices in [307]). Thus, the non-linear perturbation differential equations can be split in two parts, one for amplitude impulse response $h_a(t)$ and, second, for phase impulse response $h_\varphi(t)$, neglecting the combined amplitude-phase impulse response, which corresponds to second derivatives in the non-linear perturbation differential equations. To have the determination of ISF more practical, it was taken in [315], that the disturbance in electronic circuits is a disturbance in electrical current, which affects the charge, and thus the voltage of the capacitance C_{tot} associated with node of the circuit, where the current is applied (having the other node of the current source connected to ground, for convenience). That is $A_\delta\delta(t)=I_\delta\delta(t)=C_{tot}\Delta V=\Delta Q$, where ΔV and ΔQ are the instantaneous changes of the voltage across and the charge in the capacitance C_{tot}. By denoting with $Q_{max}=C_{tot}|V_{max}|$ the magnitude of the maximum charge that can occur in C_{tot} during one period of oscillation $T=1/f_s$, and by t_δ the exact position of the disturbance within the period T, $0\le t_\delta\le T=1/f_s$, then the phase increment due to the disturbance is

$$\begin{aligned}\Delta_\varphi(t-t_o) &= \int_{t_o}^{t} h_\varphi(t-\tau)i(\tau)d\tau = \int_{t_o}^{t} \frac{ISF[rem(t-\tau,T)]}{Q_{max}}i(\tau)d\tau \\ &= \int_{t_o}^{t} \frac{ISF(2\pi f_s t_R)}{Q_{max}}i(\tau)d\tau,\end{aligned} \tag{373}$$

where the function $rem(x/T)=x-T\times integer(x/T)$ is Remainder after division function, which is the difference between x and the last full period nT, and takes into account the periodic and cyclostationary behavior of the phase sensitivity to disturbances. Therefore, the remainder time $t_R(t)=rem(t,nT)$ corresponds to the time t_δ of the disturbance. So, the definition of impulse sensitivity function (ISF) becomes as

$$ISF(2\pi f_s t_R) = \frac{Q_{max}h_\varphi(t,t_R)}{1(t-t_R)} = Q_{max}h_\varphi(t,t_R), \tag{374}$$

since after the moment t_δ of the impulse disturbance with Dirac impulse, the unity step function $1(t-t_R)=1$ for $t\ge t_R\equiv t_\delta$. For the dual case of a voltage noise source in series with an inductor, Q_{max} should be replaced with $\Phi_{max}=L\times I_{max}$, where Φ_{max} represents the maximum magnetic flux swing in the inductor by swing I_{max} for the current in the inductor of inductance L.

As discussed in more details in [315], ISF can be relatively easily obtained from transient simulations by injecting impulse at different relative phases of the oscillation waveform and the oscillator is simulated for a few cycles afterwards. By sweeping the impulse injection time t_δ across one cycle of the waveform, $t_\delta \equiv t_R$, and recording the resulting time shift $\Delta t(t_R)$, h_φ can be calculated as function of t_R noting that $\Delta_\varphi = 2\pi f_s \Delta t$. There also other methods given in [315], based on phase (state) variables by $\varphi_R = 2\pi f_s t_R$, which use the derivatives $\partial v_i / \partial(\varphi_R)$ of the normalized nodal voltages $v = V_i(\varphi_R)/|V_a|$ at each node (i) in the circuit. As the first approximation,

$$ISF_i(2\pi f_s t_R) = \frac{\dfrac{\partial v_i(2\pi f_s t_R)}{\partial(2\pi f_s t_R)}}{\sum_j \left[\dfrac{\partial v_j(2\pi f_s t_R)}{\partial(2\pi f_s t_R)}\right]^2} \approx \frac{\dfrac{\partial v_i(\varphi_R)}{\partial \varphi_R}}{\max\left(\dfrac{\partial v_j}{\partial \varphi_R}\right)^2}, \text{ for } i^{th} \text{ circuit node and all j circuit nodes.} \quad (375)$$

Having ISF for one period of oscillation, one can use the phase variable $\varphi_R = 2\pi f_s t_R$, and calculates the DC (average) value ISF_{dc} and the root-mean-square (effective) ISF_{rms} from

$$ISF_{dc} = \frac{1}{2\pi}\int_0^{2\pi} ISF(\varphi_R)\, d\varphi_R = \frac{c_o}{2}, \quad (376)$$

and

$$ISF_{rms}^2 = \frac{1}{2\pi}\int_0^{2\pi} [ISF(\varphi_R)]^2\, d\varphi_R = \frac{1}{2}\sum_{k=0}^{\infty} c_k^2, \quad (377)$$

where c_k are the Fourier coefficients of ISF, and the later equation is the Parseval relation.

It is shown in [315], that the phase noise, as defined by eq. (339), due to white noise sources with power spectrum density S_{white} of the current white noise is then

$$PhN_{1/\Delta f^2}(\Delta f) = \frac{S_{SSB}(\Delta f)}{P_{carrier}} = \frac{ISF_{rms}^2}{2Q_{max}^2}\frac{S_{white}}{(2\pi\Delta f)^2} = \frac{\sum_{k=0}^{\infty} c_k^2}{4Q_{max}^2}\frac{S_{white}}{(2\pi\Delta f)^2}, \text{ with } S_{white} \text{ in units } A^2/Hz, \quad (378)$$

while the phase noise due to up-converted flicker low-frequency current noise sources with power spectrum density $S_{1/f} \propto 1/f$ depends only on ISF_{dc}, and the corner frequency $\Delta f_{c/\Delta f^3}$ between $1/\Delta f^2$ and $1/\Delta f^3$ components in the phase noise spectrum is reduced, as compared to the corner frequency f_c between 1/f and white noise, and these corner frequencies are related by

$$\frac{\Delta f_{c/\Delta f^3}}{f_c} = \frac{ISF_{dc}^2}{ISF_{rms}^2} = \frac{1}{2}\frac{c_o^2}{c_o^2 + \sum_{k=1}^{\infty} c_k^2} \approx \frac{1}{2}\left(\frac{c_o}{c_1}\right)^2 \le 0.5. \quad (379)$$

Thus, combining eqs. (378) and (379) one gets that

$$\begin{aligned}\mathrm{PhN}(\Delta f)&=\frac{S_{SSB}(\Delta f)}{P_{carrier}}=\mathrm{PhN}_{1/\Delta f^2}(\Delta f)\left(1+\frac{\Delta f_{c/\Delta f^3}}{\Delta f}\right)\\&=\frac{ISF_{rms}^2}{2Q_{max}^2}\frac{S_{white}}{(2\pi\Delta f)^2}\left(1+\frac{ISF_{dc}^2}{ISF_{rms}^2}\frac{f_c}{\Delta f}\right),\end{aligned}\tag{380}$$

which is another way to write the Leeson's formula of eq. (351) for near-carrier phase noise, by using other quantities, however.

From the last equation, also provided that $S_{white}\times f_c=S_{1/f}(1\mathrm{Hz})$ is the flicker low-frequency noise at frequency 1Hz, see again the discussion in section IV.4.5, then one gets a second useful form for phase noise in terms of ISF

$$\begin{aligned}\mathrm{PhN}(\Delta f)&=\frac{ISF_{rms}^2}{2Q_{max}^2}\frac{S_{white}}{(2\pi\Delta f)^2}+\frac{ISF_{dc}^2}{2Q_{max}^2}\frac{S_{white}2\pi f_c}{(2\pi\Delta f)^3}\\&=\frac{ISF_{rms}^2}{8\pi^2Q_{max}^2}\frac{S_{white}}{(\Delta f)^2}+\frac{ISF_{dc}^2}{8\pi^2Q_{max}^2}\frac{S_{1/f}(1\mathrm{Hz})}{(\Delta f)^3},\end{aligned}\tag{381}$$

showing that the phase noise follows the evolution of the white and flicker noise in the circuit components, for example, with the bias, although in different proportions, as compared to low frequency, since $ISF_{dc}<ISF_{rms}$, as mentioned above.

Next interesting arrangement of the last equation, which can be used to separate the up-conversion of low-frequency noise from cyclostationary noise high frequency noise sources, is

$$\begin{aligned}\mathrm{PhN}(\Delta f)&=\frac{\left(ISF_{rms}^2-ISF_{dc}^2\right)}{8\pi^2Q_{max}^2}\frac{S_{white}(\mathrm{RF})}{(\Delta f)^2}+\frac{ISF_{dc}^2}{8\pi^2Q_{max}^2}\frac{S_{white}(1\mathrm{Hz})}{(\Delta f)^2}+\frac{ISF_{dc}^2}{8\pi^2Q_{max}^2}\frac{S_{1/f}(1\mathrm{Hz})}{(\Delta f)^3}\\&=\frac{c_o^2}{32\pi^2Q_{max}^2(\Delta f)^2}S_{LF}(\Delta f),\ \text{from up-conversion of low frequency noise}\ S_{LF}=S_{white}+S_{1/f}(\Delta f),\\&+\frac{\left(c_o^2+2\sum_{k=1}^{\infty}c_k^2\right)}{32\pi^2Q_{max}^2(\Delta f)^2}\sum_{k=1}^{\infty}S_{RF}(kf_s),\ \text{from high frequency noise sources at oscillator harmonics}\ kf_s.\end{aligned}\tag{382}$$

The high frequency noise sources S_{RF} are sources of white noise, in principle, with an exception, the gate induced noise in MOS transistors, which increases with frequency. They have stationary components, given by the quiescent (average) operation point, and also cyclostationary components that follow the variations around the quiescent point. It is clear from eq. (382) that the up-conversion of low-frequency (or base-band) noise is only a portion of the total phase noise, as mentioned by the discussion of harmonic balance method earlier. Interestingly, if re-arranging eq. (382), one will obtain an equation in the form of eq. (352) for harmonic balance analysis of phase noise, demonstrating convergence between the ISF-based perturbation method and harmonic balance method.

Considering the different terms in eq. (382), the conversion of circuit noise into phase fluctuation and side-band noise around oscillator harmonics is illustrated in Figure 60. The top spectrum S_{NOISE} is the noise from different

noise sources connected to one node of the oscillator circuit, e.g. sum of noise currents of pMOS and nMOS transistors, which drain terminals are connected together in each inverter in CMOS ring oscillator. The low-frequency portion of S_{NOISE} , which corresponds to S_{LF} in eq. (352) and has white and 1/f components, causes phase fluctuation S_{PHASE} that is proportional to $(c_o)^2$, and respectively, S_{PHASE} due to S_{LF} has 1/f and white noise components. The stationary and cyclostationary high frequency portions of S_{NOISE} around oscillator harmonics kf_s (which correspond in eq. (352) to white noise S_{RF}) cause phase fluctuations S_{PHASE} that are proportional to $(c_o)^2$ and $(c_k)^2$, respectively. In the figure, only (c_k) are shown for clarity. (To say bluntly, this is down-conversion of RF amplitude noise into LF phase noise S_{PHASE}; the term usually used is "folding of noise spectra" [316], in analogy to what happens when sampling signals with frequency spectrum higher than the Nyquist frequency. Actually, the process is similar.) Since the high frequency portions of S_{NOISE} are with white spectra, then the noise in S_{PHASE} due to S_{RF} is also white. Thus, S_{PHASE} "collects" white noise from low and high frequencies via all (c_k), whereas, the 1/f noise in S_{PHASE} "comes only once" from low-frequency noise S_{LF} via (c_o). Therefore, the ratio 1/f noise to white noise is decreased in S_{PHASE}, as compared to S_{NOISE}, and the corner frequency between 1/f and white noise decreases in S_{PHASE}, again as compared to the corner frequency in S_{NOISE}. Note that the decrease of the corner frequency is due to increase of the white noise contribution in S_{PHASE}, rather than due to decrease of the contribution of 1/f noise, which one may mistakenly suggest. The bottom spectrum in Figure 60 illustrates the conversion of S_{PHASE} in the spectrum of the oscillator S_{OSC}. As follows from the general relation $2\pi(df)=\partial(d\varphi)/\partial t$ between phase φ and frequency f, see between eqs. (346) and (347), each component in S_{OSC} is obtained from the corresponding component in S_{PHASE} by multiplying with $1/(2\pi\Delta f)^2$, resulting in side-band phase noise lobes in the spectrum of the oscillator. This is illustrated in the figure by arrows for the fundamental harmonic f_s of the oscillator (the side-band lobes around f_s in S_{OSC} are shaded for clarity), and the same happens around higher harmonics, scaled by the amplitude of the particular harmonic.

To take into account the cyclostationary components without making harmonic balance, it is suggested in [315] that one can modify $ISF(2\pi f_s t_R)$ defined by eq. (374) and calculated from eq. (375), for example, with the normalized evolution $\alpha(2\pi f_s t_R)$ of cyclostationary noise (in respect to stationary noise) with t_R, $0\le t_R\le T$, within the period T of oscillation, $T=1/f_s$, obtaining effective impulse sensitive function $effISF(2\pi f_s t_R)$, given by

$$effISF(2\pi f_s t_R) = ISF(2\pi f_s t_R)\times\alpha(2\pi f_s t_R),$$

$$\text{with } \alpha(2\pi f_s t_R) \approx \sqrt{\frac{S_{NOISE}(2\pi f_s t_R)}{\frac{1}{T}\int_0^T S_{NOISE}(2\pi f_s t)dt}} \text{ or with } \alpha(2\pi f_s t_R) \approx \frac{\sqrt{S_{NOISE}(2\pi f_s t_R)}}{\frac{1}{T}\int_0^T \sqrt{S_{NOISE}(2\pi f_s t)}dt}, \qquad (383)$$

where the quantities in the numerator of α are cyclostationary noise and the quantities in the denominator are stationary noise. The corresponding DC and RMS values of effISF are then obtained from (376) and (377), respectively for $effISF_{dc}$ and $effISF_{rms}$.

The relation between effISF and ISF depends on how the devices operate during oscillation cycle. In ring oscillators, effISF≈ISF, but in Colpitts oscillators, effISF is very different from ISF [315]. The rule is simple. If the peaks in ISF and cyclostationary noise source are in phase, that is, the peaks occur at the same time in the oscillation cycle, then effISF≈ISF. This is the case of ring oscillators, in which the current and its noise are at maximum during the signal transitions between high and low, and also the derivatives of nodal voltages are with maximum absolute values, $|\partial v_i/\partial\varphi_R|$=max by $\Sigma(\partial v_j/\partial\varphi_R)^2\approx$constant in eq .(375), during the signal transitions, and,

thus, from eq. (375), ISF also peaks during transitions. Detailed investigations of phase noise in ring oscillators can be found in [315, 317, 318, 319, 320], from which one can have simple analytical expressions for phase noise as function of number of stages, supply, frequency, transistor sizes, etc., and comparisons to other type of oscillators, and regimes of operation, as well as relations for jitter, that match with the discussion on eqs. (371) and (372) earlier. In contrast to ring oscillators, in LC oscillators, the cyclostationary currents and voltages are not in phase, which results in large differences between ISF and effISF, especially in single transistor oscillators, such as the Colpitts oscillators, in which the cyclostationary transistor current is only for a fraction of the half period of oscillation, and $\max(|\text{effISF}|)=\max(|\text{ISF}|\times\alpha)<\max(|\text{ISF}|)\times\max(\alpha)$.

VII.2.3. Suggestions for the design of oscillators

Several useful considerations in design of oscillators can be deduced from eq. (382).

Minimize the up-conversion of low-frequency noise

The first is that the DC component $\text{ISF}_{dc}=c_o/2$ has to be minimized in order to minimize the up-conversion of low-frequency noise, according to the middle line of eq. (382). This will also reduce the conversion of stationary RF white noise (e.g. thermal noise or shot noise due to quiescent current) by the term $(c_o)^2\sum S_{RF}(kf_s)$ in the bottom line of the equation. As shown in [321], ISF_{dc} is minimized when having particular symmetry in the waveforms of the oscillator, either even periodic waveforms $f(t)=f(-t\pm nT)$, or half-wave symmetric waveforms $f(t)=-f(t+T/2\pm nT)$. From practical point of view, both symmetries put requirements for identical transitions in oscillator waveform. This is depicted in Figure 61, in which the ideal waveforms are given with solid lines, and the distorted waveforms are drawn with black dash-lines, giving a rise of non-zero ISF_{dc}, and thus, of up-conversion of low-frequency noise into phase noise. The even symmetry requires identical transitions by mirroring in time, and the half-wave symmetry requires identical transitions by mirroring in amplitude, the latter also at delay of exactly half period. From spectral point of view, the even symmetry results in harmonic expansion of the oscillator waveform, $\sum\cos(2k\pi f_s t+k\varphi_o)$ with φ_o=constant, that is in-phase or equally delayed harmonics. The half-wave symmetry requires that the oscillator signal is free from even harmonics and the duty cycle of the signal is 50%. As seen from Figure 61, differences in rise and fall times of the circuit signals cause deviation from desired ideal waveforms. The circuits shown with gray color in the figure help to improve the signal symmetry. In particular, the complementary pMOS cross-coupled transistor pair works in anti-phase of the nMOS transistor pair. Therefore when nMOS switches on, pMOS switches off (and vice versa), which provides that one has always a transistor switching on and other transistor switching off during the time of signal transitions, and thus, smaller difference between rise and fall times and better even periodic symmetry, as illustrated by gray dash-lines in the bottom-left figure. Experiments that confirm the effect of reduction of phase noise when using complementary cross-coupled pairs can be found in [322]. The introduction of current limiting ("starving") transistors in series with the main switching transistors in the ring oscillator minimizes the difference between the charging (I_P, from pMOS) and discharging (I_N, to nMOS) currents that flow through the node capacitance during transitions, and, thus, the rates dv/dt during signal rise and fall are equalized, resulting in signal with better half-wave symmetry, as illustrated by the gray dash-line in the bottom-right figure. Obviously, limiting the current, the circuit is slower, and the oscillation frequency in the ring oscillator with current "starving" is expected to decrease, as compared to the initial circuit without current "starving". Experiments and analysis that confirm the effect of reduction of phase noise in ring oscillators when using current limiting ("starving") transistors to adjust the currents of nMOS and pMOS transistors can be found in

[317]. Thus, the symmetry of the up and down transitions between the signal levels is important consideration for reduction of phase noise due to up-converted low-frequency noise [315].

The occurrence of high-order harmonics is not favorable for oscillators

The second useful consideration in design of oscillators, which can be deduced from the bottom line of eq. (382), is that the occurrence of high-order harmonics is not favorable for oscillators with low phase noise. In the presence of harmonics with numbers k=1, 2, 3…K_{max} in oscillation signal, one can estimate the AC component of ISF_{rms} from eq. (375) [315], as follows. Assume that the signal is

$$V(\varphi)=\sum_{k=1}^{K_{max}} V_k \cos(k2\pi f_s t)=\sum_{k=1}^{K_{max}} V_k \cos(k\varphi)=V_1\sum_{k=1}^{K_{max}} \frac{V_k}{V_1}\cos(k\varphi)=V_1 v(\varphi), \text{ with } \varphi=2\pi f_s t, \quad (384)$$

where V_k are the harmonic amplitudes and $v(\varphi)=V(\varphi)/V_1$ is the normalized signal in eq. (375), from which the impulse sensitivity function ISF(φ) is

$$ISF(\varphi)=\frac{\frac{\partial v(\varphi)}{\partial(\varphi)}}{\sum_k\left[\left|\frac{\partial v_k(\varphi)}{\partial(\varphi)}\right|\right]^2}=-\frac{\sum_k k\frac{V_k}{V_1}\sin(k\varphi)}{\sum_k k^2\left(\frac{V_k}{V_1}\right)^2}. \quad (385)$$

To obtain ISF_{rms}, one can transform ISF(φ) in the space of harmonic numbers, using Fourier transformation by making analogy k→kω and φ→t, and then using Parseval relation, see again eq. (377). Assume constant norm of the derivative in the denominator of eq. (385), which seems reasonable for cyclostationary process in LC oscillator. Next equation shows the result, but note that the simple transformation misses important fact of convolution with noise, as it will be corrected afterwards.

$$ISF_{rms}^2(\text{incomplete})=\frac{1}{2}\frac{\sum_k k^2\left(\frac{V_k}{V_1}\right)^2}{\left[\sum_k k^2\left(\frac{V_k}{V_1}\right)^2\right]^2}=\frac{1}{2\sum_k k^2\left(\frac{V_k}{V_1}\right)^2}. \quad (386)$$

When substituting in eq. (378) we obtain the following incomplete result for phase noise

$$\text{incomplete: } PhN_{1/\Delta f^2}(\Delta f)=\frac{S_{SSB}(\Delta f)}{P_{carrier}}=\frac{ISF_{rms}^2}{2Q_{max}^2}\frac{S_{white}}{(2\pi\Delta f)^2}$$
$$=\frac{1}{4Q_{max}^2\left[\sum_{k=1}^{K_{max}} k^2\left(\frac{V_k}{V_1}\right)^2\right]}\frac{S_{white}}{(2\pi\Delta f)^2}, \quad (387)$$

where reminding that the white noise S_{white} is in units A^2/Hz. This incomplete result has wrong behavior when the number and the amplitudes of the harmonics increase. In particular, the incomplete result suggests that the phase noise will decrease when K_{max} and V_k increase, which is the opposite to what one would observe in experiments.

The correct result will be obtained when considering that the white noise at different harmonics is uncorrelated. Therefore ISFk of each harmonic k can be separately estimated from eq. (385), taking $\varphi_k=k\varphi$, $v_k(\varphi_k)=\cos(\varphi_k)$, which is the same as for $K_{max}=1$ in eq. (385), and gives

$$\mathrm{ISFk}(\varphi_k \equiv k\varphi)=\frac{\dfrac{\partial v_k(\varphi_k+\psi_k)}{\partial(\varphi_k)}}{\left|\dfrac{\partial v_k(\varphi_k+\psi_k)}{\partial(\varphi_k)}\right|^2}=-\frac{\sin(\varphi_k+\psi_k)}{\sin^2(\varphi_k+\psi_k)+\cos^2(\varphi_k+\psi_k)}\Rightarrow \mathrm{ISFk}_{rms}^2=\frac{1}{2}, \tag{388}$$

where ψ_k is the initial phase of the k^{th} harmonic at $\varphi=0$. Note that for a cycle $\varphi=0\ldots2\pi$ of the oscillator, φ_k makes k cycles, thus adding k times the noise from k^{th} harmonic, which results in phase displacement $\Delta\psi_k=k^{½}\Delta\varphi_k$ at the end of the oscillator cycle $\varphi=2\pi$. (We use $k^{½}$, because the contributions $\Delta\varphi_k$ are statistically independent, since the white noise is uncorrelated, and the variance $(\sigma_{\Delta t,wh})^2\propto t\propto k$ is proportional to the time t and, thus, number of cycles k, see eqs. (355), (369)–(372) earlier). The phase displacement $\Delta\psi_k$ causes the "pulling" of the white (or other) noise around the k^{th} harmonic in the phase noise, as depicted in Figure 60 earlier, by changing of the state of all other harmonics, including the fundamental one, as illustrated in Figure 62 in simplified form. Provided that the phase "pulling-pushing" between the oscillator harmonics via the change ΔL of the state on the oscillation contour is small, then one can write for the oscillator contour, k^{th} and p^{th} harmonics that

$$\begin{aligned}\Delta L &\approx V_k\sin(\Delta\psi_k)=V_k\sin\left(\sqrt{k}\Delta\varphi_k\right)\approx V_k\sqrt{k}\Delta\varphi_k, \text{ for } k^{th} \text{ harmonic},\\ &\approx V_p\sin(\Delta\psi_p)=V_p\sin\left(\sqrt{p}\Delta\varphi_p\right)\approx V_p\sqrt{p}\Delta\varphi_p, \text{ for } p^{th} \text{ harmonic},\\ &\approx\sqrt{\sum_k V_k^2}\sin(\Delta\varphi)\approx\sqrt{\sum_k V_k^2}\Delta\varphi, \text{ for oscillator contour}.\end{aligned} \tag{389}$$

Rewritten in terms of noise, $\Delta\varphi_k\rightarrow(\mathrm{ISFk}_{rms})^2S_{white}(kf_s)$ and $\Delta\varphi_p\rightarrow S_{\varphi p}$, from the first and second lines of the last equation, one gets

$$S_{\varphi_p,k}\left(\text{from } k^{th} \text{ harmonic}\right)=\frac{kV_k^2}{pV_p^2}\mathrm{ISFk}_{rms}^2S_{white}(kf_s), \tag{390}$$

where $S_{\varphi p,k}$ is the noise in the phase in the p^{th} harmonic caused by the k^{th} harmonic, and $(\mathrm{ISFk}_{rms})^2=½$, as follows from eq. (388). Neglecting the cyclostationary component in the white noise, we can assume that S_{white} is nearly constant within oscillation cycle. It is of particular interest the fundamental harmonic, p=1. By adding the contributions from all harmonics, each given by $S_{\varphi 1,k}$ in eq. (390), and according to eq. (382), the phase noise component in the fundamental harmonic f_s of the oscillator due to RF white noise becomes, as

$$\begin{aligned}\text{complete: } \mathrm{PhN}_{1/\Delta f^2}(\Delta f, \text{at } f_s, \text{from RF noise})&=\frac{\left(\mathrm{ISF}_{rms}^2-\mathrm{ISF}_{dc}^2\right)}{2Q_{max}^2}\frac{S_{white}(RF)}{(2\pi\Delta f)^2}\\ &\approx\frac{S_{white}(RF)}{2Q_{max}^2(2\pi\Delta f)^2}\sum_{k=1}^{K_{max}}k\frac{V_k^2}{V_1^2}\mathrm{ISFk}_{rms}^2=\frac{S_{white}(RF)}{2Q_{max}^2(2\pi\Delta f)^2}\frac{1}{2}\sum_{k=1}^{K_{max}}k\frac{V_k^2}{V_1^2},\end{aligned} \tag{391}$$

where reminding that the white noise S_{white} is in units A^2/Hz. This is the corrected version of the incomplete eq. (387). Consequently, instead of eq. (386), the correct expression for ISF in presence of harmonics in the

oscillation signal is

$$\mathrm{ISF}_{\mathrm{rms}}^2 - \mathrm{ISF}_{\mathrm{dc}}^2 = \mathrm{ISF}_{\mathrm{rms}}^2(\text{complete, at } f_s\text{, for RF noise}) = \frac{1}{2}\sum_{k=1}^{K_{\max}} k\frac{V_k^2}{V_1^2}. \quad (392)$$

Now, both eqs. (391) and (392) demonstrate in a simple manner that the phase noise in oscillators increase when the high-order harmonics occur and increase, while the incomplete eqs. (386) and (387) predicted the opposite incorrectly. The problem in eqs. (386) and (387) is mathematical. In particular, is was neglected the phase variation of the denominator in the preceding eq. (385) and it was assumed constant by Fourier transformation. The correct way of making the transformation is complicated, and it uses characteristic multipliers and characteristic exponents (Floquet theorem) in order to invert and find reciprocal of the state transition matrix of differential equations of the oscillator – details on how to do this in time and frequency domains can be found in [306, 307]. It is noted in [315] that ISF is inversely proportional to the sum of signal derivatives (in time domain) in the oscillator circuit, which consequently, is replicated squared in phase noise. This observation, however, should be not extrapolated in frequency domain, assuming that the larger are the high-order harmonics the higher are the derivatives ISF in time domain. In fact, at high frequency and large signals, the amplifier slew rate limits the speed of the transitions, which causes increase of high-order harmonics, but the maximum values of the derivatives are hardly increased.

Circuits with more than one node. Worth mentioning, eqs. (391) and (392) are for single node circuit calculation. If the circuit is with more nodes, then the calculations has to be performed for each node and the results has to be properly referred to output of the oscillator. Usually, there is circuit symmetry in LC oscillators and circuit repetition in ring oscillators, which allows easily to scale the result for one circuit node to the whole circuit. Furthermore, if the operation of the transistors in the circuit is non-linear, obviously this is the case in oscillators, then the noise sources are cyclostationary and correction for the value of S_{white} will be necessary. Nevertheless, eqs. (391) and (392) are so simple that are appropriate even for approximate manual and qualitative calculations, without using complicated simulators and in cases when the circuit details are not available, by measurements, for example. For such cases, since $(\mathrm{ISF1}_{\mathrm{rms}})^2 = ½$ for the ideal sinusoidal signal, then it might be helpful to use the excess value of ISF that is caused by the occurrence of the harmonics, e.g. as the figure of merit. The excess ISF, $(\mathrm{eISF}_{\mathrm{rms}})^2$, is a difference between $(\mathrm{ISF}_{\mathrm{rms}})^2$ of the signal with harmonics and $(\mathrm{ISF1}_{\mathrm{rms}})^2 = ½$ for the ideal sinusoidal signal; and $\mathrm{eISF}_{\mathrm{rms}}$ is defined as

$$\mathrm{eISF}_{\mathrm{rms}}^2 = \mathrm{ISF}_{\mathrm{rms}}^2 - \mathrm{ISF1}_{\mathrm{rms}}^2 = \mathrm{ISF}_{\mathrm{rms}}^2 - \frac{1}{2} = \frac{1}{2}\sum_{k=2}^{K_{\max}} k\frac{V_k^2}{V_1^2} \approx \frac{1}{2}\left[\frac{\mathrm{PhN}_{1/\Delta f^2}(\text{harmonics})}{\mathrm{PhN}_{1/\Delta f^2}(\text{sinusoidal})} - 1\right], \quad (393)$$

by assuming no change in other parameter, e.g. $Q_{\max}$, Δf and S_{white} are the same for the two signals. To illustrate the contribution of signal harmonics to the phase noise, we plot $(\mathrm{ISF}_{\mathrm{rms}})^2$ and $(\mathrm{eISF}_{\mathrm{rms}})^2$ for several harmonics in Figure 63a versus the harmonic distortion $(\mathrm{HD}_k)^2 = (V_k/V_1)^2$. Observe that higher order harmonics cause stronger increase of phase noise for the same level harmonic distortion. Since the quantity $\mathrm{THD}^2 = \sum(V_k/V_1)^2 - 1$ is the popular total harmonic distortion factor that is used for quantification of the deviation of signals from ideal sinusoidal waveform, we show the dependence of phase noise for two typical waveform distortions, in which the harmonic amplitudes vary as the reciprocal of the harmonic number, $V_k \propto 1/k$. In Figure 63b, only the odd harmonics are present, resulting in a "rectangular" type of distortion. In Figure 63c, all harmonics are present,

which results in “peaking” distortion, if the harmonic co-sinusoids are in phase, or “saw tooth” distortion, is the harmonic sinusoids are in phase.

Observe in all plots of Figure 63 that a harmonic distortion in the range −10dB to −4dB causes 3dB (two times) increase of phase noise. These are high levels of waveform distortion, especially for LC oscillators. However, the transistors that drive the LC tank operate usually in highly non-linear mode, close to regime of switching, and consequently, the cyclostationary noise associated with current pulses replicates the harmonic of the cyclostationary current of the transistors. As explicitly stated in [315], ISF(φ) multiplies the transistor noise in phase (time) domain, which results in convolution in the domain of harmonic numbers after Fourier transformation. Then, even if the oscillation signal is free of harmonics, the product ISF×S_{white} will have harmonics, since $S_{white}(k\varphi)\propto I_{cyclo}(k\varphi)$ repeats the harmonics in the transistor cyclostationary current I_{cyclo}. For example, as follows from eq. (263), for cyclostationary shot noise in the collector of BJT one defines noise generating current i_{cyclo}, given by

$$i_{cyclo}(\varphi)=\sqrt{S_{I_C,cyclo}df}\approx\sqrt{2q(1+\beta_f)df}\sum_{k=1}^{K_{max}}\sqrt{I_k}\exp(j\psi_k-jk\psi_d), \tag{394}$$

where β_f is high frequency current gain, normally $\beta_f\sim f_s/f_T<\beta_{DC}$, I_k are the amplitudes of the cyclostationary collector current, and $\psi_d\approx 2\pi f_s/f_T$ is the phase delay due to transit time of the BJT. In phase (time) domain, one writes for phase noise deviation that

$$\Delta_\varphi(\varphi)=ISF(\varphi)\times i_{cyclo}(\varphi), \tag{395}$$

which is a multiplication, and it results in convolution after Fourier transformation in harmonic domain

$$\begin{aligned}\Delta_\varphi(k)&=\int_{k=0}^{K_{max}}ISF(x)\times i_{cyclo}(k-x)dx\\&=\sqrt{2q(1+\beta_f)df}\sum_{k=1}^{K_{max}}\sqrt{k}\frac{V_k}{V_1}\exp(j\psi_k)\sum_{m=0}^{K_{max}}\sqrt{I_m}\exp(-jm\psi_d)\end{aligned}. \tag{396}$$

where we assume that the DC component in ISF is zero at the waveform conditions discussed earlier. Converted in noise power spectrum density, one gets

$$\begin{aligned}S_\varphi&\approx\left[\frac{1}{2}\sum_{k=1}^{K_{max}}k\frac{V_k}{V_1}\right]2q(1+\beta_f)I_{C,DC}\left\{1+\sum_{m=1}^{K_{max}}\frac{I_m}{I_{C,DC}}\cos^2(m\psi_d)\right\}\\&\approx\left[\frac{1}{2}\sum_{k=1}^{K_{max}}k\frac{V_k}{V_1}\right]2q(1+\beta_f)I_{C,DC}\left\{1+\sum_{m=1}^{K_{max}}\frac{I_m}{I_{C,DC}}\right\}\end{aligned} \tag{397}$$

The term in the square brackets is the same as eq. (392). The term between the square and large brackets is the stationary white noise. The sum in the large brackets is the relative contribution of cyclostationary noise. In the last line we have neglected the delay in cyclostationary current, assuming $f_s<<f_T$, to make the comparison to eqs. (390) and (392) clear that the harmonics I_m in the transistor cyclostationary current contribute on the top of the noise from the stationary current $I_{C,DC}$, and also, to show the origin of the last sum $\sum S_{RF}(kf_s)$ in eq. (382). The precise expression for the cyclostationary noise with $\psi_d\neq 0$ is obtained from harmonic balance method for circuit analyses with computer simulators, and it was shown earlier by eq. (352). To conclude the discussion on the

importance of oscillator harmonics for phase noise we combine the last line of eq. (397) with eq. (391), and obtain the following expression, which shows in explicit generic form the harmonic contributions to oscillator phase noise.

$$\begin{aligned}\mathrm{PhN}_{1/\Delta f^2}(\Delta f, \text{at } f_s, \text{from RF noise}) &= \frac{\left(\mathrm{ISF}_{\mathrm{rms}}^2 - \mathrm{ISF}_{\mathrm{dc}}^2\right)}{2Q_{\max}^2}\frac{S_{\mathrm{white}}(\mathrm{RF})}{(2\pi\Delta f)^2} \\ &\approx \frac{S_{\mathrm{white}}(\mathrm{RF}, \text{stationary current noise at } I_{DC})}{2Q_{\max}^2(2\pi\Delta f)^2}\left\{1+\sum_{k=1}^{K_{\max}}\frac{I_k}{I_{DC}}\cos^2\left(k\frac{f_s}{f_T}\right)\right\}\left[\frac{1}{2}\sum_{k=1}^{K_{\max}} k^{(1+\mathrm{corr})}\frac{V_k^2}{V_1^2}\right].\end{aligned} \tag{398}$$

Here, $Q_{max}=V_{max}C_{tot}$ is the swing of the charge during oscillation by voltage swing $V_{max}=V_{pk-pk}$ across the node capacitance C_{tot}, presumably the capacitance is the same as the capacitance of oscillator LC tank; Δf is the frequency offset from fundamental harmonic f_s of oscillation, $k=1\ldots K_{max}$ are harmonic numbers; V_k are the harmonic amplitudes of the voltage across the LC tank, the term in the square brackets is impulse sensitivity function (ISF) corresponding to the waveform of the voltage of the oscillator; I_{DC} and I_k are DC and harmonic amplitudes in the current of the transistor, which drives the LC tank with negligible delay $t_d/T\sim f_s/f_T$, thus $\cos^2(x)\approx 1$, the sum in the large brackets is the relative contribution of cyclostationary noise; and S_{white} is stationary white current noise, associated with I_{DC}, and S_{white} is with negligible correlation coefficient $1>\mathrm{corr}\approx 0$ at oscillation harmonics. We are reminding that the white noise S_{white} is in units A^2/Hz. Now, eq. (398) allows to carry out the discussion on the next design consideration for phase noise, the oscillator power.

Optimum magnitude of the oscillation

The third useful consideration in design of oscillators is to find optimum magnitude of the oscillation, so that the phase noise is minimized. From the Leeson formulas, eqs. (350) or (351), it is clear that, if everything else being the same, increasing the power $P_{carrier}$ of the fundamental harmonic signal, the phase noise decreases. Therefore, one would increase the oscillation amplitude V_1 in order to minimize the phase noise PhN. However, the increase of the amplitude will increase the non-linearity in the oscillator circuit, and thus, the high order harmonic content will increase, which will increase the phase noise, as discussed just above, and summarized by eq. (398). Experiments reported in the literature, such as in [297, 323], show that the relation phase noise – oscillator power is non-monotonic, and the relation has a minimum at certain optimum power. Deviations of $P_{carrier}$ from P_{opt} result in increase of phase noise and in v-shaped plots of the relation PhN–$P_{carrier}$, as shown in Figure 64.

Consider the LC oscillator. The quality factor of the LC tank is $Q=2\pi f_s R_s C_{tot}$, where R_s is the loss resistance in parallel to the LC tank. Then, the charge swing $Q_{max}=V_{max}C_{tot}=V_{max}Q/(2\pi f_s R_s)$ and $(V_{max})^2\approx 8(V_{rms})^2=4(V_1)^2(1+\mathrm{THD}^2)=8R_sP_{tot}$, where the total power of the signal $P_{tot}=P_{carrier}(1+\mathrm{THD}^2)$ is the sum of the power of the fundamental harmonic $P_{carrier}$ and the power of high-order harmonics $P_{carrier}\mathrm{THD}^2$, the latter is given by the total harmonic distortion factor THD. Assume that $\mathrm{THD}<0.3$, then $P_{tot}\approx P_{carrier}$. Inserting these relations in eq. (398), we get

$$\mathrm{PhN}_{1/\Delta f^2} \approx \left(\frac{S_{\mathrm{white}}(\mathrm{RF}, \text{stationary})R_s}{8P_{\mathrm{carrier}}}\frac{f_s^2}{2Q^2(\Delta f)^2}\right)\left\{1+\sum_{k=1}^{K_{\max}}\frac{I_k}{I_{DC}}\right\}\left[\frac{1}{2}\sum_{k=1}^{K_{\max}} k\frac{V_k^2}{V_1^2}\right], \tag{399}$$

where $P_{noise}=S_{white}R_s=(4kT/R_s)NF_{wh}$ is the spectral density of the noise power that is applied on R_s from all noise sources connected to the LC tank, e.g. from transistors and R_s itself. Evidently, we have obtained the Leeson

formula for phase noise from white noise, see eq. (351), with the impact of harmonics disclosed. Note, we are assuming that the DC component of ISF is zero for clarity, and one should add $(ISF_{dc})^2$ to the sum in the square brackets, to obtain the complete expression for $PhN_{1/\Delta f^2}$.

In the special case of LC oscillators, I_k can be large, in the order of I_{DC}/k when the transistor is operating close to switching mode, but the LC tank suppresses the high order harmonics as $V_k \approx I_k R_s/(kQ)$ for $k \geq 2$ and $Q>3$. Since $V_1=I_1R_s$, then the terms $k(V_k/V_1)^2 \approx 1/(k^3Q^2)$ are small, and for $k \geq 2$, the sum $\sum k(V_k/V_1)^2 \approx 1/(2k^2Q^2) < 1/72 << 1 = 1(V_1/V_1)^2$ for $k=1$. Thus, one can take value ½ for the expression in the square brackets of eq. (399), and also $P_{carrier} \approx (V_1)^2/(2R_s) = (I_1R_s)^2/(2R_s) = (I_{DC})^2(I_1/I_{DC})^2R_s/2$ for the power of the oscillation signal. Therefore, for practical cases of single transistor LC oscillators, such as Colpitts oscillators, eq. (399) can be reduced and arranged as

$$PhN_{1/\Delta f^2} \approx \frac{f_s^2}{16Q^2(\Delta f)^2}\left(\frac{S_{white}(RF, I_{DC})}{I_{DC}^2}\right)\frac{\left\{1+\sum_{k=1}^{K_{max}}\frac{I_k}{I_{DC}}\right\}}{\left[\frac{I_1}{I_{DC}}\right]^2}, \text{ single transistor LC oscillator.} \quad (400)$$

For the mostly used topology of LC oscillators with cross-coupled pair transistors, see Figure 61a, the DC current of each transistor is a half of the supply current, $I_{DC}=I_{SUP}/2$, and both transistors contribute with their noise, thus $S_{white}=2S_{white}(I_{SUP}/2)$. Eq. (400) becomes as

$$PhN_{1/\Delta f^2} \approx \frac{f_s^2}{2Q^2(\Delta f)^2}\left(\frac{S_{white}(RF, I_{SUP}/2)}{I_{SUP}^2}\right)\frac{\left\{1+\sum_{k=1}^{K_{max}}\frac{2I_k}{I_{SUP}}\right\}}{\left[\frac{2I_1}{I_{SUP}}\right]^2}, \text{ pair-transistor LC oscillator,} \quad (401)$$

where note that the harmonic amplitudes I_k of the currents in each transistor are taken doubled, owing to $I_{DC}=I_{SUP}/2$.

Eqs. (400) and (401) show in convenient form the impact of different factors on $1/\Delta f^2$ phase noise in LC oscillators. First, PhN increases with oscillation frequency as $(f_s)^2$. Second, PhN can be decreased using LC tank of higher quality factor Q. Third, PhN can be decreased, increasing the consumption I_{DC} or I_{SUP} of the circuit, since the term in the round brackets is the normalized white current noise in transistors, which is inversely or nearly inversely proportional to the bias current – see eqs. (248)–(250) for MOS transistors and eq. (263) for bipolar transistors. Fourth, PhN can be minimized by increasing the oscillation amplitude, or precisely, maximizing the use of the bias current I_1/I_{DC}, or $2I_1/I_{SUP}$, as shown by the squared ratio in the square brackets. Fifth, however, maximizing I_1/I_{DC}, one should not exceed $I_1>I_{DC}$ much, because the transistor current will saturate and distortions will occur in the transistor drive current. The harmonics I_k associated with this distortion will increase the sum $\sum I_k/I_{DC}$ in the large brackets, and PhN will increase, consequently.

The overall behavior of eqs. (400) and (401) shows that there is a minimum for phase noise as function of the level of use I_1/I_{DC} of the bias current. This is clearly depicted in Figure 64a for a CMOS oscillator [297], where the minimum phase noise is not at the maximum oscillation voltage (V_1), but at maximum ratio I_1/I_{SUP}.

To illustrate further, we assume "rectangular" distortion of the transistor current. That is, odd harmonics I_k, $k=3,5,\ldots$, occur when increasing the amplitude of the fundamental harmonic I_1 above I_{DC}. We shall assume that

the high-order harmonic power $(I_k)^2/(2R_s)$ is proportional to the power of the fundamental harmonic $[(I_k)^2/(2R_s)]^k$ on exponent k, which is well established from the intermodulation analyses of RF circuits, e.g. IIP3 at which extrapolated power of first and third harmonics are equal. The second assumption, which follows from Fourier series of rectangular waveforms, is that at particular value $I_{1R}=(4/\pi)I_{DC}$ of the fundamental harmonic, the amplitude of the rectangular signal is I_{DC} and the amplitudes of the harmonics are $I_{kR}=I_{1R}/k$. With these assumptions, it can be shown that for a given value I_1, one has

$$\frac{I_k}{I_{DC}}=\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\left(\frac{I_1}{I_{DC}}\right)^k. \tag{402}$$

This is because

$$\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\left(\frac{I_{1R}}{I_{DC}}\right)^k=\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\left(\frac{\frac{4}{\pi}I_{DC}}{I_{DC}}\right)^k=\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\left(\frac{4}{\pi}\right)^{(k-1)}\frac{\frac{4}{\pi}I_{DC}}{I_{DC}}=\frac{I_{1R}}{k}\frac{1}{I_{DC}}=\frac{I_{kR}}{I_{DC}}$$

and

$$\left(\frac{I_k}{I_{DC}}\right)^2=\left[\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\left(\frac{I_1}{I_{DC}}\right)^k\right]^2=\left[\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\right]^2\frac{\left(I_1^2\right)^k}{\left(I_{DC}^2\right)^k},$$

from which it follows that

$$I_k^2=\left[\frac{1}{k}\left(\frac{\pi}{4}\right)^{(k-1)}\left(\frac{I_1}{I_{DC}}\right)^k\right]^2=\left[\frac{1}{k}\left(\frac{\pi}{4I_{DC}}\right)^{(k-1)}\right]^2\left(I_1^2\right)^k\propto\left(I_1^2\right)^k,$$

since I_{DC} is constant.

Using these rations, we have obtained the phase noise from eq. (401), shown with a solid curve in Figure 64b, considering shot noise $2qI_{SUP}/2$ in BJT and quality factor Q=16 and loss resistance 720Ω, as reported for the circuit in [323]. We also put in the figure the data from [323]. Evidently, the calculation with eq. (401) for the minimum phase noise matches well with the experimental data, shown by squares, and again, the minimum phase noise is at maximum ratio I_1/I_{SUP}, just below the severe harmonic distortion in the current begins. In the particular circuit, the automatic amplitude control keeps the oscillator amplitude constant, and the minimum phase noise is observed when the current consumption is minimum, as shown by circles in the figure, thus, the usage of the supply current is at maximum. Worth mentioning, the distortion in the transistor current is accompanied also with increase of the current consumption, and ratio $I_1/I_{DC}>1.2$ (or $I_1/I_{SUP}>0.6$) is practically not accessible by real circuits, since "hyper-rectangular" distortion ($I_1/I_{DC}>4/\pi$) is very unlikely.

VII.2.4. Summary

To summarize, the models for phase noise converge each to other. The simplest Leeson formula, given by eq.(350) or eq. (351), captures the essence, and eqs. (400), (401) enhance it to predict the optimum operation point for minimum phase noise. However, the accuracy of these simple equations is not very high, since the non-linear operation and the phase of harmonics are neglected. The consequence is that the slopes, when deviating from optimum, cannot be calculated accurately, as seen from Figure 64b. The accurate calculations can be

obtained only by circuit simulators, by employing harmonic balance or perturbation analysis. The computer simulators have achieved maturity in phase noise analysis last decade. The analyses showed that the up-conversion of low-frequency noise in phase noise occurs when asymmetry in the transitions in the waveforms are present. The RF noise, however, contributes always, resulting in unavoidable $1/\Delta f^2$ component in the phase noise, which can be minimized only by high quality factor resonators. Important insights for phase noise through experiments and several analytical theories have been obtained by many researchers and reported in the literature at varying levels of generality and rigor. The following commented list of publications traces the efforts and achievements over the years. Despite the tremendous progress made during the past decades on the theory, modeling, analysis, and characterization of phase noise, certain gaps still remain and a "standard" theory for phase noise is not established at present and an incomplete collection of references is provided below.

- Leeson, 1966 [298] – simple generic expression for phase noise in LC oscillators.

- Barnes, 1966 [312] – delay-difference approach for evaluation infinite variances, and used in Allan variance of frequency.

- Allan, 1966 [311] – introduction of finite time variance measurements as the measure of infinite time variance of frequency. The approach is known as determination of Allan variance now.

- Lax, 1967 [324] – nonlinear, nonstationary oscillator noise analysis based on Langevin theory and Fokker–Planck equations.

- Abidi, Meyer, 1983 [325] – jitter in relaxation oscillators.

- Vanicola, Varshney, 1983 [326] – dispersion of modulated signals in Lorentzian noise spectra due to oscillator white and random walk phase noise.

- Chen, Ziel, Amberiadis, 1984 [299] – identification of circuit non-linearity as additional contributor to phase noise.

- Kartner, 1990 [327] – nonstationary phase noise model based on Floquet theory.

- Weigandt, Kim, Gray, 1994 [328] – CMOS jitter in ring oscillators with time variant noise sources and confirmed by Monte Carlo circuit simulations.

- McNeill, 1994 [329], 1997 [330] – jitter in ring oscillators.

- Razavi, 1996 [331] – practical aspects of time invariant analysis of noise in CMOS oscillators.

- Poore, 1997 [304] – method for harmonic balance in calculation of phase noise in mixers and oscillators.

- Takagi, Serikawa, Kurita, 1997 [332] – experimental evidence for correlation between 1/f and phase noise in BJT amplifier, showing that 1/f noise can be reduced, detecting phase noise. The correlation is via the diffusion coefficient in BJT and was modeled with transmission line model in [333] in 2005.

- Samori, Lacaita, Villa, Zappa, 1998 [316] – contribution from harmonics to phase noise in LC oscillators, by spectral "folding", caused by convolution between broadband noise and oscillator harmonics.

- Samori, Lacaita, Zanchi, Pizzolato, 1998 [310] – verification of the relation between phase noise and cycle-to-cycle jitter – see eq. (356).

- Herzel, 1998 [334] – phase noise is treated as spectral optical linewidth broadening of lasers. At low-frequency offset, due to 1/f noise, the phase noise spectrum has Gaussian probability shape $\exp[-(\Delta f/\sigma)^2]$ similar to Doppler broadened laser light lines, while at larger frequency offset the phase noise spectrum has Lorentzian shape $2D_\varphi/[(D_\varphi)^2+(2\pi\Delta f)^2]$, where D_φ is "diffusivity" of phase in the circuit. The latter Lorentzian shape for phase noise from white noise is basically in the same form as eq. (354) when removing from the equation the sum $\sum c_f S_f$ that was related to "colored" noise. The convolution in frequency domain between circuit impulse response and noise is addressed.

- Hajimiri, Lee, 1998 [315] – introduction of impulse sensitive function (ISF) for the phase in oscillators in perturbation analysis of phase noise. The method emphasizes the dominant role of phase perturbation and neglects the amplitude

perturbation in oscillators. This simplification made the phase noise analysis straightforward, relaxing the complexity in computer simulations. The simplification was subject of criticism, but the results from ISF analysis converge to other methods of perturbation analysis.

- Post, Linscott, Oslick, 1998 [321] – requirements for waveform symmetry that minimize phase noise, according to the ISF method for analysis of phase noise.

- Hajimiri, (Limotyrakis), Lee, 1998 [322], 1999 [317] – application and verification of the ISF method for analysis of phase noise, with analysis of circuit asymmetry that causes extra phase noise in ring oscillators.

- Margarit, Tham, Meyer, Deen, 1999 [323] – identification of non-monotonic behavior of phase noise level as function of bias and gain in the oscillator amplifier. The optimum condition is when the ratio harmonic amplitude/DC bias is at maximum, as we have shown here in Figure 64 and follows from eqs. (398)–(401).

- Demir, Mehrotra, Roychowdhury, 2000 [307] – full perturbation theory for phase noise and jitter in oscillators from white noise. The method is mathematically complex and it is appropriate for computer simulations of circuits.

- Demir, 1998 [306], 2002 [308] – inclusion of 1/f and other "colored" noise sources in perturbation theory [307] for phase noise and jitter in oscillators.

- Coram, 2001 [335] – a critical discussion that the ISF method for phase noise analysis misses the contribution of amplitude noise in phase noise. That is true, when simplifying the vector norm in the denominator of eq. (375) to constant and assuming that the vector of disturbance is always in the direction of the amplitude of the oscillation, which is never the case, since the noise has random phase, while the oscillation signal and its derivatives are cyclostationary, thus, the disturbance is never well in phase with the amplitude of the oscillation. Therefore, the potential underestimation by ISF method perhaps is not more of 3dB, which, practically, is similar to (or less than) the measurement accuracy, and this underestimation is not a serious issue at present, while the vast linearization and computational difficulties by numerical integration may cause unforeseen instability of solutions when using Floquet theory mapped in discrete numerical methods. Indeed, the demand of resources (CPU time, memory) and the accuracy of the different methods for phase noise analysis in computer simulators are not consistently addressed in the literature, especially in relation to accuracy of device models.

- Vanassche, Gielen, Sansen, 2002 [336] –ISF and full perturbation methods are identical for stationary noise sources, and ISF method fails for long non-stationary noise (or systems with locking), since no period of cyclostationary process is present, which violates the assumption for ISF. On the other hand, stationary noise was used in the derivation of closed expressions for phase noise from full perturbation in [307, 308], so, perhaps the assumption that the amplitude noise can be neglected in ISF method is not very critical.

- Dai, Harjani, 2002 [318] – effective quality Q-factor for phase noise in ring oscillators and coupled ring oscillators.

- Vanassche, Gielen, Sansen, 2003 [337] – a generalized semi-analytical method for derivation of phase noise expressions obtained by perturbation theory, and the results also are similar to that in [334] by assumption of spectral line widening. This method indicates that phase noise expressions can be derived by different means, but the final result will be the same. Various issues in oscillator noise analysis, ranging from "averaging" of differential equations, through split of process into fast and slow parts for behavioral modeling, to near-carrier spectrum.

- Grozing, Berroth, 2004 [319] – calculation of minimum possible phase noise in CMOS ring oscillators with relations to transistor geometry.

- Navid, Lee, Dutton, 2005 [320] – calculation of minimum possible phase noise in relaxation (RC, ring, no inductors) oscillators.

- Demir, 2006 [309] – use of frequency integrator macro model for phase that reflects and simplifies the perturbation methods for analysis of phase noise and jitter.

- Chorti, Brookes, 2006 [314] – formal mathematical modeling of phase noise when the phase noise spectrum could have several different power-law slopes ($PhN(\Delta f)=\sum(a_n/\Delta f)^n$, where a_n are constants).

- Zhao, 2009 [338] – differential Colpitts voltage controlled oscillator based upon the understanding on symbolic expressions of negative resistance and phase noise theory.

- Liu, 2016 [339] – Theoretical analysis using time-varying phase noise theory derives closed-form symbolic formulas for the $1/f^2$ phase noise region, showing that this feedback path could improve the phase noise performance

- Siddiq, 2019 [340] – phase noise theory developed for FMCW radar systems. New design equation derived to specify the maximum bound on the allowable source phase noise level in radar systems.

From the above publications, one could see that the results from different theories and methods for phase noise analysis converge, but a really general and unique theory for phase noise might be not possible to achieve. The bottom line is that the up-conversion of low-frequency noise is not the only process, which generates phase noise, although the 1/f and other "colored" noise sources dominate in near-carrier spectrum of phase noise, and the contribution of up-converted low-frequency noise is higher when the transitions in oscillating process are asymmetrical. Thus, the vertical symmetry in circuits and less harmonics in the non-linear oscillator circuit help to minimize the up-conversion, but this is a trade-off with the level of power supply used. Nevertheless, better designs [341] can achieve micro-power VCO with oscillation frequency in the GHz range and phase noise -115 dBc/Hz to -125 dBc/Hz at frequency offset of 1 MHz, by the simultaneous optimization of contradicting factors - reduction of phase noise level and power consumption, even when the VCO frequency increases.

VII.3. Noise in sensors

Electronic circuits are widely used in sensors for different purposes, e.g., for measurement of pressure, temperature, bio-chemical concentrations, light detection and imagers, to mention some sensor applications. The variety of sensors is large and it is impossible to address the implication of the noise in all types of sensors. Therefore, we briefly discuss the noise in sensors with examples for electrochemical and photo sensors, in order to illustrate the significance of noise in transistors and other electronic devices in sensors. The sensor's figures of merit, e.g., signal-to-noise ratio (SNR) are defined in respect to the sensed quantity. Consequently, the sensor definitions differ from the definitions for electronic devices presented earlier in Sec. IV.4.3. ***Noise factor, noise resistance, noise temperature***.

VII.3.1. Noise in electrochemical sensors

The electrochemical sensors acquire the potential difference (voltage ΔV_{Nion}) between the sensing electrode and the chemical solution of a given ion concentration, relying on the Nernst relation that the change in the potential difference $\Delta V_{Nion} \propto (\varphi_t/\eta)\ln(N_{ion})$ is a component proportional to the logarithm of the ion concentration N_{ion} in the chemical solution, where $\varphi_t = kT/q \approx 0.026V$ is the thermal voltage at room temperature T=300K and η is the valence of the ion, e.g., $\eta = \pm 1$ for Na^+ and Cl^- ions, and $\eta = \pm 2$ for Ca^{2+} and O^{2-} ions. If the chemical does not dissociate in the solution, then the chemical molecules have to be polarized dipoles, which is usually the case for bio molecules, such as DNA, η can be fractional number, and the dipoles have to be properly aligned in order the potential difference (voltage ΔV_{Nion}) to occur. Thus, the sensitivity of the electrochemical sensors, e.g., Si nano-wire pH sensor [342] at room temperature T=300K with $\eta = 1$ and 1pH=1decade of N_{ion}, is

$$K_{Nion} = \Delta V_{Nion} / \log_{10}(N_{ion}) = (\varphi_t/\eta)\ln(10) \approx 60mV/pH = 60mV/dec\ of\ N_{ion} \quad (403)$$

In practice, the sensitivity K_{Nion} of the electrochemical sensor varies below the value predicted by eq. (403), e.g., between 30mV/dec and 55mV/dec for the Si nano-wire pH sensor in [342], because the ion activity, which actually participates in the Nernst equation in place of N_{ion}, is a fraction of the ion concentration N_{ion}.

If the chemical does not dissociate in ions and does not have polarized molecules, then it is electrochemically inactive, the Nernst equation does not apply, the solution behaves as an electrical insulator and the chemical concentration cannot be acquired by means of electrical potential difference (voltage ΔV_{Nion}) measurement. Other requirement for the electrochemical sensors is that the liquid solution has to be in equilibrium, that is, chemical reactions have finished (e.g., reduction and oxidation, if any, have reached stationary rates), the temperature is constant (and known), and there is not a directed flow of ions, that is, there is no or at least negligible electrical current flow through the chemical solution.

The above discussion implies a structure of reference electrode (RE), a solution with ion/dipole concentration N_{ion}, an ion selective layer, which adsorbs the desired chemical on an insulating layer (and repulses other chemicals), and the second electrode, e.g., a silicon layer. Adding two contacts of width W and at distance L, one obtains a MOSFET-like structure, with the gate being the reference electrode, a gate stack of (solution + ion selective layer + SiO_2 insulator), and transistor channel in the semiconductor between the two contacts in the silicon layer. This structure is known as the ion-selective field-effect transistor (ISFET). One advantage of this structure is that it can be manufactured by the usual MOS/CMOS fabrication processes, optionally omitting deposition of the polysilicon or metal gate and adding only two extra steps for deposition of the Ag/AgCl reference electrode and the ion-selective layer for functionalization of the surface of the SiO_2 gate dielectric of the ISFET. Thus, a second advantage of the ISFET is that the ISFET is of micrometer-size and it can be integrated in silicon and microfluidic chips as a single sensor or an array of sensors. A third advantage is that the ISFET avoids wiring (and environmental noise disturbances, thereof) by stacking the $N_{ion}-\Delta V_{Nion}$ sensing interface (liquid-ion selective layer-dielectric-semiconductor) in the gate stack of the ISFET. A fourth advantage is that ΔV_{Nion} is equivalent of a change ΔV_G of gate biasing of the ISFET, thus, the ISFET provides a signal gain through the transistor trans-conductance $g_m=\Delta I_D/\Delta V_G=\Delta I_{Nion}/\Delta V_{Nion}$.

From the fourth advantage, one deduces that the gate voltage noise S_{VG} of the MOS transistor can be referred as a noise S_{Nion} of the sensed quantity through the derivative of the Nernst equation, and vice-versa, depending which noise source is considered.

$$S_{VG} = (\varphi_t/\eta)^2 \, S_{Nion} / N_{ion}^2 = K_{Nion}^2 \, S_{log10(Nion)} \, , \tag{404}$$

where $S_{log10(Nion)}$ is PSD in unit [(decades of N_{ion})²/Hz] when S_{VG} is PSD in unit [V²/Hz], or $S_{log10(Nion)}$ is a squared RMS value in unit [(decades of N_{ion})²] for a given frequency band $f_{min}-f_{max}$ when S_{VG} is a squared RMS value in unit [V²/Hz] in the same frequency band. The noise from other ISFET parts (reference electrode, liquid solution) is negligible [343]. One should use low-noise voltage sourcing circuit for biasing the ISFET gate (the reference electrode), because the noise of this source sums with and might be larger than the voltage noise S_{VG} of the MOS transistor in the ISFET.

Note in eq. (404) that the voltage noise in absolute unit of volts at the ISFET gate causes normalized noise for the concentration N_{ion}, which means that the voltage noise causes multiplicative error for N_{ion} and uncertainty for the sensitivity (conversion "coefficient") of the electrochemical sensor. Therefore, the voltage noise limits the minimum relative change $\Delta N_{ion}/N_{ion}$ that can be sensed, but it does not cause a threshold for minimum N_{ion}, that can be detected. Consequently, the signal-to-noise ratio (SNR) of the electrochemical sensor is defined in [342] for the sensor sensitivity K_{Nion} (but not for the concentration N_{ion}), as a ratio of the RMS value of the gate voltage noise S_{VG} of the MOS transistor for a frequency band $f_{min}-f_{max}$ and the voltage for one decade change of N_{ion}, (that is one pH unit in [342]), which is the value of K_{Nion} in unit of volts. Thus, the first line of the next eq. (405)

shows that the definition is ratio of the sensor sensitivity K_{Nion} to the root-mean-square (RMS) value of the gate voltage noise S_{VG} of the MOS transistor for a frequency band f_{min}–f_{max}. Furthermore, using the relations for the noise in the electrochemical sensor in the previous eq. (404), one sees in the second line of eq. (405) that this definition of SNR is actually a reciprocal of the RMS value of the normalized noise of the ion concentration N_{ion} in the frequency band f_{min}–f_{max}. Thus, SNR and the normalized noise for N_{ion} are figures of merit for the noise in the electrochemical sensor that have reciprocal values, but hold the same information for the noise.

$$\begin{aligned} SNR &= \frac{K_{Nion}}{\sqrt{\int_{f_{min}}^{f_{max}} \frac{S_{VG}(1Hz)}{f} df}} = \frac{K_{Nion}}{\sqrt{S_{VG}(1Hz) \times \ln\left(\frac{f_{max}}{f_{min}}\right)}} = \frac{K_{Nion}}{RMS_{VG}} \\ &= \frac{K_{Nion}}{K_{Nion} \times RMS_{\log 10(Nion)}} = \frac{1}{RMS_{\log 10(Nion)}} = \sqrt{\frac{N_{ion}^2}{\int_{f_{min}}^{f_{max}} S_{Nion}(f) df}} = \frac{1}{\sqrt{\text{norm.noise of } N_{ion} \text{ in frequency band } f_{min} - f_{max}}} \end{aligned} \quad . \quad (405)$$

We now provide for expected typical values for the noise in electrochemical sensors. The chemical processes have time constant in the range of milliseconds, so one desires the acquisition of chemical concentration with a rate of more than readings per 10 seconds for changes in the chemical concentration to be detected. Accordingly, the frequency band for noise of the electrochemical sensor is from 0.01Hz to 100Hz, in which the 1/f noise of the MOS transistor dominates in the gate-referred voltage noise S_{VG}. The sensor usually covers up to 5 decades for N_{ion}, which results in approximately 0.3V span of the gate voltage $\Delta V_G = \Delta V_{Nion} = K_{Nion} \times 5$ decades $\approx$ 60mV/dec $\times$ 5 decades = 0.3V from and above the threshold voltage of the MOS transistor. For such span of the gate voltage, the gate voltage noise S_{VG} of the MOS transistor can be considered of approximately constant value, especially for MOS transistors with not very thin oxides of thickness 5-20nm, 5nm in [342] and 17.5nm in [343], which prevent from gate currents and currents through the liquid (a requirement discussed above) and withstand possible electrostatic damages in the ISFET, when using it in the sensor with simple ESD protection. Putting all together, the RMS value of the gate voltage noise S_{VG} of the MOS transistor in the ISFET is

$$\begin{aligned} RMS_{VG} &= \sqrt{S_{VG}(1Hz) \times \ln\left(\frac{f_{max}}{f_{min}}\right)} = \sqrt{S_{VG}(1Hz) \times \ln(10) \times \log_{10}\left(\frac{f_{max}}{f_{min}}\right)} \\ &= \sqrt{S_{VG}(1Hz)} \times \sqrt{2.3 \times \log_{10}\left(\frac{100Hz}{0.01Hz}\right)} \\ &= \sqrt{S_{VG}(1Hz)} \times \sqrt{2.3 \times 4 \text{ frequency decades}} = 3.03 \times \sqrt{S_{VG}(1Hz)} \end{aligned} \quad . \quad (406)$$

where $S_{VG}(1Hz)$ is the power spectrum density (PSD) at frequency 1Hz of the 1/f gate-referred voltage noise of the MOS transistor in the ISFET.

The value of $S_{VG}(1Hz)$ can be obtained by several approaches. One approach is the experimental determination of PSD of $S_{VG}(f_{tst})$ at several test frequencies f_{tst} , and the averaging calculation of $S_{VG}(1Hz)$ by $S_{VG}(1Hz)$=average($f_{tst} \times S_{VG}(f_{tst})/1Hz$). This approach was undertaken in [342], obtaining values in the range of $S_{VG}(1Hz) = \{3 \times 10^{-8}$ V²/Hz to 3×10^{-6} V²/Hz$\}$ for a nano-wire ISFET. A second approach uses noise models, and determines $S_{VG}(1Hz)$ from simulations. This approach was undertaken in [343], obtaining values in the range of $S_{VG}(1Hz) = \{2 \times 10^{-11}$ V²/Hz to 3×10^{-10} V²/Hz$\}$ for a bioFET. A third approach uses statistical data for noise and

predicts typical values of $S_{VG}(1Hz)$ by scaling rules. Consider $FOM_{S_{VG}}$ from eq. (106) in Sec. IV.1.2. ***Coupled noise component (number fluctuation)***.and Figure 20. From the trend (thick gray line labeled by "Average" in Figure 20), $FOM_{S_{VG}}=6\times10^{-9}$ μm²V²/Hz for an oxide thickness EOT=10nm of a typical ISFET. The values of $FOM_{S_{VG}}$ can be four times smaller for EOT=5nm (for the NW ISFET in [342]), three times larger for EOT=17.5nm (for the bioFET in [343]) and four times larger for EOT=20nm. These values may also vary a decade above and below the average, since the logarithmic standard deviation of $FOM_{S_{VG}}$ is 10dB, as indicated with σ_{dB} in Figure 20 and Figure 21. As follows from eq. (106), the PSD at 1Hz of the 1/f gate-referred voltage noise of the MOS transistor in a typical ISFET with EOT=10nm gate oxide is $S_{VG}(1Hz)=FOM_{S_{VG}}/(WL)=6\times10^{-9}$ μm²V²/Hz /(1000 μm²)=6×10^{-12} V²/Hz, with gate area WL=1000 μm² for the typical ISFET.

Summarizing the approaches for the determination of $S_{VG}(1Hz)$, one expects that the PSD at 1Hz of the 1/f gate-referred voltage noise of the MOS transistor in a ISFET is in the range of $S_{VG}(1Hz)$ = {10^{-11} V²/Hz to 10^{-6} V²/Hz}, depending on gate oxide thickness and sizes of the MOS transistor. Substituting in eq. (406), the RMS value of the gate voltage noise S_{VG} of the MOS transistor in the ISFET is expected in the range

$$\begin{aligned} RMS_{VG} &= 3.03\times\sqrt{S_{VG}(1Hz)} \\ &= 3.03\times\sqrt{10^{-11}\,V^2/Hz} \text{ to } 3.03\times\sqrt{10^{-6}\,V^2/Hz}\,. \\ &= 10\mu V \text{ to } 3mV \end{aligned} \qquad (407)$$

Considering from eq. (403) that the sensitivity of the electrochemical sensors does not exceed $K_{Nion} \le 60mV/pH = 60mV/dec$ of N_{ion}, then from eq. (404), the RMS value of the noise in the electrochemical sensor is in the range

$$\begin{aligned} RMS_{\log10(Nion)} = \sqrt{S_{\log10(Nion)}} &= \frac{RMS_{VG}}{K_{Nion}} = \frac{10\mu V \text{ to } 3mV}{30mV/dec \text{ to } 60mV/dec} \\ &= (0.00017 \text{ to } 0.1) \text{ decades of } N_{ion} \text{ or pH} \\ &= (0.04\% \text{ to } 26\%) \text{ of } N_{ion} \end{aligned} \quad , \qquad (408)$$

showing that small-size and "noisy" MOS transistor in ISFET may ruin the accuracy of the electrochemical sensor, just because of high level of the low-frequency noise.

Finally, from the reciprocal dependences in eq. (405), the signal-to-noise ratio (SNR) of the electrochemical sensor with ISFET is in the range

$$SNR = \frac{K_{Nion}}{RMS_{VG}} = \frac{1}{RMS_{\log10(Nion)}} = 10 \text{ to } 6000\,. \qquad (409)$$

showing that the electrochemical sensor with ISFET can theoretically have good SNR, but similarly to above, the high level of the low-frequency noise small-size and "noisy" MOS transistor in ISFET can reduce the SNR to unacceptable low values, e.g., causing unstable readings of the pH sensor for all digits after the decimal point.

VII.3.2. Noise in photo detectors and imaging arrays sensors

Sensors for light are increasingly used at present, because they provide non-contact and remote sensing and imaging, which other types of sensors cannot, converting the optical power into photo current, followed by electrical circuit, which amplifies the photo current and convert it into useful electrical signal, e.g., voltage. Inherent for the noise in photo detectors is that the optical power is low, which results in low photo current,

which is comparable with leakage currents and noise in electronic devices and circuits. The problem is reinforced, when reducing the area of the photo sensors, which is the case in multi-million-pixel imaging arrays. Therefore, photo detectors with built-in amplification and/or impedance buffering are of main interest at present, in order to sense very low optical power by micrometer-sized opto-electrically active areas of the photo sensors, which can be integrated also in the semiconductor chip together with other electronic circuits for signal and data processing and communication.

Below, we briefly discuss the noise in the associated circuits of two types of commonly used photo detectors with built-in amplification and buffering. These photo detectors are the avalanche photo diodes (APDs), which can detect even a single photon, and the active pixel sensors (APSs), which are compatible for integration with the electronic circuits in silicon chips to create imaging arrays with millions of pixels, but still small in size and cheap to be used in handheld personal devices, e.g., cell phone cameras, and at the tip of medical devices, e.g., cameras for endoscopes, dentistry and elsewhere.

The avalanche photo diodes (APDs) are PN, PIN and more complex junction structures, reverse-biased with excess voltage V_{ex} slightly above the junction breakdown voltage V_{BR}. At this biasing condition, an avalanche process multiplies the primary carriers generated thermally and optically in the junction depletion layer. The avalanche multiplication coefficient is $M \propto \exp(V_{ex})$, resulting in larger reverse current $I_R=M(I_{leak}+I_{ph})$ compared to the primary reverse currents, the thermally generated current I_{leak} (observable when the junction is in dark) and the optically generated (photo) current $I_{ph} \propto P_{opt}$ proportional to the optical power P_{opt} absorbed in the depleted layer. Thus, the gain of the APD is the avalanche multiplication coefficient M. Since the junction is reverse biased, then the carriers traverse the junction quickly, producing short "spikes" of current, and shot noise with white power spectrum density (PSD) of magnitude [344, 345, 346, 347, 348]:

$$S_{APD,white} = 2q(I_{leak}+I_{ph})M^2F = 2qI_R \times (MF), \text{ with } F=1\ldots M, \text{ thus, } F<\text{APD gain} \quad (410)$$

where F is called "excess noise factor", please see the above cited references for details on what and how F depends on, and F is lower than the APD gain M (avalanche multiplication coefficient).

Eq. (410) is a good model for the APD noise at medium and high frequencies, e.g., when using the APD in fiber-optical communications or for RF noise sources with PIN diodes, and when the APD gain M<100 is not high. However, for photo sensors operating in the continuous regime, one observes the low-frequency 1/f noise of the APD, as one can see in [346, 347, 348]. In these references, the APDs are hetero-structures with separate light absorption and charge multiplication layers, the APD noise with 1/f power spectrum density (PSD) is described as $S_{APD,1/f}=K_F \times I_R^2/f$, and the values of the K_F are in the range from below 10^{-6} to above 10^{-4}, varying with the APD structure, processing and bias. However, noise with Lorentzian spectra are also observed in APD, occasionally at low I_R [348] and regularly at elevated V_{ex} that corresponds to high gain M>200. The problem is that a micro-plasma random switching occurs [46], when attempting elevated V_{ex} for high gain M of the APD. Therefore, V_{ex} of the APD bias in continuous operating photo sensors has to be kept so that the APD gain is low, e.g., M=20…50<100. Other low-level light applications of photo sensors require high gain M>1000, e.g., single photon detection/counting in photo sensors for Raman spectrometers and time decaying fluorescence imaging (FLIM), in which the APD operates in Geiger mode. In this single photon detection regime, V_{ex} is elevated and the micro-plasma is forced to extinguish by special high-speed (~ GHz) active quenching and reset circuits. Even though, there are problems with missing detection of photons and spurious after-pulsing of the micro-plasma. Many publications are devoted recently for single-photon APDs (SPADs), but the low-frequency noise in SPAD

is not elaborated well in these publications in order to extend a discussion here.

The active pixel sensors (APSs) are a combination of a photodiode and an electronic circuit, which allows integrating the photo sensing with essential functionality in each pixel of an imager array. The APS functionality comprises cyclic operation similar to that of a CCD (charge-coupled device). Each cycle includes "phases" of reset, intergradation and readout. The reset phase connects a voltage source to apply a certain and maximum potential/voltage across the depleted layer of the photo sensing structure (creating a potential well under the gate in the CCD, or a reverse voltage on the capacitance of the photodiode in the APS). By disconnecting the voltage source in the integration phase, the photo generation of charge reduces the depletion, through collecting photo charge in the potential well of the CCD, or through discharging the photo diode capacitance by the photo current. Then, by a set of switches that connect the pixels sequentially to a readout amplifier, the readout phase acquires the amount of charge collected in the CCD, or the reduction of voltage across the photo diode in APS, where the reduction is in respect to the maximum reverse voltage set in the reset phase.

While the phases of reset-integration-readout in cyclic operation of CCD and APS are similar, the structures and circuits for CCD and APS are very different. From the circuit perspective, the CCD requires switching of high gate voltages (in the range of 10-30V), and the CCD structure is incompatible with mainstream CMOS processes, which is a problem for the single-chip integration of CCD with other electronic circuits. The advantage of the APS is that all devices in the APS are readily available in mainstream CMOS processes, e.g., the photo diode is just one of the PN junctions used for drain, source or well for MOS transistor in the silicon substrate. Accordingly, all parts in the APS use the ordinary supply for other analog and digital circuitry in the CMOS chip. This made the APS the cost-effective and attractive choice for building imagers nowadays, and we briefly discuss the noise in APS below.

The noise in APS with a typical circuit shown in Figure 65 (a) is analyzed in [349] in terms of the voltage across the capacitance C_{PH} of the node $v_{s.}$ of the photo diode D. The pixel consists of a photodiode D and three MOS transistors. M1 is the reset transistor, which resets the photo diode voltage level close to the supply voltage VDD before integration. The transistor M2 is the buffering source follower, which isolates the high-impedance photo diode node from the low-impedance readout bus. The transistor M3 is the pixel select transistor, which connects the pixel to the readout bus. During the readout phase, the channel resistance of M3 is low, and the pixel output voltage $V_o=(V_s-V_{T.M2})$ at node v_o follows the photo diode voltage V_s at node v_s, offset-shifted with the threshold voltage $V_{T.M2}$ of M2. The transistor M4 is the pixel load shared by the pixels connected to the readout bus.

Shot noise and "KTC" noise in APS. The shot noise $2qI_R=2q(I_{leak}+I_{ph})$ in the photo diode reverse current $I_R=(I_{leak}+I_{ph})$ is mainly considered in [349] at different reset and integration times. The shot noise current is converted into voltage noise by the differential resistance $r_d=dV_s/dI_R\sim\varphi_t/I_R$, where V_s is the voltage of node $v_{s.}$ and $\varphi_t=kT/q\approx 0.026V$ is the thermal voltage at room temperature T=300K. Accordingly, as given in [349], the voltage noise of node $v_{s.}$ is

$$\begin{aligned} RMS^2_{Vs(I_R)} &= \frac{qI_R r_d}{2C_{PH}} F \approx \frac{qI_R}{2C_{PH}} r_d, \text{ with } F \approx 1 \text{ and } r_d = \frac{dV_s}{dI_R} \\ &\approx \frac{qI_R}{2C_{PH}} \times \frac{\varphi_t}{I_R} = \frac{kT}{2C_{PH}} = \text{KTC noise, assuming } r_d \approx \frac{\varphi_t}{I_R} \end{aligned} \tag{411}$$

where the factor F depends weakly on the reset and integration times. Note that due to $r_d \sim \varphi_t / I_R$, the shot noise has been converted into thermal-like noise, often termed as the "KTC" noise in imagers. Note also that the KTC noise is in principle independent of bias and photo current. Consider in addition that the voltage of the v_s node is sampled twice in the so-called correlated double sampling (CDS), once in the beginning and then at the end of the integration phase, so that the voltage swing ΔV_s of the discharge of the capacitance C_{PH} of node v_s is obtained. This swing is proportional the photo current, since the swing is given by

$$\text{CDS swing } \Delta V_s = V_{reset} - V_{integ.end} = \frac{I_{leak} + I_{ph}}{C_{PH}} t_{int} \tag{412}$$

where t_{int} is the integration time, and many offset voltages in the circuit are compensated, except for the offset due to leakage current I_{leak}. Thus, considering ΔV_s as a signal, the signal to noise ratio (SNR) due to shot noise in the photo diode in the APS is

$$\begin{aligned} SNR_{shot} &\approx \frac{\Delta V_s}{\sqrt{2 \times \text{KTC noise}}} = \frac{I_{leak} + I_{ph}}{\sqrt{kT \times C_{PH}}} t_{int} \\ &= \frac{I_{leak} + I_{ph}}{\sqrt{q \frac{kT}{q} C_{PH}}} t_{int} = \frac{I_{leak} + I_{ph}}{\sqrt{q \varphi_t C_{PH}}} t_{int} \approx \frac{I_{leak} + I_{ph}}{\sqrt{1.6 \times 10^{-19} \text{C} \times 26\text{mV} \times C_{PH}}} t_{int} , \\ &\sim \frac{\text{fA to pA}}{64 \times 10^{-12} \sqrt{\text{fF to pF}}} (\mu\text{s to ms}) \end{aligned} \tag{413}$$

The SNR_{shot} increases with the photo current and integration time, and the orders of the magnitudes of the different quantities are swing ΔVs = 100mV, noise $2 \times KTC = (0.2mV)^2$ and $SNR_{shot} \approx 500 = 54dB$ for $I_R = (I_{leak} + I_{ph}) = 1pA$, $C_{PH} = 0.1pF$ and $t_{int} = 10ms$. Eq. (413) suggests that SNR increases with the integration time, but the detailed analysis in [349] indicates non-monotonic SNR as function of the reset and integration times, as shown in Figure 65 (b) and (c), mainly due to non-linearity in the photo diode and MOS transistor I–V and C–V characteristics and saturation of the voltage of the photo diode node v_s or of the APS output node v_o near the GND supply rail at long integrations time and high photo currents.

In addition to the shot noise of the photo diode discussed above, a significant contribution to the APS noise has the buffering transistor M2 in Figure 65 (a), because it operates in the active mode of source follower. This transistor contributes with 1/f noise, and if the gate area is small, e.g., $(W \times L) < 3\mu m^2$, then it contributes also with random-telegraph-signal (RTS) noise, which can cause pixel "blinking", if a step transition between the levels of the RTS waveform occurs during the integration phase between the two samples of the correlated double sampling (CDS). The other transistors have small to negligible noise contribution, because they are operating as switches (M1 and M3) and the load transistor M4 can be of large area, being one for hundreds of pixels in the imaging array, and not affecting the fill factor of the array.

To refer ***the 1/f noise of the buffering transistor*** M2 in Figure 65 (a) to the photo diode node v_s, one considers the gate-referred 1/f voltage noise $S_{VG}(f) = S_{VG}(1Hz)/f$, for which we have scaling rules and statistical data for $FOM_{S_{VG}}$ in eq. (106) in Sec. IV.1.2. ***Coupled noise component (number fluctuation)*** and in Figure 20. From the trend (thick gray line labeled by "Average" in Figure 20), $FOM_{S_{VG}} = 6 \times 10^{-9}$ $\mu m^2 V^2/Hz$ for an effective oxide

thickness EOT=10nm of the gate dielectric in a MOS transistor. Having scaling rule $FOM_{S_{VG}} \propto EOT^2$ and typically a gate dielectric thickness EOT=3nm to 5nm for the buffering transistor M2, depending on the CMOS technology node, then the values of $FOM_{S_{VG}}=1.5\times10^{-9}$ µm²V²/Hz to 6×10^{-10} µm²V²/Hz, for EOT=5nm and 3nm, respectively, are four to ten times smaller than $FOM_{S_{VG}}=6\times10^{-9}$ µm²V²/Hz for EOT=10nm. According to eq. (106), and knowing the area (W×L) of the buffering MOS transistor M2, the power spectrum density (PSD) at 1Hz of the gate-referred 1/f voltage noise of M2 is $S_{VG}(1Hz)=FOM_{S_{VG}}/(W\times L)$, and the PSD of the 1/f noise associated with the photo diode node v_s is

$$\text{PSD of the 1/f noise in APS} \qquad \begin{aligned} S_{VG}(f) &= \frac{S_{VG}(1\text{Hz})}{f} = \frac{FOM_{SVG}/(W\times L)}{f/1\text{Hz}} \\ &= \frac{(6\text{ to }15)\times10^{-10}\left[\mu\text{m}^2\text{V}^2/\text{Hz}\right]/(3\mu\text{m}^2\text{ to }10\mu\text{m}^2)}{f/1\text{Hz}} \\ &= \frac{(0.6\text{ to }5)\times10^{-10}\,\text{V}^2/\text{Hz}}{f/1\text{Hz}} \end{aligned} \tag{414}$$

Having delay t_{int} in the correlated double sampling (CDS), the squared magnitude $|H_{CDS}(f)|^2$ of the transfer function of the delay-subtraction operation of the CDS is a cycling spectrum (spectral "comb") given in [350] by

$$\left|H_{CDS}(f)\right|^2 = 4\sin^2(2\pi f t_{int}/2) \approx 2\times\text{HPF}(f \geq 0.25/t_{int}) \tag{415}$$

which, for the purpose of 1/f noise analyses, can be approximated by a high-pass filter HPF($f\geq f_{min}$) with power gain 2 and corner frequency $f_{min}=0.25/t_{int}$. The values of the HPF parameters are easily deduced, considering the two samples in CDS, thus, the power gain=2, and f_{min} as the min frequency, at which

$$|H_{CDS}(f_{min})|^2=\text{average}(|H_{CDS}(f)|^2)=\text{gain}=2.$$

The cycling of $|H_{CDS}(f)|^2=4\sin^2(2\pi f t_{int}/2)=2(1-\cos(2\pi f t_{in}))$ can be understood, considering that the subtraction in CDS suppresses fully a DC signal and "full-period" signals with frequencies of $f=n/t_{int}$, n=0, 1, 2,…, but adds "half-period" signals with $f=(n+½)/t_{int}$, because $\cos(2\pi((n+½)/t_{int})t_{in})=\cos(\pi)=-1$. Then, the identity $(1-\cos(x))=2\sin^2(x/2)$ explains the "twice faster" cycling of $|H_{CDS}(f)|^2$, that is $|H_{CDS}(f)|^2$ is a spectral "comb" filter with frequency step of $\Delta f=0.5/t_{int}$. Finally, the increase of the signal (actually, the 1/f noise) frequency from 0=DC to $0.5/t_{int}$ causes reduction of the signal suppression toward signal addition, passing through the corner frequency $f_{min}=0.25/t_{int}$ of the high-pass filter approximation, at which the CDS does neither suppresses nor enhances the average power gain=2 of the high-pass filter approximation. Comparing the frequency step $\Delta f=0.5/t_{int}$ of $|H_{CDS}(f)|^2$ of the spectral "comb" with the slower-varying spectrum of the 1/f noise, the given in [350] Riemann-like integration at the $|H_{CDS}(f)|^2$ spectral "comb" for determination of the root-means-square (RMS) value of the 1/f noise can be replaced with continuous integration, that is

$$
\begin{aligned}
\mathrm{RMS}^2_{1/f,CDS} &= \frac{1}{2\pi}\int_{-\omega}^{\infty} S_{1/f}(\omega)\left|H_{CDS}(\omega)\right|^2 \mathrm{LPF}(\omega \le \omega_{max})\,d\omega, \text{ but } S_{1/f}(\omega<0)=0 \\
&= \frac{1}{2\pi}\int_{0}^{\infty} \frac{S_{1/f}(1\mathrm{Hz})}{f}\times\frac{4\sin^2(\pi f t_{int})}{1+(f/f_{max})^2}\,d(2\pi f), \text{ the known noise PSD} = 2\times\text{double-sided} \\
&= S_{1/f}(1\mathrm{Hz})\int_{0}^{\infty} \frac{4\sin^2(\pi f t_{int})}{f\left[1+(f/f_{max})^2\right]}\,df = \text{ complicated integral} \approx \int_{0}^{\infty}\frac{S_{1/f}(1\mathrm{Hz})}{f}\times 2\times \mathrm{HPF}(f\ge f_{min})\times\mathrm{LPF}(f\le f_{max})\,df \\
&\approx 2S_{1/f}(1\mathrm{Hz})\int_{f_{min}}^{f_{max}}\frac{df}{f}, \text{ with } f_{min}=\frac{0.25}{t_{int}},\ f_{max}=\frac{1}{2\pi\tau_{acq}} \text{ and } (3 \text{ to } 5)\tau_{acq}\approx\frac{t_{int}}{10 \text{ to } 20} \\
&\approx 2S_{1/f}(1\mathrm{Hz})\times 1\mathrm{Hz}\times\ln\left(\frac{f_{max}}{f_{min}}\right) = 2S_{1/f}(1\mathrm{Hz})\times 1\mathrm{Hz}\times\ln\left[\frac{(30 \text{ to } 100)/(2\pi t_{int})}{0.25/t_{int}}\right] \\
&\approx 2\times(0.6 \text{ to } 5)\times 10^{-10}\,\mathrm{V^2/Hz}\times 1\mathrm{Hz}\times\ln(19 \text{ to } 64) = (0.019\ \mathrm{mV} \text{ to } 0.065\ \mathrm{mV})^2 < (0.2\ \mathrm{mV})^2 = 2\times\mathrm{KTC\ noise}
\end{aligned}
\tag{416}
$$

where $\omega_{max}=1/\tau_{acq}=2\pi f_{max}$ is the bandwidth of the readout amplifier for acquisition of the pixel output voltage at node v_o, considered as a first-order low-pass filter LPF with pole at f_{max} and time constant τ_{acq}. As shown in the fifth line in eq. (416), the acquisition time constant τ_{acq} has to have low value to instantly sample the pixel output voltage at node v_o for a short time of $t_{int}/(10 \text{ to } 20)$, and with low dynamic error, that is, the acquisition time constant τ_{acq} has to be (3 to 5) times shorter than the short acquisition time short time of $t_{int}/(10 \text{ to } 20)$.

The lines in eq. (416) indicate the following. The first line explains the method for determination of the squared $\mathrm{RMS}^2_{1/f,CDS}$ value of the noise by power-spectrum transfer functions $|H_{CDS}(\omega)|^2$ of the CDS and $\mathrm{LPF}(\omega<\omega_{max})$ of the band-limited to ω_{max} acquisition channel for the pixel output voltage at node v_o. The second line substitutes the equations for the 1/f noise PSD and the power spectrum transfer functions, also converting the angular frequency ω into regular frequency f in unit Hz The third line indicates that the integration results in a complicated function, which is a mixture of trigonometric and hyperbolic functions even for white noise, as one can see in eq. (10) in [350], and in eq. (11) in the same reference, also a sinc(x) function occurs in the mixture, when considering a finite time in the rectangular waveform of sampling in the CDS acquisition channel. Such complicated functions (although seemingly accurate) are not worth for analyses, since the scatter in the noise PSD around the geomean value of the PSD is in excess of ±6dB, as seen by the rocky curve for (and labeled by) Single acquisition in Figure 76 later in Sec. VIII.6. **Consequences from statistical nature of LFN – distributions in spectra, techniques of averaging, data volume and coordinates, instrumentation**.

Therefore, approximations are undertaken in the fourth line of eq. (416) with an ideal high-pass filter $\mathrm{HPF}(f>f_{min})$ of power gain=2 for the power-spectrum transfer functions $|H_{CDS}(\omega)|^2$ of the CDS and with an ideal low-pass filter $\mathrm{LPF}(f<f_{max})$ for the power-spectrum transfer function of the acquisition channel, which allows to write a simple definite integral in the fifth line, with integrand 1/f for the noise PSD and limits f_{min} to f_{max} (the values are discussed above) for the frequency pass-band determined by the CDS and the bandwidth of the acquisition channel. This pass-band applies for the calculation of $\mathrm{RMS}^2_{1/f,CDS}$ value of the 1/f noise in the APS. The sixth line of eq. (416) shows the resulting formulas for the calculation of $\mathrm{RMS}^2_{1/f,CDS}$ value of the 1/f noise in the APS.

The seventh line substitutes the values from eq. (414) for the PSD at 1Hz of the gate-referred 1/f voltage noise of the MOS transistor (gate oxide thickness EOT=3nm to 5nm, gate area (W×L)=3µm² to 10µm²) typically used for

the buffering transistor in APS, see again M2 the circuit in Figure 65 (a). The last, the eighth line shows that the $RMS_{1/f,CDS}$ due to the 1/f noise in APS is expected to be 3 to 10 times (10dB to 20dB) lower than the 2×KTC noise of the shot noise in the photo diode reverse current, estimated above after eq. (413). However, noticing that the scatter of $FOM_{S_{VG}}$ is σ_{dB}=10dB in Figure 20, which may cause the same 10dB increase in S_{VG} and in $RMS_{1/f,CDS}$, and if the buffering transistor area is reduced 10 times (10dB) in the range (W×L)=0.3μm² to 1μm², which would increase the 1/f noise of the MOS transistor with the same 10dB, then $RMS_{1/f,CDS}$ of the 1/f noise could be increased 20dB and may become comparable or even dominant in the APS. Other problem in APS with buffering transistor of sub−μm² gate area is the RTS noise, which can cause having "blinking" pixels in the imager array; and the problem is now discussed.

Mentioned above, ***the RTS noise in the buffering transistor***, M2 in the APS circuit in Figure 65 (a), could cause pixel "blinking", if a step transition between the levels of the RTS waveform occurs during the integration phase between the two samples of the correlated double-sampling (CDS). Similarly to above for the KTC and 1/f noise, one refers the RTS noise as a voltage noise at the photo diode node v_s, see Figure 65 (a), although it occurs actually at the output node v_o., since the pixel output voltage $V_o=(V_s-V_{T.M2})$ at node v_o follows the photo diode voltage V_s at node v_s, offset-shifted with the threshold voltage $V_{T.M2}$ of M2. Thus, the RTS noise voltage ΔV_{RTS} of the APS is the RTS noise voltage $\Delta V_{T,M2,RTS}$ in the threshold voltage of M2, it occurs as voltage noise $\Delta V_{o,RTS}=\Delta V_{RTS}=-\Delta V_{T,M2,RTS}$ at the APS output, and the output RTS noise voltage is identically referred to the photo diode node v_s, that is $\Delta V_{s,RTS}=\Delta V_{o,RTS}=\Delta V_{RTS}=-\Delta V_{T,M2,RTS}$.

As discussed in length Sec. IV.3 **RTS noise in MOS transistors**, the RTS change of the threshold voltage of the MOS transistor is due to RTS capture/emission of single electron in the gate capacitance WLC_{ox} of the MOS transistor. Therefore, from eq. (178), the voltage of the RTS change is $\Delta V_{RTS}=-\Delta V_{T,M2,RTS}=\pm q/(WLC_{ox})$, where the signs ± are alternating at random intervals between the changes. Considering also the CDS and eq. (412) for the voltage swing ΔV_s of the discharge of the capacitance C_{PH} of node v_s (due to the photo diode reverse current) during the integration phase, one writes for the CDS-acquired voltage swing ΔV_{APS} at the output of the APS (node v_o) that

$$\Delta V_{APS}=\Delta V_s+\Delta V_{RTS}=\Delta V_s\pm q/(WLC_{ox}),\text{ when RTS occurs with probability } PR_{RTS},\text{ otherwise } \Delta V_{APS}=\Delta V_s \quad (417)$$

This equation corresponds to the histogram [351] for single RTS shown in Figure 66 (a), heuristically explained in Figure 66 (b). There is a central peak corresponding to $\Delta V_{APS}=\Delta V_s$, and two lobes at $\Delta V_s\pm q/(WLC_{ox})$. The probability $PR_{RTS}=(\Sigma CNT_{lobe})/(CNT_{peak}+\Sigma CNT_{lobe})$ for occurrence of the RTS noise in the APS is the ratio of the counts ΣCNT_{lobe} of occurrences (often called frequency in the histogram) in the lobes over the total count $(CNT_{peak}+\Sigma CNT_{lobe})$ in the histogram, including the central peak. The RTS probability PR_{RTS} is split between the two lobes, but not necessarily equally split.

If two or more traps are involved in the RTS noise, then one has multi-level RTS, the number of lobes is 2×(number of traps), as shown heuristically in Figure 66 (c) for a triple RTS in an APS with a larger-area buffering transistor (M2 in Figure 65 (a)), having reduced spacing $q/(WLC_{ox})$ between the lobes. The reduced spacing results in overlaps of the peaks in the histogram. The overlaps eventually lead to inability to distinguish the peaks in the histogram envelope (gray curve), the individual RTS noises superimpose and, consequently, attribute into 1/f noise.

While the spacing $q/(WLC_{ox})$ between the lobes is certainly predictable in the histograms for the CDS-acquired voltage swing ΔV_{APS} at the output of the APS, the probabilities $CNT_{lobes}/(CNT_{peak}+\Sigma CNT_{lobe})$ of the lobes cannot

be elaborated well theoretically by diverse reasons, e.g., the RTS transitions are random, they randomly occur (or not) during the integration time, the occurrences of high-to-low and low-to-high level transitions of RTS are not necessarily alternating in consecutive CDS, the RTS time constants vary between the pixels and with the bias and illumination in imaging arrays, etc. Therefore, although the RTS noise in the buffering transistor causes RTS noise in APS, the rates of the pixel "blinking" are impossible to predict. The RTS "blinking" of the APS can be slow or fast, or in bursts, and one can arbitrary percept the RTS noise in APS as a bi-stable "blinking" or as a "flickering" noise, which is a counterpart of the fact that noise waveforms cannot be reconstructed from the noise spectra, due to omitting the random phase in the noise spectra.

Following from above, one can state amplidude and RMS figures-of-merit (FOMs) for the RTS noise in imagers. The amplidude FOM is the signal-to-noise ratio, which uses the amplidude $q/(WLC_{ox})$ of the RTS noise

$$\begin{aligned} \mathrm{SNR}_{\mathrm{RTS,blink\_APS}} &= \frac{\text{CDS swing } \Delta V_s}{\Delta V_{\mathrm{APS}}} = \frac{\left(I_{\mathrm{leak}} + I_{\mathrm{ph}}\right) t_{\mathrm{int}} / C_{\mathrm{PH}}}{q/\left(WLC_{\mathrm{ox}}\right)} \\ &\approx \frac{(10\mathrm{fA\ to\ 10pA}) \times (10\mu\mathrm{s\ to\ 10ms})}{3 \times 1.6 \times 10^{-19}\mathrm{C}}, \text{ with } WLC_{\mathrm{ox}} \approx \frac{C_{\mathrm{PH}}}{3} \\ &\approx 0.2 \text{ to } 200\,000 = -14 \text{ dB to } +106 \text{ dB} \end{aligned} \tag{418}$$

The calculation shows very wide range, from complete dominance (SNR=−14 dB) of the RTS noise at low photo current in small-area APS and short integration time t_{int}, to insignificant RTS noise at high photo current in large-area APS under strong illumination and long integration time t_{int}. The published data imply that the RTS noise is dominant in APS arrays with pitch lower than 3μm, in which the buffering transistors have an area lower than $(W\times L)<1\mu m^2$, and many pixels are "blinking". The problem of pixel blinking is rare for imaging arrays with pixel pitch 10μm and larger, since the buffering transistors are of large area $(W\times L)>3\mu m^2$.

The RMS FOM for RTS noise in APS uses the root-mean-square value $q/(WLC_{ox})\times(\sqrt{2}/2)$ of the RTS noise and the probability PR_{RTS} for the CDS readings in the lobes of the histogram. There is not really an established RMS FOM for the RTS noise in APS, but the FOM can be constructed, as follows

$$\begin{aligned} \mathrm{SNR}_{\mathrm{RMS,flickering\_APS}} &= \frac{\text{CDS swing } \Delta V_s}{\mathrm{RMS}\left(\Delta V_{\mathrm{APS}}\right)} \\ &\approx \frac{\left(I_{\mathrm{leak}} + I_{\mathrm{ph}}\right) t_{\mathrm{int}} / C_{\mathrm{PH}}}{\left[q/\left(WLC_{\mathrm{ox}}\right)\right] \times \frac{\sqrt{2}}{2} \times \sqrt{\left(1 - \sum PR_{\mathrm{RTS}}\right)^2 + \sum PR_{\mathrm{RTS}}^2}} \\ &\approx \frac{\mathrm{SNR}_{\mathrm{RTS,blink\_APS}}}{\sqrt{2}\sqrt{1 - \sum PR_{\mathrm{RTS}}}} < \mathrm{SNR}_{\mathrm{RTS,blink\_APS}}, \\ &\text{assuming } \sum PR_{\mathrm{RTS}} \approx \left(\sum PR_{\mathrm{RTS}}\right)^2 + \sum PR_{\mathrm{RTS}}^2 \text{ and } \sum PR_{\mathrm{RTS}} < 50\% \end{aligned} \tag{419}$$

The RMS FOM is usually a scaled version of the amplitude FOM, but the scaling coefficient (the expressions in denominator of the equation with the probability PR_{RTS} of the lobes in the histogram) is uncertainly varying around unity, and the expressions can be written differently by different authors. Again, there is not really an established RMS FOM for the RTS noise in APS, although many works use RTS time constants [352, 353] and spectra for the RTS noise [353]; thus, we do not extend further the discussion on RMS FOM for the RTS noise in APS.

Diverse approaches have been also undertaken to remedy the RTS noise in imagers, e.g., trapped charge "flush" by the reset phase followed by a short integration phase [353] and multiple sampling for CDS [354]. However, all these complicate the operation, increase the circuit overhead and the chip area of the imager, without remedying the inability of the CDS to remove RTS transitions during the integration phase. In fact, the only sustainable for the practice approach for reduction of the RTS noise in APS remains the use of larger-area buffering transistor (M2 in Figure 65 (a)), and consequently, limiting the imaging array to pixel pitch larger than 3µm.

In summary for the noise in APS, the dominant noise sources are the KTC noise due to shot noise in the photo diode, and the 1/f noise and RTS noise due to a small area of the transistor buffering the photo diode from the readout amplifier in the imaging array. The levels of these noises are inversely proportional to the photo detector and transistor areas, limiting the APS pitch to min 3-5 µm in the imaging arrays.

VIII. Outlook for the LFN

Looking about a century back, one can see that the noise in electronic devices was always in the scope of investigations, hand by hand with the development of these devices and the range of their applications. Therefore, obviously, we begin the outlook by the extraction of trends from the past.

VIII.1. Trend for the LFN level and variations

In the era of vacuum electronic devices, which is the first half of the 20th century, the fundamental relations for white noise as shot noise [355], $S_I=2qI$, and thermal noise [172, 173], $S_V=4kTR$, were established by the charge quantization of the current flow and thermodynamic approaches. The presence of flicker noise was noted, attempting to explain it by thermodynamic approaches, e.g. in [356], as it was done successfully for the white noise, and thus, it was promising approach for the flicker noise, but it was not possible. Therefore, the approach of superposition of Lorentzian noise sources was developed in the period from late 1930's [357], during 1940's [358] to late 1950's [85].

Nevertheless, the universality of thermal noise provided suitable reference, and the noise factor (or noise figure), NF – see eqs. (226) and (228), became a norm in measurement and reporting the results for 1/f noise, which continued to mid-1970's, and caused some inconvenience and delay to establish that the factors for 1/f noise are different from the factors that determine the white noise. Apart from the shape of the spectra, the main difference between white and 1/f noise is that the white noise can be described by electrical DC quantities, current for shot noise or resistance for thermal noise, while the 1/f noise is in addition inversely proportional to the size of statistical populations of carriers and/or traps. In particular, taking given densities of charge carriers and traps, the 1/f noise normalized value $S_{norm}=S/DC^2=K_F/f\propto 1/Area$ - see eq. (14), is inversely proportional to the noise generating Area. The noise generating Area is the cathode area in the case of vacuum tubes [359], the emitter junction area in the case of bipolar transistors and diodes, and the gate area in the case of field-effect transistors, as discussed in details in previous sections. Accumulating experience from about two preceding decades, experiments and analyses, such as those in [359, 360, 361], established the areal dependence of 1/f noise. This dependence was used for the design of low-noise amplifiers, in which the input transistors are large [362], despite the critical statements in some publications, e.g. in [363], that the area is irrelevant to 1/f noise, using by inertia the noise factor as a figure of merit, and overlooking that the resistance for the minimum NF increases when the Area decreases.

Thus, the factor ($Area\times K_F$), see eq. (2), can be used as the figure of merit (FOM) to compare the data for 1/f

noise in different electronic devices. The evolution of the factor (Area×K_F) and the normalized RMS noise (i_{noise}/I_{DC}) is shown in Figure 67 over almost of a century of research on 1/f noise. In this figure, the data are from [359, 364] for vacuum tubes. The data for Si, Ge and SiGe bipolar transistors is aggregated from Figure 2, Figure 14 and Figure 40, and from [365, 366, 367] for the period before year 1980. The aggregated data from Figure 15 for MOS transistors and from Figure 54 for silicon nanowires and carbon nanotubes are also shown, as well as data from [360, 362] for the period before year 1980 for MOS transistors. Data from [368, 369] for III-V semiconductor HBTs and from [370] for optical noise of light-emitting diodes are also added for comparison. The prediction lines (top to bottom) are for MOS-Analog, MOS-RF and BJT and the Moore's law (right-hand axes) are from ITRS predictions for the period 2006-2020 [3]. Please, see eq. (3) for the conversion from ITRS FOM (μmμV²) to FOM (Area × K_F). As assumed in ITRS [3], the regression line in Figure 67a is with slope (−0.5)dB/year, reflecting improvement in device structures and fabrication, and the Moore's law line is for minimum-sized MOS transistor with gate area L², where L is ½ of DRAM pitch.

The graph in Figure 67a is arranged so that the left-hand and right-hand scales are with 6 decades difference. Once the line for Moore's law crosses with regression line, or it is below a data point, then the ratio (Area × K_F)/L² is larger than 10^{-6}, indicating significant level of 1/f noise in small area devices. To illustrate, the (Area×K_F) data from Figure 67a are recalculated in Figure 67b as ratio of noise to DC currents, according to

$$\frac{i_{noise}}{I_{DC}} = \sqrt{\frac{(\text{Area} \times K_F)}{L^2}}\text{, at 1Hz in bandwidth=1Hz,} \tag{420}$$

where i_{noise} is the RMS (root-mean-square) value of the 1/f noise at 1Hz and in bandwidth of 1Hz, and L has the values from Figure 67a. The line for Moore's law in Figure 67b is from ITRS [3] for the number of MOS transistors per cm² in microprocessor integrated circuits. Interestingly, the data around year 2000, especially for BJT, imply that the ratio i_{noise}/I_{DC} follows the slope of the Moore's law for integration, that is slope +1dB/year, which means that FOM (Area×K_F) is constant, and the 1/f noise rapidly increases with the integration level to critical for device application magnitudes. This is discussed in the next paragraphs for reliability issues. Overall, as the devices become smaller, highly integrated (and faster), the 1/f noise increases [170] in ratio to the DC current, thus circuits with minimum sized transistors become noisier. From crossing the level i_{noise}/I_{DC}~0.1% at about year 2000, it follows that one does not use minimum sized devices when the application is sensitive to noise, an obvious observation in analog and RF designs nowadays. This fact is clearly seen in the trend followed by the data collected after 2000, in which the measured noise is lower than the predictions.

Having the data from a century of research on 1/f noise arranged in Figure 67, one will be able to deduce trends and predict possible issues related with 1/f noise in circuit applications. An earlier prediction for possible future scenarios was made in [29]. An observation in Figure 67a is that (Area×K_F) is relatively higher at the beginning when new devices are introduced, and later, it decreases according to the regression line. However, note two large concentrations of data points, one located around the beginning of 21st century, and the other one around the last decade. In the first one, although the data are largely scattered, the minimum values tend to be in the range (Area×K_F)=10^{-9} to 10^{-8} μm², irrespectively of the type of the device. In the second group, the minimum values have been reduced significantly, almost two orders of magnitude below the first group of data and also well below the regression line. The devices that have contributed to this reduction are NWs and CNTs, but mainly HBTs, in which the introduction of C in the SiGe semiconductor has reduced the noise significantly. The consequence is that the normalized 1/f noise i_{noise}/I_{DC} (at 1Hz in bandwidth 1Hz) is also reduced, not following

the regression line prediction, as depicted in Figure 67b.

Despite the large scattering of data, MOS transistors do seem to agree with the assumption in ITRS [3], that (Area×K_F) decreases, following the slope −0.5dB/year of the regression line in Figure 67a. Considering the data from ITRS [3], as given by the lines for MOS-Analog and MOS-RF in Figure 67a and b, and using $\int x^{-1}dx=\ln(x)$, then one can write for the RMS (root-mean-square) value of the 1/f noise that

$$\frac{\text{RMS LF Noise}}{\text{DC}} = \sqrt{\frac{(\text{Area}\times K_F)}{WL} 2.3\log_{10}\left(\frac{f_{max}}{f_{min}}\right)}, \tag{421}$$

where (Area×K_F) is the FOM taken from the ITRS prediction lines for MOS-Analog and MOS-RF in Figure 67a, the gate area (WL) depends on the application of MOS transistor and it is greater than the minimum L^2, depicted as Moore's law in Figure 67a, the constant 2.3=ln(10) follows from changing natural to decimal logarithm, and $\log_{10}(f_{max}/f_{min})$ is the number of frequency decades in the bandwidth that device application considers. Let us consider (f_{max}/f_{min})=100kHz/1Hz, 1MHz/1Hz and 10^{12}=1/(bit error rate) for analog, RF and digital applications, respectively, as shown in Table 4.

In the majority of applications, one needs the peak-to-peak magnitude of the noise. Assuming the common belief that the noise is Gaussian, as follows from central limit theorem for superposition of random events, then one can write for the peak-to-peak magnitude (LF Noise)

$$\frac{\text{LF Noise}}{\text{DC}} = 2N_\sigma \frac{\text{RMS LF Noise}}{\text{DC}} = 2N_\sigma \sqrt{\frac{(\text{Area}\times K_F)}{WL} 2.3\log_{10}\left(\frac{f_{max}}{f_{min}}\right)}, \tag{422}$$

where N_σ is the number of standard deviations that one needs to take in order to achieve a desired Confidence Probability, given by

$$\text{Confidence Probability for LF Noise} \approx \int_{-N_\sigma}^{+N_\sigma} \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx = \text{erf}\left(\frac{N_\sigma}{\sqrt{2}}\right), \text{ if normal distribution.} \tag{423}$$

For analog applications, one usually takes N_σ=2, which represents the noise "peak power" (LF Noise)² with sufficient confidence of ~95%, considering repetition and averaging of analog signals. Also, one usually uses large transistors (WL)~500×20L^2 to avoid short channel effects. Substituting the FOM (Area×K_F) from ITRS for MOS-Analog, one obtains the data shown with diamonds in Figure 68. Provided that one requires high accuracy in analog applications, e.g. not more than 0.01% error from noise, the 1/f noise causes larger uncertainty, and it is a concern for analog applications, which always consider noise reduction techniques, such as integration, averaging, offset tracking and cancelling, etc.

For RF applications, especially with the explosion of digital wireless communications, there are tough requirements for increased carrier frequencies, wider bandwidth and small channel spacing and phase instability. Therefore, one usually takes N_σ=3, which guaranties 99.7% confidence and implements error correction techniques that successfully maintain the communication channel operating at error rates of 0.1%-0.2%. The high frequency usually requires large aspect ratio W/L~1000 in MOS transistors and use of short channel devices normally at the minimum gate length L. Thus, (WL)~1000×1L^2, and substituting the FOM (Area×K_F) from ITRS for MOS-RF, one obtains the data shown with squares in Figure 68. Provided that correction of more than 0.2% error rates increases considerably the overhead in the communications channels, the 1/f became a

concern for RF applications, and after year 2005, it is always in the focus in the design of low power oscillators and mixers, since the 1/f noise degrades the phase noise in RF front-ends.

In the past, 1/f noise was not assumed to be a problem for digital applications, since the immunity of digital circuits against noise is relatively high, having SRAM Static Noise Margin (SNM) more than 10%-20% of the DC supply, as estimated in [371] for 0.5-0.8μm CMOS technologies, that is SNM~0.3−1V at V_{DD}=2−5V. Simulations of sub-50nm CMOS [372], however, indicate that the opening in the "butterfly" graph of SRAM is reduced to few tens mV, since the DC supply is less than 1V, and process variations are becoming more important. On the other hand, one relies that digital circuits do not do random error. This might be not true for circuits with very small MOS transistors, unfortunately. A normal requirement for SRAM is bit error rate of 10^{-12}, the reciprocal value of which is 12 frequency decades. One usually takes at least N_σ=4, which guaranties confidence 99.994% that double error will not occur. To save chip area, and to comply with Moore's law, one uses small MOS transistors in SRAM cells, e.g. (WL)=3×1L². Since there is no specification in ITRS for 1/f noise in digital MOS, we take the specification for MOS-RF (note that in virtually all domains of technology, RF electronics is going digital via analog-to-digital converters (ADC) and digital-to-analog converters (DAC) [373]). Then, using eq. (422), we obtain the data shown with circles in Figure 68. Provided that the SRAM Static Noise Margin (SNM) can be as low as 10%, we see from the figure that the reliability of digital circuits is also a topic of concern after year 2013 [374, 375, 376, 377]. In other words, the 1/f noise might become unforeseen limit in device down-scaling.

Connecting the points for (LF Noise/DC) with 0.01% for MOS-Analog (it was some time in year 1995), 0.2% for MOS-RF (it was circa year 2005) and 20% for MOS-Digital (around 2015), one draws the solid line in Figure 68, which illustrates the escalating issues with low-frequency noise by the device down-scaling with MOS technology nowadays. And this is on the top of the many other difficulties that accompany the device miniaturization, such as lithography, alignment, power density, wiring, accessing, volume of designs, testing, cost and complexity of the technology, etc., which are addressed in ITRS [3].

It is important to mention that average data for 1/f noise are used in the calculations shown in Figure 68. As we have presented in previous sections, and also seen in Figure 67, individual devices may have large deviations for the 1/f noise, normally 6−10dB, that is, about a decade for noise power spectrum density, or about 3-4 times for the ratio (LF Noise/DC). The issues with the scattering of the noise level around the average were put forward in the 1990's, when the device areas became less than 1μm², by publications such as [74], from which Figure 11 is adopted for illustration earlier in section III.4. "**Measurement and characterization uncertainty – experimental accuracy, fitting and averaging**". These large deviations for the 1/f noise also imply that the "average" noise does not describe completely what the noise in the devices is. Looking again at shaded areas in Figure 13, one clearly sees that the relative variation becomes more than 100% (3dB) of the "average" noise in BJTs, when the emitter area is less than 1μm², consequently, with "average" $K_F>10^{-8}$. Similar data for MOS transistors can be found in [72, 378]. In this situation, there are two-three simple, but confusing, questions: "What is the meaning of noise variation larger than average noise, since the noise power is always greater than zero?"; "Why the noise variation is greater than the noise itself, assuming that the average is representative for the noise in the devices?"; and "What we should use or do when the noise variation is greater than the average noise?". Surprisingly, these simple questions are related to the fundamental understanding for low-frequency noise in electronic structures (and other systems [160]), and to the principles and techniques of gathering

information for noise and modeling in these devices.

VIII.2. Common assumptions for noise

Let us list with A1, A2, …A6 the assumptions for noise, used in the literature, starting from fundamental to specific, and briefly discuss them from present point of view.

A1. The noise is not deterministic, that is, random variables describe mathematically the noise currents and voltages, and statistical parameters (e.g. average, variance, etc.) over a population of data (realizations, samples, time of observation) can only describe the noise.

A2. The noise is statistically invariant in wide sense, that is, the average rate (e.g. DC current, which is number of electrons passing a cross-section per unit time) and the variance (standard deviation in square, σ^2) are constants in wide sense, both in time and for identical samples and at identical conditions. This assumption is established by the first half of the 20th century, it seems for mathematical convenience in order to use the Carson theorem, as one can see in several publications, e.g. in [379, 380]. The Carson theorem states that, if identical pulses with duration, or lifetime τ, and shape F(t) occur randomly at average rate R_{avg} (with Poisson distribution), then the power spectrum density of the noise is

$$S(f) \approx 2R_{avg}|F(f)|^2, \text{ with } F(2\pi f) = \int_{-\infty}^{\infty} F(t)\exp(-j2\pi ft)dt, \tag{424}$$

where F(f) is the Fourier transformation of F(t). Taking an example for current shots ΔI due to electrons traversing a PN junction or the base of a BJT for transit time $\tau=\tau_t$ at average rate $R_{avg}=I_{DC}/q$, with q being the electronic charge, for $\tau=0$ we have Dirac current pulses $F(t)=q\delta(t)\Rightarrow F(f)=q$, and one obtains the relation for shot noise

$$S(f) \approx 2\frac{I_{DC}}{q}|q|^2 = 2qI_{DC}, \text{ with } \tau=0. \tag{425}$$

For a finite value $\tau>0$, at an additional assumption for exponential decay in the autocorrelation in the pulse [379] in order to satisfy Langevin equation $\partial\Delta F/\partial t=-\Delta F(t)/\tau+F(t)$, with $\Delta F(t)=F(t)-F_{avg}$, one gets the Lorentzian noise power spectrum density

$$S(f) \approx 4\sigma^2\frac{\tau}{1+(2\pi f\tau)^2}, \text{ with } \tau>0, \tag{426}$$

where the variance σ^2 might be difficult to be found, but it is assumed to be constant, as stated above.

A3. The noise is small in magnitude (e.g. as compared to DC), and therefore, it can be described with linear AC models. In addition, the DC component of the noise (e.g. average of noise waveform in time domain) is zero. Combined with the assumption above for constant variance σ^2, and according to Wiener-Khintchine theorem, the noise power in time and frequency domains are equal. A direct consequence is that the power spectrum and the square of noise waveform are equally representative for the noise properties. Other consequences are that the noise power is additive to any other power in the device, and the DC is always deterministic.

A4. The microscopic noise sources may have autocorrelation different from zero, but only in time. At mesoscopic scale (device level), the microscopic noise sources are independent, that is, there is no spatial, time or frequency correlation between the individual microscopic noise sources. This assumption implies that the size

of the microscopic noise sources is negligible, as compared to the distance between the noise sources or the sizes of the device. Also, combined with the assumption above for linearity, it allows for superposition of noise sources, see next.

A5. The population (the number of) of microscopic noise sources is sufficiently large from statistical point of view, and the realization of all microscopic sources is time invariant, that is, it depends on the duration of the observation (measurement), but not when the observation was made. Combined with the previous assumption, this allows to sum or integrate microscopic noise sources over any variable, e.g. along sizes of the device, energy, time constants, number of the observations (measurements), independently from integration over time or frequency associated with individual observations (single capture of data during measurement). Also, due to the additive property of the noise power assumed in A3, and according to the central limit theorem for this case, the noise in the device should be Gaussian.

A6. Following from the last assumption, continuous mathematical treatment of the noise is allowable, by using effective quantities, such as energy, velocity, mobility, time constants, etc., that represent the particle-wave-quantized nature of the solid matter on average, and in deterministic manner. Thus, the frame for noise analyses, modeling and measurement is consistently defined, and the science, physics and practice on electronic noise grows rapidly in proportion with the growth of the electronics.

VIII.3. The fabrication "frozen noise" – from spatial variations to yield problems

While the above assumptions (at the end of the previous sub-section) draw a coherent picture from physical and mathematical points of view, the device downscaling actually has questioned several of them. Some of the issues are now discussed. In this discussion, we pay particular attention to the numbers, because they are where the problems usually come from when dealing with random quantities, thus, with statistics for noise.

VIII.3.1. Length uncertainty variances

Please, look again at Figure 67a. Except for the uncertain data (solid circles) for nanowires, when their cross-section area is used to calculate the figure of merit (Area×K_F), all other data are above (Area×K_F)=10^{-9}μm², mostly in the range of 10^{-8}μm². Then, we calculate from eq. (421) a characteristic distance σ_d of random uncertainty for a frequency decade, taking f_{max}/f_{min}=10, (RMS LF Noise)/DC=1, and setting WL=$(\sigma_d)^2$. So, we get

$$\sigma_d = \sqrt{2.3(\text{Area}\times K_F)} \sim \sqrt{2.3\times10^{-8}\,\frac{\mu m^2}{\text{decade}}} = (0.15 \geq 0.047)\frac{\text{nm}}{\sqrt{\text{decade}}}. \qquad (427)$$

The meaning of σ_d is that after 10 transitions between atoms in the semiconductor, the electron (or hole) carrier will be displaced at distance σ_d from the ideal position, which would be, if no randomness exists in electron motion, and the electron travels "smoothly" at the deterministic velocity and trajectory, which one can calculate by the continuous equations, e.g. for drift and diffusion, for example. For comparison, the atomic radius of silicon is 0.11nm and the average atomic distance in silicon crystals is 0.27nm. Thus, the ten transitions correspond to 2.7nm travelling distance, and for a given length L of the device, the uncertainty length σ_{L1e} for one electron traversing the device is

$$\sigma_{L1e} = \sigma_d\sqrt{\log_{10}\left(\frac{L}{2.7\text{nm}}\right)} = \sqrt{2.3(\text{Area}\times K_F)}\sqrt{\log_{10}\left(\frac{L}{2.7\text{nm}}\right)}. \qquad (428)$$

Here, we note that the number of transitions, which the independent electrons do, when traversing the device, is proportional to the product of number of electrons (n) in the sample and to the ratio t/τ_t of the time of observation t and transit time τ_t. Therefore, the uncertainty length σ_{acc} accumulated from all electrons during the observation is

$$\begin{aligned}\sigma_{acc} &= \sigma_{L1e}\sqrt{\log_{10}\left(n\frac{t}{\tau_t}\right)} = \sigma_d\sqrt{\log_{10}\left(\frac{L}{2.7nm}n\frac{t}{\tau_t}\right)} \\ &= \sqrt{2.3(Area\times K_F)}\sqrt{\log_{10}\left(\frac{L}{2.7nm}n\frac{t}{\tau_t}\right)}.\end{aligned} \tag{429}$$

Note that the logarithmic functions are result from the constraint that the noise is 1/f, and they do not imply particular physics. We will discuss this later by eq. (438).

On the other hand, one "sees" all (n) electrons, when measuring the device, so, the measurement is integration over the number of electrons, and the observed length uncertainty variance is

$$\begin{aligned}\sigma_L^2 &= \frac{\sigma_{acc}^2}{n} = \frac{\sigma_d^2}{n}\log_{10}\left(\frac{L}{2.7nm}n\frac{t}{\tau_t}\right) \\ &= \frac{(Area\times K_F)}{n}2.3\log_{10}\left(\frac{L}{2.7nm}n\frac{t}{\tau_t}\right) \\ &= \frac{(Area\times K_F)}{n}\ln\left(\frac{n}{2.7nm}\nu t\right),\end{aligned} \tag{430}$$

where $\nu=L/\tau_t$ is the carrier velocity. The differentiation against the measurement time t gives

$$d\left(\sigma_m^2\right) = \frac{(Area\times K_F)}{n}d[\ln(t)] \tag{431}$$

We use now that 1/t corresponds both the resolution df and the minimum non-zero frequency f_{min} in the spectrum in frequency domain.

$$\begin{aligned}&d\left(\sigma_L^2\right) = \frac{(Area\times K_F)}{n}d\left[\ln\left(\frac{1}{f_{min}}\right)\right] = -\frac{(Area\times K_F)}{n}d[\ln(f_{min})]\Rightarrow \\ &\frac{d\left(\sigma_L^2\right)}{-df} = S_L(f_{min}) = \frac{(Area\times K_F)}{f_{min}n}\end{aligned} \tag{432}$$

Here, S_L is the power spectrum density of the length uncertainty σ_L, as promoted by random deviation of electron motion from deterministic trajectory, and the minus sign is due to the fact that the longer is the time of the observation, the lower is f_{min}, and (−df) denotes increase in the resolution in frequency domain. Since the observation time can be chosen with various values, then we obtain 1/f noise, given by

$$S_L(f) = \frac{(Area\times K_F)}{f\ n} = \frac{\sigma_d^2}{2.3\ f\ n}, \tag{433}$$

$$\text{with } \sigma_d \approx \begin{cases} 0.05\text{nm (1/2 atomic radius), when } (\text{Area}\times K_F) = 10^{-9}\,\mu\text{m}^2 \\ 0.15\text{nm (between atomic radius and atomic distance), when } (\text{Area}\times K_F) = 10^{-8}\,\mu\text{m}^2 \\ 0.5\text{nm (almost lattice constant), when } (\text{Area}\times K_F) = 10^{-7}\,\mu\text{m}^2 \end{cases}$$

and σ_d being random displacement of electron (uncertainty of electron position) after 10 transitions between atoms in the semiconductor, as defined by eq. (427). The total number of carriers n=n'WL is proportional to the carrier concentration (n') and device area WL, e.g., as in MOS transistors, for example.

Surprisingly, eq. (433) is for spatial noise, the unit is area/frequency, and this unit is in direct disagreement with assumption A4 above, but on the other hand, the values for the random displacement of electron σ_d are meaningful. For example, the momentum of electrons, p_{th}, due to thermal motion in silicon is

$$p_{th} = \nu_{th} m_o m_e = 10^5 \left[\frac{\text{m}}{\text{s}}\right] \times 9.11\times 10^{-31}[\text{kg}] \times 0.29 \left[\frac{\text{kg}}{\text{kg}}\right] = 2.46\times 10^{-26}\left[\text{kg}\frac{\text{m}}{\text{s}}\right], \tag{434}$$

where, with the values given above, ν_{th} is the thermal velocity, m_o is the electron mass, and m_e is the relative electron mass for conductivity calculations. From Heisenberg uncertainty principle, we have

$$\begin{aligned} &\frac{h}{4\pi} \le \sigma_{p_{th}} \sigma_{th} \Rightarrow \frac{h}{4\pi p_{th}} \le \frac{\sigma_{\nu_{th}} m_o m_e}{\nu_{th} m_o m_e}\sigma_{th} = \frac{\sigma_{\nu_{th}}}{\nu_{th}}\sigma_{th} = \sigma_{th}\frac{\text{Noise in } \nu_{th}}{\nu_{th}} = \sigma_{th} \\ &\sigma_{th} \ge \frac{6.63\times 10^{-34}[\text{Js}]}{12.6\times 2.46\times 10^{-26}\left[\text{kg}\frac{\text{m}}{\text{s}}\right]} \approx 2\text{nm}, \end{aligned} \tag{435}$$

where, with the values given above, we use that ν_{th} is the same as RMS of the thermal velocity (the thermal motion is random), σ_{th} is the uncertainty for the position of the electron, and h is Planck constant. One "sees" the thermal noise only in the direction of the current flow, e.g. along x-coordinate, so, in this direction, the instantaneous uncertainty for the position of the electron is

$$\sigma_{xth} = \frac{\sigma_{th}}{\sqrt{3}} \approx 1.47\text{nm}, \tag{436}$$

but after 10 transitions, which is equivalent to average of ten, or reduction of the bandwidth 10 times, one expects that the electron is closer to its average equilibrium position, thus, the uncertainty for the electron position should be

$$\sigma_{dth} = \frac{\sigma_{xth}}{\sqrt{10}} \approx 0.36\text{nm}. \tag{437}$$

By comparing to eq. (433), we observe that the values for σ_d and σ_{dth} are in the same order of magnitude, which justifies the values for the random displacement of electron σ_d, that causes 1/f noise. The smaller values for σ_d indicate that the 1/f noise might be "smoother", or more deterministic, than the thermal noise in short time scale, which is obvious from practical point of view, since the power spectrum density of the 1/f noise decreases at high frequencies.

Note that the variance σ_{acc} in eq. (429) for 1/f noise increases with the time of observation t, whereas σ_{dth} in eq. (429) is independent of t. Qualitatively, this is similar to what is established by Allan variance, recalling the

corresponding eqs. (372) and (355) and the discussion on them. The increase of the variance for 1/f noise is in contrast with the assumption A2 that the noise is statistically invariant in wide sense, from which it followed that the variance is constant in time. Also, the range for σ_d is at atomic distances, which does not lend much confidence to assumption A6 that the particle-wave-quantized nature of the solid matter can be neglected in noise analyses and models. Explanation of the 1/f noise in terms of quantum effects was attempted [381], and reviewed together with other fundamental 1/f noise sources [382], but the quantum 1/f noise was also strongly criticized [2], by objecting the assumption for the "power loss" introduced in the wave equations for the cases beam in vacuum and metal at zero absolute temperature; and at present, the concept for quantum 1/f noise is not followed up widely, except for special cases, e.g. for lasers, by different mathematical treatment. Later works [383, 384, 385], however, indicated that vibration-wave nature of the crystals can produce 1/f noise, and, perhaps, there will soon be a reassessment of the approach for quantum noise, since some of the down-scaled devices are with dimensions similar to quantum devices, such as quantum dots and wires (e.g. CNT), and the noise in these devices is very large, see again Figure 51.

VIII.3.2. Consequences from the length uncertainty variances

Now, we discuss the consequences from the length uncertainty variances and 1/f noise, which were introduced by eqs. (427) to (433). First, we observe that the logarithmic function disappeared at the step of differentiation after eq.(430), hiding all device and bias dependent quantities in the logarithmic function, but giving a rise to 1/f noise for the power spectrum density in eq. (433) by the observation time t. Otherwise, eq. (433) is in very regular form, as expected, and we will use this after the following remark.

Instead of σ_d, the random displacement of electron (uncertainty of electron position) after 10 transitions, one can choose characteristic standard deviations σ_{Fo} for different quantities F in eq. (427), e.g., F can be mobility, carrier concentration, velocity, etc., and a spectral density S_F for the corresponding quantity will be obtained in eq.(433), as long as the variance of the chosen quantity is a logarithmic function of the time. This is nothing, but to write

$$\sigma_F(t) = \sigma_{Fo} \ln\left(\frac{t}{t_{Fo}}\right) \Leftrightarrow \frac{d\sigma_F}{dt} = \frac{\sigma_{Fo}}{t} \Leftrightarrow \exp\left[\frac{\sigma_F(t)}{\sigma_{Fo}}\right] = \frac{t}{t_{Fo}}, \tag{438}$$

where the characteristic standard deviation σ_{Fo} and reference time constant t_{Fo} has to be chosen by physical and other reasons. Obviously, the middle equation implies 1/time distribution for σ_F. For the physically plausible relaxation processes, in which $F(t)/Fo=\exp(-t/\tau)$ occurs randomly at average rate R_{avg}, one has Lorentzian noise spectrum, according the Carson theorem – see again eqs. (424) and (426), but by distributing the variance as $1/\tau$, also assuming proper time for observation, one obtains the superposition integral for 1/f noise from bistable RTS fluctuation, as we have used many times in the previous sections – see eqs. (72)–(74) in section III.3 for BJT and eqs. (97) and (100) section IV.1 and later for MOS transistors, for example. Indeed, distributing amplitudes as 1/time at sufficiently high average rate will produce also $\sigma_F \propto \log(t)$, and therefore, 1/f noise spectrum. Thus, one has to be careful using the observation of 1/f noise as justification for existence of particular random phenomena in the device, the justification should be provided by other method too. The 1/f noise only provides that the variance of the fluctuation is a cumulative function of time, in particular a logarithmic function, if the spectrum is exactly with slope 1/f. Other conclusions require additional physical justification.

The regular form of eq.(433) allows to arrange it in normalized form, by adding it as another term S_Z/Z^2 to the canonic form of the Hooge equation (6).

$$\frac{S_I}{I_{DC}{}^2}=\frac{S_{Qn}}{Q_n{}^2}=\frac{S_\mu}{\mu^2}=...=S_{norm}=\frac{\alpha_H}{n\ f}=\frac{\sigma_d^2}{2.3WL\ f\ n} \tag{439}$$

Here, n is the total number of carriers, α_H is the Hooge parameter, and WL is the area of the device, in which the 1/f noise is generated. Surprisingly, α_H becomes a function of the area of the device, as

$$\alpha_H=\frac{\sigma_d^2}{2.3WL}=\frac{(\text{Area}\times K_F)}{WL}\sim\begin{cases}10^{-9},\text{for }(\text{Area}\times K_F)=10^{-9}\,\mu m^2\text{and WL}=(1\times1)\mu m^2\\10^{-7},\text{for }(\text{Area}\times K_F)=10^{-8}\,\mu m^2\text{and WL}=(1\times0.1)\mu m^2\\10^{-5},\text{for }(\text{Area}\times K_F)=10^{-7}\,\mu m^2\text{and WL}=(0.1\times0.1)\mu m^2\end{cases} \tag{440}$$

spanning pretty well the "good" values for α_H reported in the literature. One of the questions from the past, whether one should use the number of atoms, or number of carriers in Hooge equation, and answered in favor of number of carriers [2] in the 1980s, seems is not completely answered. The noise in few-nanometer-sized devices may help to find whether the sample size matters for Hooge parameter.

VIII.3.3. Fabrication frozen noise. Definition

Another difficult problem, which emerged in sub-100 nm devices, is the "contrast" of doping profiles and lithography [3]. For maximum resolution, the doping profiles are achieved by ion implantation, usually using the gate mask or the polysilicon gate itself as the implantation masks, with the purpose of self-alignment in MOS transistor fabrication. To minimize the drain-induced barrier lowering (DIBL), but not increasing the threshold voltage [252], the MOS transistor body doping profiles are retrograded, or even δ–doped, with a peak concentration in the range of 10^{18} cm^{-3} at depth around 30-60nm (or less) from the gate oxide interface.

We assume that Poisson distribution will hold for the ideal case of ion implantation, and several δ–doping profiles are given in Figure 69 on left. With every next generation, the depth of the δ–doping needs to decrease, so that the depletion distance t_{dep} is reduced to minimize DIBL, in order to satisfy the semi-empirical rule for minimum gate length [386]

$$L_{GATE,min}=A\ \sqrt[3]{t_j t_{ox} t_{dep}^2}\,, \tag{441}$$

where t_j is the depth of source and drain junctions, t_{ox} is the thickness of gate dielectric (silicon dioxide), and A is a fitting constant. Assuming that t_{ox} might be difficult to scale down nowadays, and that $t_j \propto t_{dep}$, then the depletion distance t_{dep} needs to scale proportionally with gate length. For δ–doping, from electrostatics of point charges, we also assume that the effective depletion depth t_{dep} at each atomic column under the gate is $(1/t_{dep})^2 \sim \sum(1/t_{ia})^2$, where t_{ia} is the implantation depth of individual impurity atoms. Then, we have generated random numbers with Poisson distributions that correspond to the desired t_{dep} in atomic distances (0.2715nm for Si), and also, corresponding to peak concentration in 10^{18} cm^{-3} at t_{dep}. Sample results are shown in Figure 69 on right. In these figures, the dots are the impurity atoms, the grid corresponds to the atomic distances in silicon, and the lines correspond to the values for the effective depth at each atomic column under the gate dielectric, obtained from abovementioned $(1/t_{dep})^2 \sim \sum(1/t_{ia})^2$. We observe strong departure from continuity toward quantization, decreasing L_{GATE} from 90nm to 12nm, accompanied with increase of the normalized standard deviation (σ_{dep}/t_{dep}) from 7% to 20% for the effective depth.

Since several parameters in MOS transistor, e.g. the threshold voltage, are proportional to t_{dep} to the first order of

approximation, then (σ_{dep}/t_{dep}) will be replicated as variation in voltage overdrive and current flow in the MOS transistor. We will present an analysis later, beginning from eq. (443). Further, in smaller size transistors, the regions of the current transport (graded areas in Figure 69 on right) become very close to the lines of the effective depth t_{dep}, which implies that the variation in t_{dep} will be replicated as variation in the current transport, thus, in noise. Since the variation in t_{dep} is due to fabrication, then we have this variation embedded in the device, thus the noise is fixed, or "frozen noise", in other words. The fabrication "frozen noise", in particular the increase of the normalized standard deviations (σ/average) in MOS transistors with L_{GATE}<20-30nm, is usually observed as increased leakage and gain reduction, smear in the device I–V curves, resulting in parameter variations from device to device and between circuits [372, 387], which should be nominally identical. The "frozen noise" from fabrication, perhaps, also increases the temporal low-frequency noise, which at present is not investigated for such relation, unfortunately.

VIII.3.4. Investigation of fabrication frozen noise

The investigations on the fabrication "frozen noise" are usually performed in three ways, namely, Monte-Carlo simulations, statistical evaluation of large number of measurements of identical samples, and theoretical extrapolation of models toward small devices. Each of these approaches provides valuable insights on the "frozen noise", but also has specific limitations.

Monte-Carlo simulations

In the Monte-Carlo simulations, one usually takes an ideal structure and adds particular randomness in the structure. Then, the device or the circuit are simulated, and from the obtained characteristics, the variations of the model parameters or characteristic quantities are estimated for magnitude, correlations and behavior. It is believed that 3D simulation of structures described at atomic level, with embedded drift-diffusion and gradient formalism in the simulation, should provide correct and sufficient information, which reflects the "frozen noise". Typical results from 3D simulations of few-nm to few-deca-nm structures are shown in Figure 70. In Figure 70a, the strong impact of single traps on carrier velocity in MOS transistor is evaluated [388, 389] for square-shaped bulk MOS transistor shown on the top of Figure 70b, where also is evaluated that the single-trap occupancy changes significantly (~18-20%) the current in 30nm×30nm transistor, and this variation is "dramatically" larger, ~50%, for 10nm×10nm transistor, thus causing unacceptably large RTS noise. In Figure 70c, the impact of 1-2 atomic layers random non-uniformity in the gate oxide and body of ultra-thin-body SOI MOS transistor (this is one of the mostly spelled alternatives for MOS transistors in future [3]) on the BSIMSOI model parameters is evaluated [372]. The scatter in the values is evident in Figure 70c, while the correlation between different model parameters is weak. In this way, the Monte-Carlo and atomistic simulations provide reach sets of values, implying forthcoming difficulties with the increasing "frozen noise", but one should note that interpretation of the data is difficult and not always very certain, because, first, the assumed structure for simulations can be unrealistic or not representative for real structures, second, the simulations generate sets of large data, which requires post-processing for extraction of statistical or other parameters, and third, the physics is beyond this point, since the numerical data sets do not "say" what, where and how the numbers depend on the structure. This is a drawback of computer simulations in principle, and it requires somewhat empirical search for correlations in the numerical data, which eventually can be related to models and physics. In addition, if one tries to obtain temporal variations and noise from them, the simulation becomes 4D, the data volumes and computation overhead become too large to be feasible for the practice. Finally, one should always keep in mind

that the quantum-wave nature of the matter, important at the few-nm scale, is lumped in continuous gradients for carrier concentrations in order to avoid the 5th D in the simulations, which would make the simulations impossible even for super computers.

Statistical evaluation of large number of measurements of identical samples

The experimental investigations on the fabrication "frozen noise" have the advantage that they are practical, although vulnerable to experimental uncertainties both due to sample fabrication and measurement errors. The obvious and main advantage of the experimental investigations is that they are "real", but not "virtual", as the simulations are. On the other hand, there are three barriers for the experiments, as compared to simulation approaches. First, the technology for sample preparation has be physically available and accessible, thus, "future" devices cannot be experimented now. Second, the devices cannot be interactively tailored, e.g. one cannot change doping profiles, oxide thicknesses etc., either because the fabrication process does not allow for this, or because it is too expensive or takes too long to do it, thus, in the space of experimental conditions, the fabrication parameters are at few or even in one fixed point, and in addition, many details from fabrication process are not available as information for experimentalist, either due to policies for proprietary information by the fabricator, or just because the fabrication does not keep track for the particular parameter. The third barrier for the experimental investigations on "frozen noise" is that the experiments are mesoscopic, that is, the device is seen from its terminals, but not inside, and one has to provide electrical access from measurement instruments to the nanostructure terminals, which, however, is accompanied with many problems for contacting the device and de-embedding the parasitics from probes, cables and etc. in the experimental setup; certainly, the simulations are completely free from the third barrier. Nevertheless, by mapping several wafers with identical devices or circuits, there are publications that experimentally characterize the "frozen noise" [387, 390, 391]. One illustrative insert from [387] is shown in Figure 71 for a 90nm CMOS SOI fabrication process, the "frozen noise" in which was evaluated using ring oscillators. The contribution of the "frozen noise" was attributed to several factors (variation in transistor threshold voltages, capacitance, resistance, interconnection, RTA) and these factors were divided in three types –variation between groups of identical ring oscillators in several chips from 2-3 wafers, and correlated and uncorrelated variations within the groups. The horizontal axis in Figure 71 for the average delay in the groups represents the first type of factors for variation between groups and chips, and the span of the inter-chip variation in this axis is about 8ns around 20ns, that is, ±20% max, or ≈10% standard deviation. The correlated and uncorrelated portions of in-group variation were deduced in the following way. First, only the data for maximum and minimum delays of individual ring oscillators in each group are considered, and two regression lines are found, one for maximum and one for minimum delays, as shown in the figure. These lines appeared to be linear functions of the average delays in the groups, and therefore, the vertical separation of about 1.6ns represents the correlated variation, which is ±2% max, or ≈1% standard deviation from the average delay of 20ns of all samples. Then, the 0.6–0.8ns scattering of the data around the regression lines represents the uncorrelated (random) variation between individual ring oscillators, which is less than 1% standard deviation from the average delay of 20ns of all samples. A close look at the publication also indicates that number of samples in this investigation is large, in the range of several thousands, and also, a variation and track to the fabrication conditions were made in order to compare to Monte Carlo simulations based on predetermined models (namely "Ring Oscillator Delay Model to Hardware Correlation", and "Back-End-of-Line" interconnect model, along with Layout-vs-Schematic, SPICE simulator and BSIMSOI model with and without lattice stress equations), which indicated that the experimental statistical characterization of the "frozen

noise" is quite costly, and possibly, accessible only to industrial research groups from foundry companies – well, the publication [387] is from IBM Semiconductor Research and Development Center.

Theoretical extrapolation of models toward small devices

The third way in the investigations on the fabrication "frozen noise" is the theoretical extrapolation of models toward small devices, as mentioned earlier. The models are usually very generic, and they are based on standard deviation of transistor sizes and trends deduced from publications. Since the most of the parameters occur in products or ratios in the device equations, then, the total standard deviation becomes a quadratic sum of the normalized standard deviations of the particular set of factors for the "frozen noise"; that is for MOS transistor, for example,

$$\frac{\sigma_{tot}^2}{I_D^2} = \frac{\sigma_W^2}{W^2} + \frac{\sigma_L^2}{L^2} + \frac{\sigma_{C_{ox}}^2}{C_{ox}^2} + \frac{4\sigma_{V_T}^2}{(V_G - V_T)^2} + \ldots, \tag{442}$$

since in saturation regime

$$I_D \approx \frac{W}{2L}\mu C_{ox}(V_G - V_T)^2 \Rightarrow \frac{dI_D}{I_D} = \frac{dW}{W} - \frac{dL}{L} + \frac{dC_{ox}}{C_{ox}} + 2\frac{dV_T}{(V_G - V_T)} \pm \ldots,$$

and eq. (442) can be further "enhanced" or "reduced" with cross-correlation terms by particular assumptions for systematic dependences, e.g. the etching affects in the same manner W and L of the gate, so the ratio W/L will be affected mostly by the variation of the gate length L as $(\sigma_{etch}/L)^2$, and this effect will be compensated partially with $[-(\sigma_{etch}/W)^2 c^2 (L/W)^2]$ by variation of the gate width W in same direction, assuming that the etching error σ_{etch} is systematic for small transistors and, therefore, the correlation coefficient $c \approx 1$. One particularly important conclusion from eq. (442) is that the processing variations need to be proportionally reduced by device down-scaling, e.g. $\sigma_{etch} \propto L_{min}$, $\sigma_{tox} \propto t_{ox}$ …, so that the total "frozen noise", (σ_{tot}/I_D in the equation) is maintained unchanged. In this way, the extrapolation models provide easily an estimate for what would be necessary in order to fabricate operational nano-devices. It is another question, however, whether this will be feasible for mass-production at the growing technical, financial and other constraints.

The answer of the later question is in the statistics of the random numbers, therefore we address it as noise, in relation to the aforementioned first assumption A1 for noise, but once the device is fabricated, then the statistics is fixed in the device spatially, that is, the noise is frozen, in contrast to the other assumptions that the noise is in time-frequency domain. (From general conservation principles in physics, one has to consider conservation of mass and charge, rather than only for power, as stated in assumption A2, questioning us whether some coordinates in the noise space are overlooked, conservatively for the "frozen noise".)

As mentioned above, now we present an analysis that uses the theoretical extrapolation approach for the impact of the "frozen noise" in δ-doping on the yield, in order to illustrate the importance of the statistics of the random numbers. Recall the values for depletion depth, t_{dep}, and its standard deviation σ_{dep} from Figure 69. For each implanted atom, at the assumed Poisson distribution for the implantation, so that $\sigma_{dep} = \sqrt{t_{dep}}$, these values are

$$\begin{aligned}
&L = 90\text{nm} \rightarrow t_{dep} = 200\,\text{atomic distances} \rightarrow \sigma_{dep}/t_{dep} = 7.07\% \\
&L = 45\text{nm} \rightarrow t_{dep} = 100\,\text{atomic distances} \rightarrow \sigma_{dep}/t_{dep} = 10.0\% \\
&L = 22\text{nm} \rightarrow t_{dep} = 50\,\text{atomic distances} \rightarrow \sigma_{dep}/t_{dep} = 14.1\% \\
&L = 12\text{nm} \rightarrow t_{dep} = 20\,\text{atomic distances} \rightarrow \sigma_{dep}/t_{dep} = 22.4\%
\end{aligned} \tag{443}$$

The corresponding average rates of occurrence of impurity atoms per atomic column are 0.963, 0.681, 0.482 and 0.306, and these correspond the desired peak concentration of 10^{18} 1/cm³ in Figure 69 on left. Taking square shaped transistors, the total number of impurities N_{imp} in these transistors, on average, will be

$$\begin{aligned}
&WL = 90\text{nm}\times 90\text{nm} \rightarrow N_{imp} \approx 106\,000 \rightarrow \left(\sigma_{dep}/t_{dep}\right)_{avg} \approx 0.022\% \\
&WL = 45\text{nm}\times 45\text{nm} \rightarrow N_{imp} \approx 19\,000 \rightarrow \left(\sigma_{dep}/t_{dep}\right)_{avg} \approx 0.073\% \\
&WL = 22\text{nm}\times 22\text{nm} \rightarrow N_{imp} \approx 3\,200 \rightarrow \left(\sigma_{dep}/t_{dep}\right)_{avg} \approx 0.25\% \\
&WL = 12\text{nm}\times 12\text{nm} \rightarrow N_{imp} \approx 600 \rightarrow \left(\sigma_{dep}/t_{dep}\right)_{avg} \approx 0.92\%
\end{aligned} \tag{444}$$

Here, the average standard deviation in the device is $[(\sigma_{dep}/t_{dep})_{avg}]^2=(\sigma_{dep}/t_{dep})^2/N_{imp}$, and this deviation is only due to Poisson distribution in ion implantation for δ-doping. The threshold voltage of the transistors, on the other hand, is given by [252]

$$\begin{aligned}
&V_T = V_{FB} + 2\varphi_F + \frac{C_d}{C_{ox}}2\varphi_F \approx V_{FB} + 2\varphi_F + \frac{1}{3}2\varphi_F \approx 0.2\text{V} - 0.4\text{V} \\
&\Rightarrow V_T \approx \frac{1}{3}2\varphi_F \sim \frac{1}{3}0.8\text{V}, \text{and } V_T \propto t_{dep} \Rightarrow \frac{dV_T}{V_T} \approx \frac{dt_{dep}}{t_{dep}},
\end{aligned} \tag{445}$$

since the ratio of depletion capacitance C_d to oxide capacitance C_{ox} is approximately C_d/C_{ox}~1/3 for MOS transistors nowadays, the Fermi potential in the body is in the range φ_F~0.4V for impurity density ~10^{17} 1/cm³ in these transistors, and also, the depletion depth t_{dep} is approximately equal to the depth of the δ-doping [252]. Therefore, the average standard deviation for the δ-doping is replicated as standard deviation for the threshold voltage of the transistors, and from eq. (444), we have

$$\begin{aligned}
&WL = 90\text{nm}\times 90\text{nm} \rightarrow \sigma_{V_T} \approx 0.022\% \\
&WL = 45\text{nm}\times 45\text{nm} \rightarrow \sigma_{V_T} \approx 0.073\% \\
&WL = 22\text{nm}\times 22\text{nm} \rightarrow \sigma_{V_T} \approx 0.25\% \\
&WL = 12\text{nm}\times 12\text{nm} \rightarrow \sigma_{V_T} \approx 0.92\%
\end{aligned} \tag{446}$$

With these values, the results from calculations using the extrapolation approach are illustrated in Figure 72. The different lines in this figure correspond to the different (W×L) sizes of the transistors (90nm×90nm), (45nm×45nm), (22nm×22nm), and (12nm×12nm), denoted with diamonds (◆), squares (■), triangles (▲) and circles (●), respectively. Figure 72a shows that at given limit for tolerance in the threshold voltage (horizontal axes in the figure), the ratio of σ_{V_T} to the tolerance for V_T increases as the device size decreases, since the number of σ_{V_T} for the given tolerance decreases. Figure 72b shows that decreasing the device size, one has to relax the design rule for variation of V_T from 0.3% for (90nm×90nm) devices to more than 10% for

(12nm×12nm), in order to achieve probability 10^{-12} for failure due to exceeding the allowed tolerance for V_T. Figure 72c shows that the impact of randomness in δ-doping is negligible for SRAM built on (90nm×90nm) transistors, since 0.25% variation of V_T is well below variations ~3% in V_T caused by other processing factors, while 3.25% variation of V_T is comparable to that variations in (22nm×22nm) transistors, which correspond roughly to 45nm MOS technologies. The required limit of 12.5% for the variation V_T in (12nm×12nm) transistors is, however, too large. Therefore, bulk MOS is generally rejected as an option for commercial MOS technologies after node 32nm [3], in which the minimum gate length should be 11nm, or less.

VIII.4. Statistical accumulation of variance "innovates" 1/f noise

The above discussion showed that the "frozen noise" is becoming a limiting factor for device downscaling, since the small populations in small devices cause larger relative variations between the devices. The question is whether the "frozen noise" also affects the temporal low-frequency noise, because all rules for the low-frequency noise imply an increase as inverse of device area. In order to answer this question, we plot in Figure 73 the "Spot frozen noise" for the implantation depth from eq. (443) and the "Device average frozen noise" for V_T from eqs. (444) and (446) together with the values for low-frequency noise predicted in ITRS [3], shown earlier in Figure 67, against the minimum device area. The observation in Figure 73 is that LFN and the "Spot frozen noise" have the same areal dependence, which implies that the low-frequency noise replicates the spatial "Spot frozen noise", but in time domain. The fact that the low-frequency noise is larger in non-uniform devices is obvious and well established semi-empirically in the 1980's – see [9], for example. The physical explanation of this fact, however, meets with difficulties, since no general theoretical approach to the problem is available.

In the origin of the problem is that the charge transport in electronic devices, in particular the transit time for electrons and holes traversing the device, is much faster than the frequency range of low-frequency noise. Therefore, other phenomena that interfere with the charge flow are attributed to the low-frequency noise, but not the charge flow itself. At present, two "schools" attribute the LFN to scattering (mobility fluctuation noise) and to trapping (number fluctuation noise). The mobility noise is usually related to the "slow" phonon scattering, see [70], for example, where, however, it was also mentioned that this might be equivalent to assume phonon number fluctuation. The number fluctuation is usually related to "slow" charge trapping in dielectric (e.g. gate oxide) with distribution of the time constants $\tau \propto 1/\tau$, and then it modulates the number of carriers in the conductive channel, replicating the distribution, and generating 1/f noise after superposition, as we have discussed several times in previous sections. However, the traps may cause also modulation of mobility, thus, again we have convergence to mobility noise. What is common in the two approaches, is that the 1/f noise is explained by distributions of time constants, strictly following assumptions A3 and A4 for small noise amplitudes and correlations only in time domain, see again the end of sub-section VIII.1. "**Trend for the LFN level and variations**", and many useful models are developed, as we have discussed in the previous sections. There are also other useful relations, such as

$$\frac{\alpha_{H,eff}}{\mu_{eff}^2} = \frac{\alpha_{H,latt}}{\mu_{latt}^2} + \frac{\alpha_{H,imp}}{\mu_{imp}^2}, \tag{447}$$

which relates the measured effective Hooge parameter $\alpha_{H,eff}$ and mobility μ_{eff} to Hooge parameters $\alpha_{H,latt}$ and $\alpha_{H,imp}$ and mobility μ_{latt} and μ_{imp} of lattice and impurity scattering – see again [70], for example, or condition for dominance of single RTS over 1/f noise [170, 392]

$$n < \frac{1}{4\pi\alpha_H}, \tag{448}$$

where (n) is number of carriers in the conductive channel, by assumption of single electron trapping, thus $\sigma_{\Delta n}/n=0.5/n$. However, there is no clear answer whether or how the spatial non-uniformity is (or not) in the origin of the distributions that generate 1/f noise. On the other hand, eqs. (428), (429) and (430) indicated a logarithmic dependence in the variance that can be obtained from spatial "frozen noise" and causes temporal 1/f noise, since this dependence is converted into 1/f noise power spectrum in eq. (433).

VIII.5. <u>Statistical origin of the temporal 1/f noise in relation with the spatial "frozen noise"</u>

We now discuss the statistical origin of the temporal 1/f noise in relation with the spatial "frozen noise", the latter probed by the fast, and thus with regular average rate, current flow, considering again the observation in Figure 73 that LFN and the "Spot frozen noise" have the same areal dependence, which implies that the low-frequency noise replicates the spatial "Spot frozen noise", but in time domain. For convenience, the standard deviation of the "Spot frozen noise" we will term with spatial "Roughness" and denote with σ_r. The spatial "Roughness" corresponds to the quantity (σ_{dep}/t_{dep}) in eq. (443), for example, e.g. $\sigma_r=(\sigma_{dep}/t_{dep})$. When the electron traverses the device, it also probes and accumulates the variations in the structure, e.g. as variation in velocity. Since the variances are statistically additive, then one writes for the electron at its exit from the device after K steps that

$$\sigma_K^2 = \sigma_r^2\left(c_K^2 + \sum_{k=1}^{K-1} c_k^2\right), \tag{449}$$

with c_k being regression coefficients that describe the transfer of the random "Roughness" into variance $(\sigma_k)^2$ at each instance k=1…K of probing, e.g. atomic position along the channel length of the MOS transistor for the "frozen noise", or scattering event in mobility models, or sampling in signal processing. Eq. (449) is a basic relation in the statistics of the so-called "innovation variance" [393, 394, 395]. The "innovation variance" was introduced as prediction for the variance in time series at the output of a correlating filter in [396], where also important finding for logarithmic spectra are given, and we will use soon. A simple introduction to "innovation variance" is given in [397], from which we insert that for a random sequence $\{X_1, X_2, \ldots, X_K\}$ with, let say, first order (one-step) auto-regression with coefficient c, one has

$$X_K = RND\_INNOV + c\ X_{K-1}, \tag{450}$$

where RND_INNOV is an uncorrelated random "innovation" added to X at step K, and RND_INNOV has variance $(\sigma_{innov})^2$, then the "innovation" is also added to the average at the previous step (K-1), and, thus to the previous random "innovations". So, the random variation ΔX_K of X_K from the average is

$$\Delta X_K = X_{K-1} - \frac{1}{K-1}\sum_{k=1}^{K-1} X_k, \tag{451}$$

and the variance $(\sigma_K)^2$ for ΔX_K is

$$\sigma_K^2 = \sigma_{innov}^2 + \sigma_{K-1}^2 = \frac{\sigma_{innov}^2}{1-c^2}. \tag{452}$$

The last term is valid for large K and $|c|<1$, owing to the convergence the power series $\sum(c^2)^k=1/(1-c^2)$, when $K\to\infty$, since k=1,2,…,K, showing also that the variance in the random sequence $\{X_k\}$will be finite for this case, but the variance $(\sigma_K)^2$ will grow, if $|c|\geq 1$. By comparing to eq. (449), we see that $(\sigma_{innov})^2=(\sigma_r)^2(c_K)^2$, and it is reasonable to assume that $c_K=c_k=c$=constant, if the sample is uniform, although it is not necessary in general, otherwise.

So, since every electron in the sample accumulates variance $(\sigma_K)^2$ given by eq. (449), but the sample has a population of (n) carriers, which we sense simultaneously at the device terminals, then the observed instantaneous variance $(\sigma_{Kn})^2$ for average of (n) carriers is

$$\sigma_{Kn}^2=\frac{\sigma_K^2}{n}=\frac{\sigma_r^2}{n}\left(c_K^2+\sum_{k=1}^{K-1}c_k^2\right)=\frac{\sigma_r^2}{n}c^2K\text{, with } c_K=c_k=c=\text{constant for uniform sample.} \tag{453}$$

Then, the carriers traverse the sample for a transit time $\tau_t=K\times\Delta t$, where Δt is the time between the probing instances k, as defined above. Therefore, for an observation time (t), the variance $(\sigma_{Knt})^2$ seen at the device terminals can be rewritten as

$$\sigma_{Knt}^2=\frac{\sigma_r^2}{n}c^2K\frac{t}{\tau_t}=\frac{\sigma_r^2}{n}c^2\frac{\tau_t}{\Delta t}\frac{t}{\tau_t}=\left(\frac{\sigma_r^2}{n}\frac{c^2}{\Delta t}\right)t=At\text{,} \tag{454}$$

where all quantities in the brackets are device parameters, which may depend on material structure, fabrication, biasing, environment and other experimental conditions, but once the experimental conditions are fixed, then these device parameters are with constant values in time, and therefore, they can be combined in a single constant parameter A, as shown by the last right-hand term in eq. (454), and $\partial A/\partial t=0$. Rewriting eq. (454) in logarithmic form, and taking derivative, we get

$$\ln\left(\sigma_{Knt}^2\right)=\ln(A)+\ln(t)=\ln(A)-\ln(f)\text{, since } f_{min}=1/t_{max}\text{, and} \tag{455}$$

$$\frac{\partial\left(\sigma_{Knt}^2\right)}{\partial f}=-\frac{\sigma_{Knt}^2}{f}\text{, because } \quad \partial A/\partial t=0, \tag{456}$$

where we have also a substitution of the time (t) with the reciprocal variable f=1/t, with the obvious meaning that twice longer time of observation gathers ½ lower frequency in the spectrum of the variance $(\sigma_{Knt})^2$. Furthermore, the variance in a given unit bandwidth Δf around frequency (f) and power spectrum density S(f) at this frequency are related by $(\sigma_{Knt})^2=S(f)\Delta f$. When substituting this relation in eq. (456), and since the unit bandwidth Δf is a given constant, then Δf is cancelled, and we get

$$\frac{\partial S}{\partial f}=-\frac{S}{f}\text{.} \tag{457}$$

It is obvious to show that the solution of this differential equation is $\ln(S)=-\ln(f)$, from the finite difference of which we obtain the expression for the noise related to accumulation of variance in the sample as

$$\frac{S(f)}{S(f_o)}=\frac{f_o}{f}\Rightarrow S(f)=\frac{f_oS(f_o)}{f}\text{,} \tag{458}$$

where the reference frequency f_o can be chosen arbitrary, and $S(f_o)$ is the power spectrum density at f_o. Noticeably, the statistical accumulation of variance produces 1/f noise, owing to the ability of the matter to probe

the built in variance, and forward the variance to next step in the time. In our example, the spatially uncorrelated "Roughness", which will be equivalent to temporal white noise, if $c_k=0$ in eq. (449), results in 1/f noise, when $|c_k|>0$, without any need of special entities, such as traps, phonons, $1/\tau$ distributions of time constants, uniform energy or spatial distributions, etc., usually assumed as the origin of the 1/f noise in electronic devices. In fact, it seems that we can make a conclusion that the common mechanism, but not necessarily the only one, in the nature that causes fluctuations with 1/f power spectrum is the statistical accumulation of variance in random sequences, known also as statistics with "innovation variance". In relation to this, we make the following notes.

* The accumulation of variance requires only non-zero cause-consequence property in the object, which is natural for deterministic objects, thus, the 1/f noise should be very common physical phenomenon, as well established from the practice [160]. The 1/f noise from accumulation of variance is completely statistical in its origin, it is based on "innovation variance" statistics, and therefore is different from the commonly used approach for superposition of independent fluctuations with $1/\tau$ distribution and accompanied with instant probing of these superimposed fluctuations. Indeed, these two approaches are not contradicting each with other, since at the time, when the superposition integral was proposed for description of 1/f noise, Du Pre has clearly stated in 1950, in the title of [358], that the superposition is "A Suggestion Regarding the Spectral Density of Flicker Noise", and such clear statements should be not overlooked.

* The derivations presented above have assumed constant rates, that is, all parameters have constant non-zero values in eq. (454) at any time scale of observation, which results in a constant relative rate of variance accumulation, and thus, in noise with exactly 1/f spectrum. If a parameter in this equation has a time scale dependence, e. g., if $\partial\ln(A)/\partial\ln(t)=\partial\ln(\sigma_r^2)/\partial\ln(t)=\beta$ in a given time scale interval, which can be caused by autocorrelation, for example, then the slope of the spectrum deviates from 1/f with the same factor β in the corresponding frequency range. This is because eqs. (455), (456) and (457) will be modified to

$$\ln\left(\sigma_{Knt}^2\right)=\ln\left(A_o t^{\beta}\right)+\ln(t)=\ln\left(A_o\right)-\ln(f^{1+\beta}), \tag{455a}$$

$$\frac{\partial\left(\sigma_{Knt}^2\right)}{\partial\left(f^{1+\beta}\right)}=-\frac{\sigma_{Knt}^2}{f^{1+\beta}}, \tag{456a}$$

$$\frac{\partial S}{\partial\left(f^{1+\beta}\right)}=-\frac{S}{f^{1+\beta}}, \tag{457a}$$

and, consequently, the solution of the last differential equation is $\ln(S)=-\ln(f^{1+\beta})$, resulting in power density spectrum

$$\frac{S(f)}{S(f_o)}=\left(\frac{f_o}{f}\right)^{1+\beta}\Rightarrow S(f)\propto\frac{1}{f^{1+\beta}} \tag{458a}$$

in the particular frequency range.

* The derivations presented above have assumed linear system of first order. Non-linear and high order effects are not considered, which however, may lead to stochastic resonance, bistability (RTS), bifurcation, chaos – all of great interest [273, 398, 399, 400, 401, 402]. The derivation provides qualitative result for 1/f noise behavior, rather than a complete quantitative description of the 1/f noise, since the value of $S(f_o)$ in eq. (458) has to be obtained by an alternative method, not cancelling the quantities in the brackets of eq. (454), as it happened in the

step from eq. (455) to eq. (456). Thus, in principle, the derivation shows that the 1/f noise could have statistical origin, but the derivation could be also qualified as heuristic to some extent, and more work is needed to obtain mature treatment of 1/f noise in terms of "innovation variance" statistics. For this point, we consider and cite the comment in [356] that Schottky made in 1926 regarding mathematical treatments of fluctuations, including his original derivation of the shot noise [355] in 1918. The citation is: "This (mathematical) procedure, though it is perhaps not mathematically the simplest and perhaps will later be replaced by a more elegant one (as in the case of small-shot effect), has the advantage of greater generality and in addition makes it possible to carry out an entirely separate investigation of the fluctuation process itself and of its action upon the circuits." Interestingly, more than a century later, the comment still applies.

* The derivation, and in particular eqs. (449), (452) and (453), can be interpreted, and by using many equivalent terms. One interpretation is "backward" auto-correlation with the history of the process, but correlation with the future is physically incorrect, although, mathematically and in terms of signal processing with FIR filters is possible. Second interpretation is in a form of convolution, e.g. $\sigma_K^2(t)=\int_0^t \sigma_r^2(\tau)c^2(t-\tau)d\tau$, but convolution of quantities in square does not have clear meaning in terms of linear transformations, such as Fourier transformation, and justification of the physical significance is unclear, although there is no problem from formal mathematical point of view, statistics and signal processing. It is highly desired to unify the terms used in different sciences, in order to achieve coherent communication and avoid misinterpretation between researchers, and to make the delivered results convertible for the practice of noise characterization, modeling and in the real world of implementation by engineers. As one can see, many terms we needed to put in quotations, since these terms are denoted differently by different authors from different fields.

Nevertheless, by choosing f_o=1Hz in eq.(458), and dividing both sides of the equation on the square of the average value, (DC^2), we obtain the expression for normalized 1/f noise

$$S_{norm}(f)=\frac{S(f)}{DC^2}=\frac{1Hz\times S(1Hz)}{f\ DC^2}=\frac{K_F}{f}, \tag{459}$$

in the generic form, as introduced earlier by eq. (15).

VIII.6. Consequences from statistical nature of LFN – distributions in spectra, techniques of averaging, data volume and coordinates, instrumentation

An important relation found in [396] for the statistical accumulation of variance in data series is that after the Fourier transformation of a random sequence $\{X_1, X_2, \ldots, X_k,\ldots,X_K\}$, representing sampling of continuous random signal at $K\rightarrow\infty$ with normalized "duration" T_n=K, the "innovation variance", (which is added at each step k=1,2,…,K of the normalized sampling "period" Δt_n=1, denoting the normalized "time" as $t_n=k\Delta t_n=k$), is given by the exponent of the integral with limits ±1/2 of natural logarithm of the spectral density $S_x(f_n)$ of {X}, where the normalized frequency is $f_n=\pm 0.5k/K\leq\pm 0.5/\Delta t_n=\pm 0.5$. Rewritten in equations, that is [393, 395]

$$\sigma_{innov}^2\left(=\sigma_r^2c^2\right)=\exp\left\{\int_{-1/2}^{1/2}\ln[S_X(f_n)]df_n\right\}, \text{ using normalized frequency is } f_n, \tag{460}$$

or, as equivalently written in [394]

$$\sigma_{innov}^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln[S_{x\omega}(\omega_n)]d\omega_n\right\}, \text{ using normalized angular frequency } \omega_n=2\pi f_n. \quad (461)$$

To the best of our knowledge, the above relations are used in statistics [403] and particularly, in economics [397], signal processing [396], and control systems [404], but newer for low-frequency noise in electronic devices and other physical systems, perhaps, because the time and the frequency are normalized, which is a departure from physical coordinate systems that makes the physical interpretations difficult, and second, because the accumulation of variance in time is somewhat overlooked in electronic circuits, by postulating finite variance in assumption A2 (see again the end of section VIII.1. "**Trend for the LFN level and variations**") in order to have invariant system, although the Allan variance for 1/f noise, jitter and phase noise (see the discussion on eq. (372) for details) and the divergence of the power of 1/f noise at zero frequency clearly indicate that the assumption for invariant system might be false. So, we put more weight on the most general assumption A1, that the statistics is what describes the noise, although the statistics may not explain all details physically, and we carry out the noise analysis further by the help of eq. (460).

By the symmetry of spectral density, $S_x(f_n)=S_x(-f_n)$, for real signals or series {X} acquired from real physical processes [393], eq. (460) can be rewritten as

$$\ln\left(\sigma_{innov}^2\right) = 2\int_0^{1/2}\ln[S_x(f_n)]df_n = 2\int_0^{1/2}\ln\left\{\frac{d\left[\sigma_x^2(f_n)\right]}{df_n}\right\}df_n, \quad (462)$$

by also taking logarithm of eq. (460) and also showing explicitly that $S_x=d(\sigma_x)^2/df_n$ is the spectral density of variance in {X}. Next, we scale the normalized frequency (f_n) to physical frequency (f), by scaling the normalized variable f_n to $f=f_{max}f_n$, where f_{max} corresponds to the sampling frequency (respectively, sampling period $t_{min}=1/f_{max}$) used in the acquisition of the random sequence {X} from the real and continuous physical noise signal X(t). Using the rules for change of integration variables, $df_n=df/f_{max}$ and $f=f_{max}/2$ when $f_n=1/2$, we get

$$\ln\left(\sigma_{innov}^2\right) = 2\int_0^{1/2}\ln\left\{\frac{d\left[\sigma_x^2(f_n)\right]}{df_n}\right\}df_n = 2\int_0^{f_{max}/2}\ln\left\{f_{max}\frac{d\left[\sigma_x^2(f)\right]}{df}\right\}\frac{df}{f_{max}}, \quad (463)$$

$$\ln\left(\sigma_{innov}^2\right) = \frac{2}{f_{max}}\int_0^{f_{max}/2}\ln[f_{max}S_x(f)]df = \ln(f_{max})\frac{2}{f_{max}}\int_0^{f_{max}/2}df + \frac{2}{f_{max}}\int_0^{f_{max}/2}\ln[S_x(f)]df, \quad (464)$$

$$\ln\left(\sigma_{innov}^2\right) = \ln(f_{max}) + \frac{2}{f_{max}}\int_0^{f_{max}/2}\ln[S_x(f)]df \Rightarrow \ln\left(\sigma_{innov}^2\right) - \ln(f_{max}) = \frac{2}{f_{max}}\int_0^{f_{max}/2}\ln[S_x(f)]df, \quad (465)$$

$$\ln\left(\frac{\sigma_{innov}^2}{f_{max}}\right) = \ln\left(S_{innov}^{average}\right) = \frac{2}{f_{max}}\int_0^{f_{max}/2}\ln[S_x(f)]df, \quad (466)$$

where the $S_{innov}=(\sigma_{innov})^2/f_{max}$ is an average measure for the spectral density of the process, which variance is accumulated with the time, so that it causes the spectral density S_x in the random time sequences {X(t)}, the

latter observed in the experiment (that is, in the measurement of X(t), for example noise current measurement, by sampling with period $t_{min}=1/f_{max}$). It can be shown that the result also corresponds to the rule for scaling between time and frequency in Fourier transformation, $Z(\alpha t)\leftrightarrow(1/\alpha)[Z(f/\alpha)]$, from which it follows that $\sigma_x(t_n=t/t_{min}=tf_{max})\leftrightarrow[\sigma_x(f_n=f/f_{max})]/f_{max}$; scaling the normalized "time" t_n with physical sampling period $\alpha=1/f_{max}$, in order to obtain physical time (t), then $\sigma_x(t)=\sigma_x(t=\alpha t_n)\leftrightarrow(1/\alpha)[\sigma_x(f_n/\alpha=f)]/f_{max}=\sigma_x(f)$. Consequently, we write

$$\text{variance increment from noise source} = \int_{f_{min}}^{f_{max}/2} S_{innov}(f)df =$$

$$= \exp\left\{\frac{2}{f_{min}}\int_{f_{min}}^{f_{max}/2}\ln\left([S_x(f)]^{[f_{min}/f_{max}]}\right)df\right\} = \tag{467}$$

$$= \exp\left\{2\,(\text{duration})\int_{(\text{frequency resolution})}^{(\text{sampling rate})/2}\ln\left(\sqrt[\text{num.points}]{\text{measured noise density}}\right)df\right\}$$

Evidently, the heuristic interpretation of the last equation is difficult, even wrong, if one tries to match physical units, missing the detail that the equation is for noise variance increment at given rate in a variance accumulation process, that is $(\sigma_{innov})^2=\sigma^2/\text{rate}$, and S_{innov} is spectral density of power rate. This is confusing, therefore, we read carefully reference [393], considering that we always have finite number of points, when measuring the noise.

The finite number of points K suggests to convert the integrals in Riemann sums. It is shown in [393] that the logarithmic function causes bias in the finite estimators, and the correct conversion of eq. (466) with the bias compensated is

$$\ln\left(S_{innov}^{average}\right)\approx 0.57721...+\frac{2f_{min}}{f_{max}}\sum_{k=1}^{f_{max}/2f_{min}}\ln[S_x(f_{min}k)],\ \text{with}\ \sigma_{ln}=\frac{\pi^2}{6}\frac{2f_{min}}{f_{max}} \tag{468}$$

where 0.57721… is the Euler constant, k=0 is excluded from the sum, since it corresponds to DC, and σ_{ln} is the standard deviation for this estimator. Then, the upper boundary of the sum $f_{max}/(2f_{min})=\text{int}(K/2)$ is the maximum integer number less than or equal to half of number of sampled points (K/2), which is the number of the Nyquist frequency, $f_{min}\text{int}(K/2)$, which corresponds to the maximum frequency discriminated in the noise spectrum by sampling rate f_{max}. We will skip many details related to the last equation, such as requirements for random variables, tapering (windowing), discussed in length for random sequences in [393, 394, 395], in order to emphasize the following important for the noise relations.

If one has i_{max} data points for noise, by setting a correspondence to above as $i_{max}=\text{int}(K/2)$, i=k and $S_x(k)=S_i$, then

$$10\text{dB}\times\log_{10}\left(S_{average}\right)\approx 2.5\text{dB}+\frac{10\text{dB}}{i_{max}}\sum_{i=1}^{i_{max}}\log_{10}[S_i],\ \text{with}\ \sigma_{ln}\approx 5.2\text{dB}-10\text{dB}\times\log_{10}(i_{max}), \tag{469}$$

$$\text{or, in absolute magnitudes,}\ S_{average}\approx 1.78\left(\sqrt[i_{max}]{\prod_{i=1}^{i_{max}}S_i}\right),$$

which imply that

* The average noise should be obtained by geometric averaging, which is arithmetic, when the noise is expressed in dB, see $S_{dB,avg}$ in eq.(83),

* The geometric average underestimates the average noise S_{avg} with 2.5dB, that is $S_{avg}=S_{dB,avg}+2.5dB$, and

* The geometric average has a standard deviation $\sigma_{geo.avg}\sim\sigma_{ln}\approx 10dB\times[0.5-\log_{10}(num.points)]$. Thus, the geometric average of 10 data points has an uncertainty of $\pm\sigma_{geo.avg}(i_{max}=10)\approx 10dB\times[0.5-1]=-5dB\approx 30\%$, whereas $\sigma_{geo.avg}$ will decrease to 10% and 3%, when increasing the number of points i_{max} to 30 and 100, respectively.

Three more findings in [393], which are relevant to physical signals, are important to mention. These are

* The magnitude distribution of $S_x(k)=S_i$ is exponential,

* The distribution of the logarithm of noise spectrum, that is $\log(S_x(k))=\log(S_i)$, is normal, which means that the distribution of the noise spectrum $S_x(k)=S_i$ is log-normal (but not normal).

* When converting to Riemann sum, a "periodogram" is used for $S_x(k)=S_i$, which is

$$S_i = S_x(f_{min}k) = \frac{1}{K}\left|\sum_{\xi=1}^{K} X_\xi \exp\left(j2\pi\xi\frac{k}{K}\right)\right|^2 \text{, "periodogram",} \tag{470}$$

which is not exactly power spectrum, the later for discrete Fourier transformation given as

$$\text{power spectrum} = \frac{1}{K}\left|\sum_{\xi=0}^{K-1} X_{(\xi+1)} \exp\left(-j2\pi\xi\frac{k}{K}\right)\right|^2, \tag{471}$$

and the small difference should be checked in future as to whether it causes unforeseen issue.

RTS noise. Averaging techniques.

Now, we inspect the above findings. We chose RTS noise, which is well known to originate to charge traps in electronic devices, and therefore, one can assume the accumulation of variance is negligible, that is, $\sigma_x=\sigma_{innov}$, since $c\approx 0$ in eq.(452). As follows from Langevin equation, see before eq. (426), one expects exponential distribution for the RTS time constants, which is also experimentally confirmed in [68], as illustrated in Figure 74. Accordingly, we generate many RTS records, e.g. as shown in Figure 75, with time constants exponentially distributed. Next, we obtain the spectra of each record, as shown in Figure 76 with light-gray line, where also the root-mean-square (RMS) average and geometric mean (Geomean) are shown with thick lines, illustrating that Geomean underestimates RMS average with 2.5dB, as stated above. Interestingly, the upper limits Geomean+σ_{dB} and RMS(average+σ) for the most probable deviations from average have virtually the same values, whereas the lower limits are different, and the RMS average is not well centered between the limits when plotting in logarithmic scale. When calculating the histograms of spectral line magnitudes around their averages, we obtain the distributions as shown in Figure 77. The histogram for RMS averaging, Figure 77a, indicates clearly the distribution of spectral line amplitudes is exponential, which means that limit lower than ($RMS_{AVG}-\sigma$) does not have physical meaning, since it is equivalent to "negative" noise. The histogram for geometric averaging, Figure 77b, indicates that amplitudes can be also well described by log-normal distribution, slightly skewed on right with 2.5dB, so that the mode of the distribution matches with RMS_{AVG}.

The two distributions in Figure 77 imply that the RMS averaging is more suitable for measurements of a device at given DC operating point, because it conserves the variance, whereas the geometric average underestimates

the variance with 2.5dB. However, the RMS average will be not very suitable when describing variations, since the exponential distribution is asymmetric, which is a problem at larger scattering of data between different operating points, owing to occurrence of bias-dependent Lorentzian components in noise spectra, or between different devices, e.g. for small area devices, as it was shown earlier in Figure 9 on left. At large scattering in the data, despite the inherent underestimation of 2.5dB, the geometric average is preferable, as illustrated further in Figure 78 [28], because the geometric averaging finds the values, where the density of data points is higher, whereas the RMS averaging suggests values above that values.

The reason for this difference between geometric and RMS averaging is in the statistical distributions in the data. This is shown in Figure 79 from [79]. The RMS averaging implies a hypothesis that the variance is finite, an assumption that is not necessary between samples, and in addition, the distribution of noise power is exponential, as discussed above, and also seen in Figure 79a. In contrary, in Figure 79c shows that large number of data points for noise tends to log-normal distribution, which was stated just before eq. (470). The distribution of noise amplitude, Figure 79b, is "between" the cases for noise power and logarithmic magnitude. The histograms, basically, confirm the statements based from the statistics of "innovation variance", given after eq. (469), which imply that the statistics of the noise is "multiplicative", according to this equation. Unfortunately, this statistics is not elaborated at present for the low-frequency noise, while the "additive" statistics is attempted to be used, as one can see [72], and these analyses deduced that the variation in the noise level is larger than the average noise, which is somewhat not acceptable, because it sounds as that the noise in the noise is larger than the noise itself. Therefore, more works should be devoted to investigate the noise variation from statistical point of view, after examination of the noise distributions, rather than to assume unproven hypotheses. The issue is that the statistics for high frequency thermal and shot noise should not be transferred at hoc to low-frequency noise, and the noise description begins and remains to be first statistics of random numbers, and then one may search for physical phenomena that are behind the noise, explaining the statistical noise behaviour as function of device size, bias, temperature, etc. The opposite approach will always fail at some point.

Instrumentation

The noise measurements for statistical evaluations, however, require large number of samples and measurement conditions, normally in the range of thousands, which will affect also the instrumentation, since the access to nanodevices is difficult, and also, the methods for noise characterization will change, in order to maintain and process the large volume of data, especially when it is expected that the nanodevices will show prominent stochastic behavior [3]. So, we return to the first sentence in this section, that the noise in electronic devices was always hand by hand with the development of these devices. The data for noise were few in the 1920's, and they were analyzed almost for one decade, while the equipment was bulky, comprising vacuum tube amplifiers, large and manually operated LC ladders. In the 1950's, the analog spectrum analyzers were the tool to measure the noise and compare to the thermal noise in terms of noise figures. With the development of microprocessors, and since the fast Fourier transformation was discovered in the mid 1960's [405], the boost in the instrumentation in 1980's allowed to obtain the data in digital format, advancing the investigations of 1/f noise and mostly RTS noise for large time intervals and to low frequencies. At present, the low-frequency noise measurements are not a problem to be performed on wafer, having low noise amplifiers that are arranged near the probes, and to acquire and store as much data, as the time budget allows. The next step will be to mount the amplifiers in the probes itself, and the problem will be to process the large volume of measured data, rather to acquire them. In not very

far future, however, one will need to change the approach to the noise, since the devices below 22nm node are expected to have distinct stochastic behavior. Perhaps, the time-spectrum measurement and analysis of low-frequency noise will be insufficient; new “coordinate” for the noise will be needed, and electrical measurement might need to be substituted with other, since the electrical contacting to the device will be impossible, or with too much overhead of parasitic noise sources, in order to be reliable. Certainly, arrangements of few atoms, as that shown in the bottom-right corner of Figure 67b must be considered as devices from now on. Such devices might not be possible to probe directly electrically, but the noise in them will be almost everything that the devices do. So, major changes in the instrumentation and approaches for low-frequency noise is expected very soon.

IX. Conclusion

Here, the themes on low-frequency noise through this work are summarized together. It was shown in section II that the main equations used for description of low-frequency 1/f noise can be easily deduced by approach of separation of the noise sources as intrinsic (Hooge noise), or these sources couple their fluctuation in the current transport of the transistors. For both noise mechanisms, the normalized noise magnitude, in ratio to DC in square, for example DC current, is inversely proportional to the device area, which implies that the noise in sub-micron area devices relatively increases with every next generation of device down-scaling. By analyzing large number of published results for bipolar junction transistors (BJTs) in section III, several issues are identified for the low-frequency noise, as follows. The scattering of the data is large even after compensating for 1/area dependence, and the variation in 1/f noise levels in the publications originate from differences in measurement setups, mostly due to impedance of biasing circuit (section III.1), differences of fabrication approaches, mostly due to interfacial oxide in the emitter at the emitter-base junction (section III.2), crossover between noise sources by changing of biasing, mostly with surface and bulk origins (section III.3) by occurrence of large Lorentzian components due to charge trapping, and measurement and characterization uncertainty (section III.4), owing to scattering in noise spectra, fitting and averaging procedures. Overall, the 1/(emitter area) dependence of the 1/f noise in BJTs is confirmed, but the scattering of the data, especially for sub-micrometer emitter area BJTs, indicates an increased variability of the noise levels as $(\text{area})^{-3/2}$, causing standard deviations for noise levels larger than the average, which puts the question whether root-mean-square (RMS) or logarithmic averaging should be preferred at such large data scattering.

Since the MOS technology has the major advances in the last decades, the noise in MOS transistors was addressed in details in section IV. Obviously, one observes input referred noise voltage power spectrum densities (PSDs) in MOS transistors about two orders of magnitude larger than in BJTs, when the emitter and gate areas are similar, although the difference is about a decade or less for the output referred normalized noise current PSD, the latter in terms of SPICE parameter K_F. Looking closer at the models and predictions for 1/f noise in MOS transistors (section IV.1), one observes that the noise in MOS transistors is a complicated mixture of interface and oxide trapping, and mobility fluctuation, both correlated and uncorrelated to the trapping. Interesting observations are that there is no single model that can describe uniquely the noise in MOS transistors, all models have advantages and issues with nano-scaled MOS transistors. Nevertheless, the trapping model provides an approach to obtain trap profiles in oxide depth, but equally, distributions in trap energy, and these cannot be distinguished each from other, as shown in section IV.2. Even the amplitude of RTS noise in identical MOS transistors can vary with the position of the trap along the channel, as follows from the analyses in section

IV.3, and has explained data published in the past. In section IV.4, we have discussed a variety of figures-of-merits (FoMs) for noise, showing the relations between low-frequency and high-frequency noise and performance of MOS transistors, also in comparison with the corresponding for bipolar transistors. From this discussion, it becomes clear that the “low-frequency” noise in modern transistors is dominating even in the range of GHz, especially for MOS transistors, which puts several questions for the scalability of low-frequency noise models, as well as, issues for the applications of these devices.

In the never pace for faster and smaller devices, many modifications of the generic BJT and MOS structures are made, and advanced transistor structures are developed, the noise in which was reviewed in section V. Germanium has returned back in SiGe transistors to improve the hole mobility, or to add strain in silicon layers (along with other techniques), again with the purpose of boosting the mobility. Many approaches in material engineering are undertaken to decrease the effective electrical oxide thickness in MOS transistors, but the data for low-frequency noise imply that the noise also increases when mixing materials and deviating from homogeneous material layers, which is somewhat in contrast to several reports made in the past that SiGe transistors and MOS transistors with higher oxide capacitance should have less noise. Among the many techniques for boosting the MOS transistor performance, only the forward body biasing consistently reduces the low-frequency noise, although this technique is limited by the maximum allowable magnitude of the forward biasing and some complications, e.g. how to provide noise-free bias, along with the layout overhead for separate wells for the transistors, extra biasing lines in already crowded with wires chips, or the extra input capacitance in gate-body tied MOS transistors.

Since it is found to be unrealistic the bulk MOS transistor to be scaled down below 32nm node, many other advanced structures have been attempted, and results from low-frequency noise characterizations of these structure are discussed in section VI. SOI MOS has been developed to thin and ultra-thin bodies, which have been surrounded with two–three gates, and even with gate all around, e.g. as a cylindrical vertical MOS transistor. The calculations for multiple gate transistors indicate that the noise should decrease as compared to the one-gate MOS transistor, which is explained by capacitive coupling between the gates, or by moving the current flow from the surface toward the “bulk”. Unfortunately, the currents flowing in the body of SOI create undesirable noise component by filtered shot noise when the currents pass through body-source and body-drain junctions, an effect very prominent in partially depleted SOI. Also, the review of ultimately down-scaled devices, such as carbon nanotube transistors, shows that the low-frequency noise increases in relative units to DC, approximately following the same 1/area dependence as deduced for much bigger bulk and SOI transistors. While this observation is useful, the result is actually not acceptable for the practical applications of CNT, since the noise becomes larger than DC even for narrow frequency bands of 1-3 decades. In other words, the single nanotube devices, seems, are not anymore deterministic, that is, they are behind the down-scaling barrier set by the 1/f noise.

The impact of low-frequency noise in circuit designs was illustrated in section VII. In the first place, section VII analyzes a typical circuit topology in the input stage of low-frequency amplifiers, for which it is shown the significance of the product (quiescent supply current)×(input referred noise voltage PSD), which suggests a general tradeoff between consumption and noise in micropower amplifiers. The contribution of low-frequency noise in up-conversion to phase noise was also discussed in this section, by reviewing the main approaches used at present for modeling of phase noise. With some simplifications, it is shown also that the phase noise can be

deduced from harmonic content in oscillator output, and the prediction of this model for minimum phase noise coincides with the practical rule to get the maximum possible amplitude from oscillator, but at not very high distortion in sinusoidal signal. Several practical considerations for phase noise are also discussed, among which are that the circuit asymmetry is responsible for up-conversion of low-frequency noise, while the white noise at every harmonic contributes to the phase noise, and also that the 1/f noise causes phase variance growing with time, so the increment of the variance (known as Allan variance) can be measured, but never the whole variance. This fact is somewhat overlooked in the treatment of low-frequency noise, and it is discussed further in the last section VIII. Section VII ends with the analysis of the impact of the LFN in sensors, in particular, electrochemical and photo sensors.

The discussion in section VIII is an attempt to summarize the achievements from investigations on low-frequency noise made for about one century, in order to outline the issues that are related with low-frequency noise in the near future. Interesting observation is that the figure of merit ($Area \times K_F$) stays in the range 10^{-8} μm^2 irrespectively of device types, e.g. vacuum tubes, BJTs, MOS transistors, carbon nanotubes, etc., and if one needs to say the possible minimum level of 1/f noise, the number is $(Area \times K_F) \geq 10^{-9}$ μm^2. Looking back in previous sections, one will also observe that the many models for 1/f noise in different types of electronic devices converge closely when the word is for numerical values, irrespectively of physical background of the models and characterization techniques used. Other observation in section VIII on the trends implies that the 1/f noise will impact soon the reliability of the operation of digital circuits, that is, the circuits with minimum sized devices might be not anymore deterministic, once the device downscaling crosses the 1/f noise barrier, which is most probably between nodes 32nm and 22nm. In other words, owing to the relatively larger low-frequency noise in nanodevices [406], the “room at the bottom” [407] for deterministic electronic devices is not anymore “plenty”. Therefore, we have inspected the assumptions for low-frequency noise, and identified that the statistical variations, both in manufacturing and due to accumulation of variance during operation of devices, cause generically 1/f noise. Again, the answers for the low-frequency noise, seems, originate from the statistics of numbers, fairly overlooked point of view at present searching for deep physics in the devices by extrapolating laws for average quantities in semiconductors toward few nanometer structures. We have shown that the statistics in the fabrication, e.g. ion implantation, results in rapid increase of uncertainty margin for sub-20nm device parameters that cause unacceptable limits for yield, by means of fabrication “frozen noise”. Consequently, the variance of the “frozen noise” is statistically accumulated by operation of the devices, by a mechanism known as “innovation variance” in the statistics, resulting in 1/f temporal noise and log-normal distributions in the noise spectra and variation in these spectra between nominally identical devices. Consequently, the geometric averaging should be preferred by characterization of noise, as suggested by “innovation variance” statistics.

The overall conclusion of this work can be given by the following citation from [2] made almost four decades ago that, “while we can describe the physical consequences of parameter fluctuations in intricate detail, we have comparatively little knowledge about the microscopic origins of voltage fluctuations in a simple resistor”. Since the low-frequency noise and variations are becoming limiting factors in device downscaling, it might be necessary to change the obvious coordinates time-amplitude and frequency-spectrum used in noise investigations, or to allow the 10nm devices to be not fully deterministic in practical applications. Many times it was needed to change the physical coordinate systems, e.g. from voltages and currents into scattering parameters (S-parameters) [196] for RF applications. At present, the many models generally coincide each with other, and

what makes the difference, are the values, which, however, scatter prominently in nanodevices.

References

[1]. G. Kirk, "Natural Change in Heraclitus", Mind (New Series), **60**(237), 35-42, 1951

[2] . P. Dutta, P. Horn, “Low-Frequency Fluctuations in Solids: 1/f Noise”, Rev. Modern Phys., **53**(3), 497-516, 1981

[3] . SEMATECH, “International Technology Roadmap for Semiconductors”, 2005 Edition, 2006 Update, http://public.itrs.net.

[4] . L. Vandamme, F. Hooge, “On the Additivity of Generation-Recombination Spectra Part 3:The McWhorter Model for 1/f noise in MOSFETs”, Physica B, **357**(3-4), 507-524, 2005

[5] . M. J. Deen, S. Rumyantsev, J. Orchard-Webb, "Low Frequency Noise in Heavily Doped Polysilicon Thin Film Resistors", J. Vac. Sci. Technol. B, **16**(4), 1881-1884, 1998.

[6] . M. J. Deen, O. Marinov, J. Yu, S. Holdcroft, W. Woods, “Low-Frequency Noise in Polymer Transistors”, IEEE Trans. Electron Devices, **48**(8), 1688-1695, 2001

[7] . O. Marinov, M. J. Deen, J. Yu, G. Vamvounis, S. Holdcroft, W. Woods, “Low-Frequency Noise in Polymer Thin-Film Transistors”,IEE Proc. - Circuits, Devices and Systems, **151**(5), 466-472, 2004

[8] . M. J. Deen, S. Rumyantsev, D. Landheer, D.-X. Xu, "Low-Frequency Noise in Cadmium-Selenide Thin-Film Transistors", Appl. Phys. Lett., **77**(14), 2234-2236, 2000.

[9] . L. Vandamme, “Noise as a Diagnostic Tool for Quality and Reliability of Electronic Devices”, IEEE Trans. Electron Devices, **41**(11), 2176- 2187, 1994

[10] . H. Markus, T. Kleinpenning, "Low-Frequency Noise in Polysilicon Emitter Bipolar Transistors", IEEE Trans. Electron Devices, **42**(4), 720-727, 1995

[11] . M. J. Deen, J. Ilowski, P. Yang, "Low Frequency Noise in Polysilicon-Emitter Bipolar Junction Transistors", J. Appl. Phys., **77**(12), 6278-6288, 1995

[12] . M. J. Deen, S. Rumyantsev, R. Bashir, R. Taylor, "Measurements and Comparison of Low Frequency Noise in npn and pnp Polysilicon Emitter Bipolar Junction Transistors", J. Appl. Phys., **84**(1), 625-633, 1998

[13] . X. Chen, M. J. Deen, Z. Yan, M. Schroter, "Effects of Emitter Dimensions on Low-Frequency Noise in Double-Polysilicon BJTs", Electronics Letters, **34**(2), 219-220, 1998

[14] . E. Simoen, S. Decoutere, C. Claeys, L. Deferm, "A Global Description of the Base Current 1/f Noise of Polysilicon Emitter Bipolar Transistors Before and After Hot-Carrier Stress", Solid-State Electronics, **42**(9), 1679-1687, 1998

[15] . C. Delseny, A. Penarier, F. Pascal, S. Jarrix, P. Llinares, "Comparison of Low Frequency Noise and High Frequency Performances of Double and Simple Polysilicon Bi-CMOS BJT", Microelectron. Reliab., **40**(11), 1869-1874, 2000

[16] . M. Sanden, M. J. Deen, O. Marinov, private communication, unpublished, 2001

[17] . M. Sanden, O. Marinov, M. J. Deen, M. Ostling, "A New Model for the Low-Frequency Noise and the Noise Level Variation in Polysilicon Emitter BJTs", IEEE Trans. Electron Devices, **49**(3), 514-520, 2002

[18] . M. J. Deen, E. Simoen, "Low-Frequency Noise in Polysilicon-Emitter Bipolar Transistors", IEE Proc.-Circuits Devices Syst., **149**(1), 40-50, 2002

[19] . M. Hoque, Z. Celik-Butler, D. Lan, D. Weiser, J. Trogolo, and K. Green, "Effect of Interfacial Oxide Thickness on 1/f Noise in Polysilicon Emitter BJTs", IEEE Trans. Electron Devices, **51**(9), 1504-1513, 2004

[20] . N. Valdaperez, J.-M. Routoure, D. Bloyet, R. Carin, S. Bardy, "Size Effects on the DC Characteristics and Low Frequency Noise of Double Polysilicon NPN Bipolar Transistors ", Microelectron. Reliab., **45**(7-8), 1167-1173, 2005

[21] . P. Benoit, J. Raoult, C. Delseny, F. Pascal, L. Snadny, J.-C. Vildeuil, M. Marin, B. Martinet, D. Cottin and O. Noblanc, "DC and Low Frequency Noise Analysis of Hot-Carrier Induced Degradation of Low Complexity 0.13 µm CMOS Bipolar Transistors ", Microelectron. Reliab., **45**(9-11), 1800-1806, 2005

[22] . I. Nam, K. Lee, "High-Performance RF Mixer and Operational Amplifier BiCMOS Circuits Using Parasitic Vertical Bipolar Transistor in CMOS Technology", IEEE J. Solid-State Circuits, **40**(2), 392-402, 2005

[23] . J. Kim, H. Oh, C. Chung, J.-H. Jeong, H. Lee, S.-H. Hwang, I.-C. Hwang, Y.-J. Kim, K. Hong, E. Jung, K.-P. Suh, "High Performance NPN BJTs in Standard CMOS Process for GSM Transceiver and DVB-H Tuner", IEEE RFIC, Symp., 4 pp., 2006

[24] . Q. Hu, S. Chen, S.-L. Zhang, P. Solomon, Z. Zhang, “Effects of substrate bias on low-frequency noise in lateral bipolar transistors fabricated on silicon-on-insulator substrate”, IEEE Electron Device Letters **41** (1), 4-7, 2020

[25] . J.-S. Syu, C. Meng, C.-L. Wang, "2.4-GHz low-noise direct-conversion receiver with deep n-well vertical-NPN BJT operating near cutoff frequency", IEEE Trans. on Microwave Theory and Techniques, **59** (12), 3195-3205, 2011
[26] . M. Fjer, S. Persson, E. Escobedo-Cousin, A. G. O'Neill, "Low Frequency Noise in Strained Si Heterojunction Bipolar Transistors" IEEE Transactions on Electron Devices, **58**, 4196-4203, 2011
[27] . S.-M. Chen, Y.-K. Fang, F.-R. Juang, C.-C. Chen, S. Liu, C.-W. Kuo, C.-P. Chao, H.-C. Tseng, "A low-flicker noise gate-controlled lateral–vertical bipolar junction transistor array with 55-nm CMOS technology", IEEE Transactions on Electron Devices **58** (10), 3276-3282, 2011
[28] . M. Jamal Deen , B. Iniguez, O. Marinov, F. Lime, "Electrical Studies of Semiconductor-Dielectric Interfaces", Springer J. Mater. Sci.: Mater. Electron., **17**(9), 663-638, 2006
[29] . M. J. Deen, O. Marinov, "Noise in Advanced Electronic Devices and Circuits", AIP Conf. Proc. **780**(1), 3-12, 2005
[30] . M. J. Deen, S. Rumyantsev, M. Schroter, "On the Origin of 1/f Noise in Polysilicon Emitter Bipolar Transistors", J. Appl. Phys., **85**(2), 1192-1195, 1999.
[31] . M.J. Deen, F. Pascal, "Review of Low-Frequency Noise Behaviour of Polysilicon Emitter Bipolar Junction Transistors", IEE Proc.-Circuits Devices Syst., **151**(2), 125- 137, 2004
[32] . M. Hoque, Z. Celik-Butler, J. Trogolo, D. Weiser, K. Green, "1/ f Noise in Positive-Negative-Positive (PNP) Polycrystalline Silicon-Emitter Bipolar Transistors", J. Appl. Phys., **97**(8), 084501-1 to 10, 2005
[33] . E. Zhao, R. Krithivasan, A. Sutton, Z. Jin, J. Cressler, B. El-Kareh, S. Balster, H. Yasuda, "Investigating the Differences in Low-Frequency Noise Behavior of npn and pnp SiGe HBTs Fabricated in a Complementary SiGe HBT BiCMOS on SOI technology", FAN-2005 Noise in Devices and Circuits III, A. Balandin, F. Danneville, M. J. Deen, D. Fleetwood, Eds., Proc. SPIE, **5844**, 132-142, 2005
[34] . E. Zhao, R. Krithivasan, A. Sutton, Z. Jin, J. Cressler, B. El-Kareh, S. Balster, H. Yasuda, "An Investigation of Low-Frequency Noise in Complementary SiGe HBTs", IEEE Trans. Electron Devices, **53**(2), 329-338, 2006
[35] . N. Lukyanchikova, N. Garbar, A. Smolanka, M. Lokshin, S. Hall, O. Buiu, I. Mitrovic, H. Mubarek, P. Ashburn, "1/f Noise and Generation/Recombination Noise in SiGe HBTs on SOI", IEEE Trans. Electron Devices, **52**(7), 1468-1477, 2006
[36] . E. Zhao, J. Cressler, M. El-Diwany, T. Krakowski, A. Sadovnikov, D. Kocoski, "On the Geometrical Dependence of Low-Frequency Noise in SiGe HBTs", Solid-State Electronics, **50**(11-12), 1748-1755, 2006
[37] . M. Hoque, Z. Celik-Butler, S. Martin, C. Knorr, C Bulucea, "Dependence of Low Frequency Noise in SiGe Heterojunction Bipolar Transistors on the Dimensional and Structural Features of Extrinsic Regions", Solid-State Electronics, **50**(7-8), 1430-1439, 2006
[38] . F. M. Puglisi, L. Larcher, P. Pavan, "Mixed-Mode Stress in Silicon-Germanium Heterostructure Bipolar Transistors: Insights From Experiments and Simulations", IEEE Transactions on Device and Materials Reliability, **19**, 275-282, 2019
[39] . B. Sagnes, F. Pascal, M. Seif, A. Hofflann, S. Haendler, P. Chevalier, D. Gloria, "Low Frequency Noise in advanced 55nm BiCMOS SiGeC Heterojunction Bipolar Transistors : impact of collector doping", International Conference on Noise and Fluctuations, **2017**
[40] . M. Seif, F. Pascal, B. Sagnes, A. Hoffmann, S. Haendler, P. Chevalier, D. Gloria, "Study of low frequency noise in advanced SiGe:C heterojunction bipolar transistors", Proceedings of the European Solid-State Device Research Conference, 373-376, 2014
[41] . O. Marinov, M. J. Deen, "1/f Noise Coefficient Kf vs. Poly-Monosilicon Interfacial Oxide", Raw Data in Electronics (www.rdie.ca), **1**(3), 73, 2007
[42] . A. Ziel, X. Zhang, A. Pawlikiewicz, "Location of l/f Noise Sources in BJT's and HBJT's-I. Theory", IEEE Trans. Electron Devices, **33**(9), 1371-1376, 1986
[43] . V. Kumar, W. Dahlke, "Low frequency noise in Cr-SiO_2-n-Si tunnel diodes", IEEE Trans. Electron Devices, **24**(2), 146-153, 1977
[44] . J. Robertson, "Interfaces and Defects of High-K Oxides on Silicon", Solid-State Electronics, **49**(3), 283-293, 2005
[45] . C. Ng, T. Chen, Y. Liu, C. Sun, "Influence of Nitrogen on Tunneling Barrier Heights and Effective Masses of Electrons and Holes at Lightly-Nitrided SiO_2/Si Interface", J. Appl. Phys., **96**(10), 5912-5914, 2004
[46] . O. Marinov, M. J. Deen, J. Tejada, "Theory of Microplasma Fluctuations and Noise in Silicon Diode in Avalanche Breakdown", J. Appl. Phys., **101**(6), 064515-1 to 21, 2007
[47] . B. Min, S. Devireddy, Z. Celik-Butler, F. Wang, A. Zlotnicka, H.-H. Tseng, P. Tobin, "Low-Frequency Noise in Submicrometer MOSFETs With HfO_2, HfO_2/Al_2O_3 and $HfAlO_x$ Gate Stacks", IEEE Trans Electron Devices, **51**(10), 1679-1687, 2004

[48] . B. Min, S. Devireddy, Z. Celik-Butler, A. Shanware, K. Green, J. Chambers, M. Visokay, L. Colombo, "Low-Frequency Noise Characteristics of HfSiON Gate Dielectric Metal-Oxide-Semiconductor-Field-Effect Transistors", Appl. Phys. Lett., **86**(8), 082102-1 to 3, 2005

[49] . B. Min, S. Devireddy, Z. Celik-Butler, A. Shanware, L. Colombo, K. Green, J. Chambers, M. Visokay, A. Rotondaro, "Impact of Interfacial Layer on Low-Frequency Noise of HfSiON Dielectric MOSFETs", IEEE Trans. Electron Devices, **53**(6), 1459-1466, 2006

[50] . S. Devireddy, B. Min, Z. Celik-Butler, H.-H. Tseng, P. Tobin, F. Wang, A. Zlotnicka, "Low-Frequency Noise in TaSiN/HfO_2 nMOSFETs and the Effect of Stress-Relieved Preoxide Interfacial Layer", IEEE Trans. Electron Devices, **53**(3), 538-544, 2006

[51] . A. Ahsan, D. Schroder, "Impact of Channel Carrier Displacement and Barrier Height Lowering on the Low-Frequency Noise Characteristics of Surface-Channel n-MOSFETs ", Solid-State Electronics, **49**(4), 654-662 , 2005

[52] . M. Haartman, B. Malm, M. Ostling, "Comprehensive Study on Low-Frequency Noise and Mobility in Si and SiGe pMOSFETs with High-κ Gate Dielectrics and TiN Gate", IEEE Trans. Electron Devices, **53**(4), 836-843, 2006

[53] . E. Simoen, A. Mercha, C. Claeys, E. Young, "Correlation Between the 1/ f Noise Parameters and the Effective Low-Field Mobility in HfO_2 Gate Dielectric n-Channel Metal-Oxide-Semiconductor Field-Effect Transistors", Appl. Phys. Lett., **85**(6), 1057-1059, 2004

[54] . K. Hung, P. Ko, C. Hu, Y. Cheng, "A Physics-Based MOSFET Noise Model for Circuit Simulators", IEEE Trans. Electron Devices, **37**(5), 1323-1333, 1990

[55] . K. Hung, P. Ko, C. Hu, Y. Cheng, "A Unified Model for the Flicker Noise in Metal-Oxide-Semiconductor Field-Effect Transistors", IEEE Trans. Electron Devices, **37**(3), 654-665, 1990

[56] . R. Jayaraman, C. Sodini, "A 1/f Noise Technique to Extract the Oxide Trap Density Near the Conduction Band Edge of Silicon", IEEE Trans. Electron Devices, **36**(9), 1773-1782, 1989

[57] . M. von Haartman, J. Westlinder, D. Wu, B. Malm, P.-E. Hellstrom, J. Olsson, M. Ostling, "Low-Frequency Noise and Coulomb Scattering in Si0.8Ge0.2 Surface Channel pMOSFETs with ALD Al2O3 Gate Dielectrics", Solid-State Electronics, **49**(6), 907-914, 2005

[58] . S. Martin, G. Li, E. Worley, J. White, "The Gate Bias and Geometry Dependence of Random Telegraph Signal Amplitudes (MOSFET)", IEEE Electron Device Letters, **18**(9), 444-446, 1997

[59] . C. Surya, T. Hsiang, "Surface Mobility Fluctuation in Meal-Oxide-Semiconductor Field-Effect Transistors", Phys. Rev. B, **35**(12), 6343-6347, 1987

[60] . E. Zhao, A. Sutton, B. Haugerud, J. Cressler, P. Marshall, R. Reed, B. El-Kareh, S. Balster, H. Yasuda, "The Effects of Radiation on 1/f Noise in Complementary (npn + pnp) SiGe HBTs", IEEE Trans. Nuclear Science, **51**(6), 3243-3249, 2004

[61] . T. Kleinpenning, "Low-Frequency Noise in Modern Bipolar Transistors: Impact of Intrinsic Transistor and Parasitic Series Resistances", IEEE Trans. Electron Devices, **41**(11), 1981-1991, 1994

[62] . X. Zhu, A. Ziel, "The Hooge Parameters of n+-p-n and p+-n-p Silicon Bipolar 'Transistors", IEEE Trans. Electron Devices, **32**(3), 658-661, 1985

[63] . C. Mukherjee, T. Jacquet, T. Zimmer, C. Maneux, A. Chakravorty, J. Boeck, K. Aufinger, K. "Comprehensive Study of Random Telegraph Noise in Base and Collector of advanced SiGe HBT: Bias, Geometry and Trap Locations", Proceedings of the European Solid-State Device Research Conference*,* 2016

[64] . C. Mukherjee, T. Jacquet, A. Chakravorty, T. Zimmer, J. Bock, K. Aufinger, C. Maneux, "Low-Frequency Noise in Advanced SiGe:C HBTs-Part I: Analysis", IEEE Transactions on Electron Devices, (IEEE*),* **63**, 3649-3656, 2016

[65] . J. Raoult, F. Pascal, C. Delseny, M. Marin, M. J. Deen, "Impact of carbon concentration on 1⁄f noise and random telegraph signal noise in SiGe:C heterojunction bipolar transistors", Journal of Applied Physics, AIP Publishing*,* **103**, 114508, 2008

[66] . M. Sanden, O. Marinov, M. J. Deen, M. Ostling, "Modeling the Variation of the Low-Frequency Noise in Polysilicon Emitter Bipolar Junction Transistors", IEEE Electron Device Letters, **22**(5), 242-244, 2001

[67] . M. Sanden, M. Ostling, O. Marinov, M. J. Deen, "Statistical Simulations of the Low-Frequency Noise in Polysilicon Emitter Bipolar Transistors Using a Model Based on Generation-Recombination Centers", Fluctuation and Noise Letters, **1**(2), L51-L60, 2001

[68] . M. Kirton, M. Uren, "Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States and Low-Frequency (1/f) Noise", Adv. Phys., **38**(4), 367-468, 1989

[69] . C. Surya, S.-H. Ng, E. Brown, P. Maki, "Spectral and Random Telegraph Noise Characterization of Low-Frequency Fluctuations in GaAs/$Al_{0.4}Ga_{0.6}As$ Resonant Tunneling Diodes", IEEE Trans. Electron Devices, **41**(11), 2016-2022, 1994

[70] . F. Hooge, "1/f Noise Sources", IEEE Trans. Electron Devices, **41**(11), 1926-1935, 1994

[71] . B. Streetman, S. Banerjee, “Solid State Electronic Devices”, 5th edition, Prentice Hall, 2000
[72] . G. Wirth, J. Koh, R. Silva, R. Thewes, R. Brederlow, "Modeling of Statistical Low-Frequency Noise of Deep-Submicrometer MOSFETs", IEEE Trans. Electron Devices, **52**(7), 1576-1588, 2005
[73] . C. Surya, T. Hsiang, "Theory and Experiment on the $1/f^{\gamma}$ Noise in p-Channel Metal-Oxide-Semiconductor Field-Effect Transistors at Low Drain Bias", Phys. Rev. B, **33**(7), 4898-4905, 1986
[74] . T. Boutchacha, G. Ghibaudo, “Low Frequency Noise Characterization of 0.18 μm Si CMOS Transistors”, Phys. Stat. Sol. (a), **167**(1), 261-270, 1998
[75] . MIL-STD-202G, “Test Method Standard Electronic and Electrical Component Parts - Method 308: Current-Noise Test for Fixed Resistors”, Defense Supply Center Columbus, http://www.dscc.dla.mil/Downloads/MilSpec/Docs/MIL-STD-202/std202mthd308.pdf
[76] . Z. Stanimirovic, M. Jevtic, I. Stanimirovic, "Simultaneous Mechanical and Electrical Straining of Conventional Thick-Film Resistors", Microelectron. Reliab., **48**(1), 59-67, 2008
[77] . A. Dziedzic, A. Kolek, W. Ehrhardt, H. Thust, "Advanced Electrical and Stability Characterization of Untrimmed and Variously Trimmed Thick-Film and LTCC Resistors", Microelectron. Reliab., **46**(2-4), 352-359, 2006
[78] . K. Arshak, L. Cavanagh, C. Cunniffe, "Excess Noise in a Drop-Coated Poly(vinyl butyral)\Carbon Black Nanocomposite Gas Sensitive Resistor", Thin Solid Films, **495**(1-2), 97-103, 2006
[79] . M. J. Deen, O. Marinov, D. Onsongo, S. Dey, S. Banerjee, "Low-Frequency Noise in SiGeC-Based pMOSFETs", Proc. SPIE, **5470**, Noise in Devices and Circuits II; F. Danneville; Ed., 215-225, 2004
[80] . M. Erturk, T. Xia, R. Anna, K. Newton, E. Adler, "Statistical BSIM Model for MOSFET 1/f Noise", Electronics Letters, **41**(22), 1208-1210, 2005
[81] . Z. Jin, M. Erturk, J. Cressler, A. Joseph, "The impact of low-frequency noise variations on the modeling and operation of SiGe circuits", Proc. Bipolar/BiCMOS Circuits and Technology Meeting, BCTM, 192-195, 2005
[82] . W.-C. Hua, M. Lee, P. Chen, M.-J. Tsai, C. Liu, "Threading Dislocation Induced Low Frequency Noise in Strained-Si nMOSFETs", IEEE Electron Device Letters, **26**(9), 667-669, 2005
[83] . J. Babcock, B. Loftin, P. Madhani, C. Xinfen, A. Pinto, D. Schroder, "Comparative Low Frequency Noise Analysis of Bipolar and MOS Transistors Using an Advanced Complementary BiCMOS Technology", IEEE Conf. Custom Integrated Circuits, CIS'01, 385-388, 2001
[84] . O. Marinov, M. J. Deen, “1/f Noise Coefficient Kf vs. Emitter Area of p-n-p Bipolar Junction Transistors”, Raw Data in Electronics (www.rdie.ca), **1**(3), 72, 2007
[85] . A. McWhorter, “l/f Noise and Germanium Surface Properties”, in Semiconductor Surface Physics, R. Kingston, Ed., Univ. of Pennsylvania Press, Philadelphia, p. 207, 1957
[86] . D. Binkley, C. Hopper, J. Cressler, M. Mojarradi, and B. J. Blalock, "Noise Performance of 0.35-mm SOI CMOS Devices and Micropower Preamplifier Following 63-MeV, 1-Mrad (Si) Proton Irradiation", IEEE Trans Nuclear Science, **51**(6), 3788-3794, 2004
[87] . E. Simoen, A. Mercha, L. Pantisano, C. Claeys, and E. Young, "Low-Frequency Noise Behavior of SiO_2–HfO_2 Dual-Layer Gate Dielectric nMOSFETs With Different Interfacial Oxide Thickness", IEEE Trans. Electron Devices, **51**(5), 780-784, 2004
[88] . E. Simoen, A. Mercha, C. Claeys, N. Lukyanchikova, N. Garbar, "Critical Discussion of the Front–Back Gate Coupling Effect on the Low-Frequency Noise in Fully Depleted SOI MOSFETs", IEEE Trans. Electron Devices, **51**(6), 1008-1016, 2004
[89] . K. Chew, K. Yeo, S.-F. Chu, "Impact of Technology Scaling on the 1/f Noise of Thin and Thick Gate Oxide Deep Submicron NMOS Transistors", IEE Proc. Circuits Devices Systems, **151**(5), 415-421, 2004
[90] R. Wilcox, J. Chang, C. Viswanathan, "Low-Temperature Characterization of Buried-Channel NMOST", IEEE Trans Electron Devices, **36**(8), 1440-1447, 1989
[91] . J. Chang, A. Abidi, C. Viswanathan, "Flicker Noise in CMOS Transistors from Subthreshold to Strong Inversion at Various Temperatures", IEEE Trans. Electron Devices, **41**(11), 1965-1971, 1994
[92] . C. Claeys, E. Simoen, A. Mercha, L. Pantisano, E. Younga, "Low-Frequency Noise Performance of HfO_2-Based Gate Stacks", J. Electrochemical Society, **152**(9), F115-F123, 2005
[93] . T. Elewa, B. Boukriss, H. Haddara, A. Chovet, S. Cristoloveanu, "Low-Frequency Noise in Depletion-Mode SIMOX MOS Transistors", IEEE Trans. Electron Devices, **38**(2), 323-327, 1991
[94] . J.-S. Lee, Y.-K. Choi, D. Ha, T.-J. King, J. Bokor, "Low-Frequency Noise Characteristics in p-Channel FinFETs", IEEE Electron Device Letters, **23**(2), 722-724, 2002
[95] . K. O, N. Park, D.-J. Yang, "1/f Noise on NMOS and PMOS Transistors and their Implications to Design of Voltage Controlled Oscillators", IEEE RFIC Symp., 59-62, 2002

[96] . R. Yang, W. Loh, M. Yu, Y.-Z. Xiong, S. Choy, Y. Jiang, D. Chan, Y. Lim, L. Bera, L. Wong, W. Li, A. Du, C. Tung, K. Hoe, G. Lo, N. Balasubramanian, D.-L. Kwong, "Reduction of Leakage and Low-Frequency Noise in MOS Transistors Through Two-Step RTA of NiSi-Silicide Technology", IEEE Electron Device Letters, **27**(10), 824-826, 2006

[97] . N. Lukyanchikova, N. Garbar, A. Smolanka, V. Kudina, C. Claeys, E. Simoen, "Analytical Model for the Impact of the Twin Gate on the Floating-Body-Related Low-Frequency Noise Overshoot in Silicon-on-Insulator MOSFETs", IEEE Trans. Electron Devices, **53**(12), 3118-3128, 2006

[98] . M. Yang, C. Kuo, A. Chang, Y. Wang, S. Liu, "Statistical Characterization and Monte-Carlo Simulation of Low-Frequency Noise Variations in Foundry AMS/RF CMOS Technology", Silicon Monolithic Integrated Circuits in RF Systems, Digest of Papers SiRF, 312-315, 2006

[99] . Z. Rittersma, M. Vertregt, W. Deweerd, S. van Elshocht, P. Srinivasan, E. Simoen, "Characterization of Mixed-Signal Properties of MOSFETs With High-k (SiON/HfSiON/TaN) Gate Stacks", IEEE Trans. Electron Devices, **53**(5), 1216-1225, 2006

[100] . F. Crupi, P. Srinivasan, P. Magnone, E. Simoen, C. Pace, D. Misra, C. Claeys, "Impact of the Interfacial Layer on the Low-Frequency Noise (1/f) Behavior of MOSFETs with Advanced Gate Stacks", IEEE Electron Device Letters, **27**(8), 688-691, 2006

[101] . G. Giusi, F. Crupi, C. Pace, C. Ciofi, G. Groeseneken, "Comparative Study of Drain and Gate Low-Frequency Noise in nMOSFETs with Hafnium-Based Gate Dielectrics", IEEE Trans. Electron Devices, **53**(4), 823-827, 2006

[102] . V. Re, M. Manghisoni, L. Ratti, V. Speziali, G. Traversi, "Total Ionizing Dose Effects on the Noise Performances of a 0.13 μm CMOS Technology", IEEE Trans. Nuclear Science, **53**(3), 1599-1606, 2006

[103] . K. Kakushima, P. Ahmet, N. Sugii, K. Tsutsui, T. Hattori, H. Iwai, "Lanthanum Oxides for Gate Insulator Applications", 209th ECS Meeting, **2**(1), 115, 2006

[104] . V. Re, M. Manghisoni, L. Ratti, V. Speziali, G. Traversi, "Design Criteria for Low Noise Front-End Electronics in the 0.13 μm CMOS Generation", Nuclear Instrum. Meth. Phys. Res. A, **568**(1), 343-349, 2006

[105] . M. Manghisoni, L. Ratti, V. Re, V. Speziali, G. Traversi, "130 and 90nm CMOS Technologies for Detector Front-End Applications", Nuclear Instrum. Meth. Phys. Res. A, **572**(1), 368-370, 2007

[106] . A. Ahsan, S. Ahmed, "Degradation of 1/f Noise in Short Channel MOSFETs due to Halo Angle Induced VT Non-Uniformity and Extra Trap States at Interface", Solid-State Electronics, **50**(11-12), 1705-1709, 2006

[107] . N. Sghaier, M. Trabelsi, Ne. Sghaier, L. Militaru, A. Souifi, A. Kalboussi, N. Yacoubi, "Static and Low-Frequency Noise Characterization in Submicron MOSFETs for Memories Cells Applications", Microelectronics Journal, **37**(11), 1399-1403, 2006

[108] . C.-M. Lai, Y.-K. Fang, C.-T. Lin, C.-W. Hsu, W.-K. Yeh, "The Impacts of High Tensile Stress CESL and Geometry Design on Device Performance and Reliability for 90 nm SOI nMOSFETs", Microelectron. Reliab., **47**(6), 944-952, 2007

[109] . P. Srinivasan, E. Simoen, B. De Jaeger, C. Claeys, D. Misra, "1/f Noise Performance of MOSFETs with HfO_2 and Metal Gate on Ge-on-Insulator Substrates", Mater. Sci. Semicond. Process., **9**(4-5), 721-726, 2006

[110] . T. Contaret, K. Romanjek, T. Boutchacha, G. Ghibaudo, F. Boeuf, "Low Frequency Noise Characterization and Modelling in Ultrathin Oxide MOSFETs", Solid-State Electronics, **50**(1), 63-68, 2006

[111] . C. Cordier, A. Boukhenoufa, L. Pichon, J. Michaud, "Low Frequency Noise Model in N-MOS Transistors Operating from Sub-Threshold to Above-Threshold Regions ", Solid-State Electronics, **49**(8), 1376-1380, 2005

[112] . C. Leyris, F. Martinez, A. Hoffmann, M. Valenza, J. Vildeuil, "N-MOSFET Oxide Trap Characterization Induced by Nitridation Process Using RTS Noise Analysis", Microelectron. Reliab., **47**(1), 41-45, 2007

[113] . F. Balestra, G. Ghibaudo, J. Jomaah, “J. Modeling of low-frequency noise in advanced CMOS devices”, International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, Wiley, **28**, 613-627, 2015

[114] . Y. Liu, H.-B. Chen, Y.-R. Liu, X. Wang, Y.-F- En, B. Li, Y.-D. Lu, “Low frequency noise and radiation response in the partially depleted SOI MOSFETs with ion implanted buried oxide”, Chinese Physics B, IOP Publishing, **24**, 088503, 2015

[115] . E. Simoen, G. M. C. Andrade, L. M. Almeida, M. Aoulaiche, C. Caillat, M. Jurczak, C. Claeys, “On the Variability of the Low-Frequency Noise in UTBOX SOI nMOSFETs ECS Transactions”, The Electrochemical Society, **49**, 51-58, 2012

[116] . B. C. Wang, S. L. Wu, Y. Y. Lu, S. J. Chang, J. F. Chen, S. C. Tsai, C. H. Hsu, C. W. Yang, C. G. Chen, O. Cheng, P. C. Huang, “Comparison of the trap behavior between ZrO2 and HfO2 gate stack nMOSFETs by 1/f noise and random telegraph noise”, IEEE Electron Device Letters, **34**, 151-153, 2013

[117] . S. Cristoloveanu, M. Bawedin, I. Ionica, I. “A review of electrical characterization techniques for ultrathin FDSOI materials and devices Solid-State Electronics”, Elsevier BV, **117**, 10-36, 2016

[118] . C. G. Theodorou, E. G. Ioannidis, F. Andrieu, T. Poiroux, O. Faynot, C. A. Dimitriadis, G. Ghibaudo, “Low-Frequency Noise Sources in Advanced UTBB FD-SOI MOSFETs”, IEEE Trans. Electron Devices, **61**, 1161-1167, 2014

[119] . M. A. S. de Souza, R. T. Doria, M. de Souza, J. A. Martino, M. A. Pavanello, “Comparative Study of Biaxial and Uniaxial Mechanical Stress Influence on the Low Frequency Noise of Fully Depleted SOI nMOSFETs Operating in Triode and Saturation Regime”, ECS Transactions, The Electrochemical Society, **49**, 77-83, 2012

[120] . L. Pichon, B. Cretu, A. Boukhenoufa, “Thermal dependence of low-frequency noise in polysilicon thin film transistors”, Thin Solid Films, Elsevier BV, **517**, 6367-6370, 2009

[121] . L. Pichon, A. Boukhenoufa, B. Cretu, R. Rogel, “Improvement in the determination by 1/f noise measurements of the interface state distribution in polysilicon thin film transistors in relation with the compensation law of Meyer Neldel”, Journal of Applied Physics, AIP Publishing, **105**, 104503, 2009

[122] . T. Nguyen, A. Savio, L. Militaru, C. Plossu, “Spatial distribution of electrically active defects in dual-layer (SiO_2/HfO_2) gate dielectric n-type metal oxide semiconductor field effect transistors”, Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, American Vacuum Society, **27**, 329, 2009

[123] . P. Gaubert, A. Teramoto, W. Cheng, T. Ohmi, “Relation Between the Mobility, Noise, and Channel Direction in MOSFETs Fabricated on (100) and (110) Silicon-Oriented Wafers”, IEEE Transactions on Electron Devices, **57**, 1597-1607, 2010

[124] . A. Ahsan, D. Schroder, "Impact of Post-Oxidation Annealing on Low-Frequency Noise, Threshold Voltage, and Subthreshold Swing of p-Channel MOSFETs", IEEE Electron Device Letters, **25**(4), 211-213, 2004

[125] . P. Gaubert, A. Teramoto, T. Hamada, M. Yamamoto, K. Kotani, T. Ohmi, "1/f Noise Suppression of pMOSFETs Fabricated on Si(100) and Si(110) Using an Alkali-Free Cleaning Process", IEEE Trans. Electron Devices, **53**(4), 851-856, 2006

[126] . G. Giusi, E. Simoen, G. Eneman, P. Verheyen, F. Crupi, K. De Meyer, C. Claeys, C. Ciofi, "Low-Frequency (1/f) Noise Behavior of Locally Stressed HfO_2/TiN Gate-Stack pMOSFETs", IEEE Electron Device Letters, **27**(6), 508-510, 2006

[127] . P. Fantini, G. Ferrari, "Low Frequency Noise and Technology Induced Mechanical Stress in MOSFETs", Microelectron. Reliab., **47**(8), 1218-1221, 2007

[128] . P. Srinivasan, E. Simoen, R. Singanamalla, H. Yu, C. Claeys, D. Misra, "Gate Electrode Effects on Low-Frequency (1/f) Noise in p-MOSFETs with High-k Dielectrics", Solid-State Electronics, **50**(6), 992-998, 2006

[129] . P. Gaubert, A. Teramoto, R. Kuroda, Y. Nakao, H. Tanaka, S. Sugawa, T. Ohmi, “Analysis of the Low-Frequency Noise Reduction in Si(100) Metal-Semiconductor Field-Effect Transistors”, Japanese Journal of Applied Physics, IOP Publishing, **50**, 04DC01, 2011

[130] . C.-Y. Chen, Y. Liu, R. W. Dutton, J. Sato-Iwanaga, A. Inoue, H. Sorada, “Numerical Study of Flicker Noise in p-Type Si0.7Ge0.3/Si Heterostructure MOSFETs”, IEEE Transactions on Electron Devices, Institute of Electrical and Electronics Engineers (IEEE), **55**, 1741-1748, 2008

[131] . D. Kim, J. Kim, H. Kang, J. W. Shim, J. W. Lee, “Influence of flexible substrate in low temperature polycrystalline silicon thin-film transistors: temperature dependent characteristics and low frequency noise analysis”, Nanotechnology, IOP Publishing, **31**, 435201, 2020

[132] . P. Gaubert, A. Teramoto, W. Cheng, T. Hamada, T. Ohmi, “Different mechanism to explain the 1/f noise in n- and p-SOI-MOS transistors fabricated on (110) and (100) silicon-oriented wafers”, Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, American Vacuum Society, **27**, 394, 2009

[133] . A. Roy, C. Enz, "Critical Discussion on the Flatband Perturbation Technique for Calculating Low-Frequency Noise", IEEE Trans. Electron Devices, **53**(10), 2664-2667, 2006

[134] . O. Jantsch, "Flicker (1/f) Noise Generated by a Random Walk of Electrons in Interfaces", IEEE Trans. Electron Devices, **34**(5), 1100-1115, 1987

[135] . M. J. Deen, M. Levinshtein, S. Rumyantsev, J. Orchard-Webb, "Generation–Recombination Noise in MOSFETs", IOP Semicond. Sci. Technol., **14**(3), 298-304, 1999

[136] . A. Arnaud, C. Galup-Montoro, "Consistent Noise Models for Analysis and Design of CMOS Circuits", IEEE Trans. Circuits Systems, **51**(10), 1909-1915, 2004

[137] . T. Kramer, R. Pease, "Low Frequency Noise in Sub-100 nm MOSFETs", Physica E, **19**(1-2), 13-17, 2003

[138] . J. Chang, C. Viswanathan, "Low Frequency Noise in CMOSTs at Cryogenic Temperatures", Proc. Low Temperature Semiconductor Electronics, (Workshop), 133-136, 1989

[139] . P. Morfouli, G. Ghibaudo, T. Ouisse, E. Vogel, W. Hill, V. Misra, P. McLarty, J. Wortman, "Low-Frequency Noise Characterization of n- and p-MOSFET' s with Ultrathin Oxynitride Gate Films", IEEE Electron Device Letters, **17**(8), 395-397, 1996

[140] . E. Simoen, A. Mercha, L. Pantisano, C. Claeys, E. Young, "Tunneling 1/f$^{\gamma}$ Noise in 5 nm HfO_2/2.1 nm SiO_2 Gate Stack n-MOSFETs", Solid-State Electronics, **49**(5), 702-707, 2005

[141] . K. Narasimhulu, I. Setty, V. Rao, "The Effect of Single-Halo Doping on the Low-Frequency Noise Performance of Deep Submicrometer MOSFETs", IEEE Electron Device Letters, **27**(12), 995-997, 2006

[142] . T. Ishida, N. Tega, Y. Mori, H. Miki, T. Mine, H. Kume, K. Torii, R.-i Yamada, K. Shiraishi, "Mechanism of state transition of a defect causing random-telegraph-noise-induced fluctuation in stress-induced leakage current of SiO_2 films", Japanese Journal of Applied Physics, **53**, 08LB01, 2014

[143] . M. R. Hasan, A. Motayed, M. S. Fahad, M. V. Rao, "Fabrication and comparative study of DC and low frequency noise characterization of GaN/AlGaN based MOS-HEMT and HEMT", Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena, **35**, 052202, 2017

[144] . X. Zhang, M. H. White, "A quantum mechanical treatment of low frequency noise in high-K NMOS transistors with ultra-thin gate dielectrics", Solid-State Electronics, **78**, 131-135, 2012

[145] . F. Crupi, G. Giusi, G. Iannaccone, P. Magnone, C. Pace, E. Simoen, C. Claeys, "Analytical model for the 1⁄f noise in the tunneling current through metal-oxide-semiconductor structures", J. Appl. Phys., **106**, 073710, 2009

[146] . D. Bauza, G. Ghibaudo, "Analytical Study of the Contribution of Fast and Slow Oxide Traps to the Charge Pumping Current in MOS Structures", Solid-State Electronics, **39**(4), 563-570, 1996

[147] . Y. Maneglia, F. Rahmoune, D. Bauza, "On the Si-SiO_2 Interface Trap Time Constant Distribution in Metal-Oxide-Semiconductor Transistors", J. Appl. Phys., **97**(1), 014502-1 to 8, 2005

[148] . F. Martinez, C. Leyris, G. Neau, M. Valenza, A. Hoffmann, J. Vildeuil, E. Vincent, F. Boeuf, T. Skotnicki, M. Bidaud, D. Barge, B. Tavel , "Oxide Traps Characterization of 45 nm MOS Transistors by Gate Current R.T.S. Noise Measurements ", Microelectronic Engineering, **80**(1), 54-57, 2005

[149] M. J. Deen, O. Marinov, Y. Kiyota, "Low-Frequency Noise in RF nMOS Transistors from a 70 nm CMOS Technology", AIP Conf. Proc. (ICNF 2007 - 19th International Conference on Noise and Fluctuations), **922**(1), 129-132, 2007

[150] . O. Marinov, M. J. Deen, "Input 1/f Noise Voltage: S_{V_G} vs. Oxide Thickness (EOT) of MOS Transistors; and S_{V_B} vs. IFO of BJT", Raw Data in Electronics (www.rdie.ca), **1**(3), 75, 2007

[151] . D. Yue-Hua, C. Jun-Ning, K. Dao-Ming, X. Chao, S. Jia-E, "Modeling for Reduced Gate Capacitance of Nanoscale MOSFETs", Solid-State Electronics, **50**(7-8), 1472-1474, 2006

[152] T. Sugii, "High-Performance Bulk CMOS Technology for 65/45 nm Nodes", Solid-State Electronics, **50**(1), 2-9, 2006

[153] . C. Petit, A. Meinertzhagen, D. Zander, O. Sirnonetti, M. Fadlallah, T. Maurel, "Comparison of Ultra-Thin Gate Oxide Degradation in P and N-MOSFETs", Proc. 24th Conf. Microelectronics (MIEL2004), **2**, 641-644, 2004

[154] . F. Crupi, "Threshold Voltage Instability and Low Frequency Noise in Hafnium-Based Gate Dielectrics", 210th ECS Meeting ,Eds.: S. Kar, S. De Gendt, M. Houssa, H. Iwai, D. Landheer, D. Misra, **3**(3), 205-214, 2006

[155] . E. Simoen, G. Eneman, C. Claeys, P. Verheyen, R. Delhougne, R. Loo, K. De Meyer, "Impact of Strain and Strain-Relaxation on the Low-Frequency Noise of SRB Silicon n-MOSFETs", Proc. Solid-State Device Research Conference, ESSDERC, 529-532, 2005

[156] . G. Ghibaudo, O. Roux, Ch. Nguen-Duc, F. Balestra, J. Brini, "Improved Analysis of Low Frequency Noise in Field-Effect MOS Transistors", Phys. Stat. Sol. (a), **124**(2), 571-581, 1991

[157] . S. Takagi, A. Toriumi, M. Iwase, H. Tango, "On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I-Effects of Substrate Impurity Concentration", IEEE Trans. Electron Devices, **41**(12), 2357-2362, 1994; see also pages 2363-2368

[158] . J. Koga, S. Takagi, A. Toriumi, "A Comprehensive Study of MOSFET Electron Mobility in Both Weak and Strong Inversion Regimes", IEDM 1994, Technical Digest, 475-478, 1994

[159] . E. Vandamme, L. Vandamme, "Critical Discussion on Unified 1/f Noise Models for MOSFETs", IEEE Trans. Electron Devices, **47**(11), 2146-2152, 2000

[160] . R. Voss, "1/f (Flicker) Noise: A Brief Review", Frequency Control Symp., 33rd, 40-46, 1979

[161] . J.-W. Wu, J.-W. You, H.-C. Ma, C.-C. Cheng, C.-F. Hsu, C.-S. Chang, G.-W. Huang, T. Wang, "Excess Low-Frequency Noise in Ultrathin Oxide n-MOSFETs Arising From Valence-Band Electron Tunneling", IEEE Trans. Electron Devices, **52**(9), 2061-2066, 2005

[162] . C. Leyris, J.-C. Vildeuil, F. Roy, F. Martinez, M. Valenza, A. Hoffmann, "Response of Correlated Double Sampling CMOS Imager Circuit to Random Telegraph Signal Noise", Proc. 6th Int. Caribbean Conference Devices, Circuits and Systems, 109-114, 2006
[163] . M. Ferrier, R. Clerc, G. Ghibaudo, F. Boeuf, T. Skotnicki, "Analytical Model for Quantization on Strained and Unstrained Bulk nMOSFET and its Impact on Quasi-Ballistic Current", Solid-State Electronics, **50**(1), 69-77, 2006
[164] . J. Ranuarez, M. J. Deen, C.-H. Chen, "A Review of Gate Tunneling Current in MOS Devices", Microelectron. Reliab., **46**(12), 1939-1956, 2006
[165] . H. Xiong, D. Fleetwood, B. Choi, A. Sternberg, "Temperature Dependence and Irradiation Response of 1/f-Noise in MOSFETs", IEEE Trans. Nuclear Science, **49**(6), 2718-2723, 2002
[166] . N. Amarasinghe, Z. Celik-Butler, A. Zlotnicka, F. Wang, "Model for Random Telegraph Signals in Sub-Micron MOSFETs", Solid-State Electronics, **47**(9), 1443-1449, 2003
[167] . L. Vandamme, D. Sodini, Z. Gingl, "On the Anomalous Behavior of the Relative Amplitude of RTS Noise", Solid-State Electronics, **42**(6), 901-905, 1998
[168] . E. Simoen, B. Dierickx, C. Claeys, G. Declerck, "Explaining the Amplitude of RTS Noise in Submicrometer MOSFETs", IEEE Trans. Electron Devices, **39**(2), 422-429, 1992
[169] . Y. Tsividis, "Operation and Modeling of the MOS Transistor", 2nd ed., McGraw-Hill, USA, 1999
[170] . L. Vandamme, M. Macucci, "1/f and RTS Noise in Submicron Devices: Faster is Noisier", AIP Proc. UPoN, **800**, 436-443, 2005
[171] F. Liu, M. Bao, H.-J. Kim, K. Wang, C. Li, X. Liu, C. Zhou, "Giant Random Telegraph Signals in the Carbon Nanotubes as a Single Defect Probe", Appl. Phys. Lett., **86**(16), 163102-1 to 3, 2005
[172] . J. Johnson, "Thermal Agitation of Electricity in Conductors", Phys. Rev., **32**(1), 97-109, 1928
[173] . H. Nyquist, "Thermal Agitation of Electric Charge in Conductors", Phys. Rev., **32**(1), 110-113, 1928
[174] . R. Jindal, "Compact Noise Models for MOSFETs", IEEE Trans. Electron Devices, **53**(9), 2051-2061, 2006
[175] . X. Jin, J.-J. Ou, C.-H. Chen, W. Liu, M. J. Deen, P. Gray, C. Hu, "An Effective Gate Resistance Model for CMOS RF and Noise Modeling", IEDM 1998, Technical Digest, 961-964, 1998
[176] . C.-H. Chen, M. J. Deen, "Direct Calculation of Metal-Oxide-Semiconductor Field Effect Transistor High Frequency Noise Parameters", J. Vac. Sci. Technol. A, 16(2), 850-854, 1998.
[177] . C.-H. Chen, M. J. Deen, "High Frequency Noise of MOSFETs I Modeling", Solid-State Electronics, **42**(11), 2069-2081, 1998
[178] . C.-H. Chen, M. J. Deen, Z. Yan, M. Schroter, C. Enz, "High Frequency Noise of MOSFETs. II. Experiments", Solid-State Electronics, **42**(11), 2083-2092, 1998
[179] . M. J. Deen, C.-H. Chen, "The Impact of Noise Parameter De-embedding on the High-Frequency Noise Modeling of MOSFETs", IEEE Proc. Int. Conf. Microelectronic Test Structures, **ICMTS 12**, 34-39, 1999
[180] . W. Kwan, C.-H. Chen, M. J. Deen, "Hot-Carrier Effects on Radio Frequency Noise Characteristics of LDD n-Type Metal–Oxide–Semiconductor Field Effect Transistors", J. Vac. Sci. Technol. A, **18**(2), 765-769, 2000.
[181] . C.-H. Chen, M. J. Deen, "Extraction of the Induced Gate Noise, Channel Noise, and Their Correlation in Submicron MOSFETs from RF Noise Measurements", IEEE Trans. Electron Devices, **48**(12), 2884-2892, 2001
[182] . C.-H. Chen, M. J. Deen, "A General Noise and S-Parameter Deembedding Procedure for On-Wafer High-Frequency Noise Measurements of MOSFETs", IEEE Trans. on Microwave Theory and Techniques, **49**(5), 1004=1005, 2001
[183] . C.-H. Chen, M. J. Deen, "High-Frequency Small Signal AC and Noise Modeling of MOSFETs for RF IC Design", IEEE Trans. Electron Devices, **49**(3), 400-408, 2002
[184] . C.-H. Chen, M. J. Deen, "Channel Noise Modeling of Deep Submicron MOSFETs", IEEE Trans. Electron Devices, **49**(8), 1484-1487, 2002
[185] . S. Asgaran, M. J. Deen, C.-H. Chen, "Analytical Modeling of MOSFETs Channel Noise and Noise Parameters", IEEE Trans. Electron Devices, **51**(12), 2109-2114, 2004
[186] . M. J. Deen, C.-H. Chen, S. Asgaran, G. Rezvani, J. Tao, Y. Kiyota, "High-Frequency Noise of Modern MOSFETs: Compact Modeling and Measurement Issues", IEEE Trans. Electron Devices, **53**(9), 2062-2081, 2006
[187] . S. Asgaran, M. J. Deen, C.-H. Chen, G. Rezvani, Y. Kamali, Y. Kiyota "Analytical Determination of MOSFET's High-Frequency Noise Parameters From NF50 Measurements and Its Application in RFIC Design", IEEE J. Solid-State Circuits, **42**(5), 1034-1043, 2007
[188] . K. Kandiah, M. Deighton, F. Whiting, "A Physical Model for Random Telegraph Signal Currents in Semiconductor Devices", J. Appl. Phys., **66**(28), 937-948, 1989
[189] . M. Manghisoni, L. Ratti, V. Re, V. Speziali, G. Traversi, G. Fallica, S. Leonardi, "Noise Analysis of NPN SOI Bipolar Transistors for the Design of Charge Measuring Systems", IEEE Trans. Nuclear Science, **51**(3), 980-986, 2004

[190] . M. Weatherspoon, L. Dunleavy, "Vector Corrected On-Wafer Measurements of Noise Temperature", IEEE Trans. Instrum. Meas., **54**(3), 1327-1332, 2005
[191] . R. Hu, "Analysis of the Input Noise Contribution in the Noise Temperature Measurements", IEEE Microwave Wireless Comp. Lett., **15**(3), 141-143, 2005
[192] . J. Randa, E. Gerecht, D. Gu, R. Billinger, "Precision Measurement Method for Cryogenic Amplifier Noise Temperatures Below 5 K", IEEE Trans. Microwave Theory Techniques, **54**(3), 1180-1189, 2006
[193] . M. Aguilar, M. Rodriguez, J. Lopez-Bonilla, P. Crozat, T. Hackbarth, J.-H. Herzog, F. Aniel, "Noise Behavior of Buried Channel SiGe HFETs for High Speed Circuit's Applications", Proc. CIE 2005, Mexico City, 419-423, 2005
[194] . H. Heffner, "The Fundamental Noise Limit of Linear Amplifiers", Proc. IRE, **50**(7), 1604-1608, 1962
[195] . H. Buckmaster, "An Alternative Derivation of Fundamental Noise Limit of Linear Amplifiers", Proc. of IEEE, **53**(2), 194-194, 1965
[196] . K. Kurokawa, "Power Waves and the Scattering Matrix", IEEE Trans. Microwave Theory Techniques, **13**(2), 194-202, 1965
[197] . D. Meer, "Noise figures [two-port linear transducers]", Proc. IEEE National Aerospace and Electronics Conf., NAECON 1989, **1**, 269-276, 1989
[198] . S. Naseh, M. J. Deen, C.-H. Chen, "Effects of Hot-Carrier Stress on the Performance of CMOS Low-Noise Amplifiers", IEEE Transactions on Device and Materials Reliability, **5**(3), 501-508, 2005
[199] . S. Naseh, M. J. Deen, C.-H. Chen, "Hot-Carrier Reliability of Submicron NMOSFETs and Integrated NMOS Low Noise Amplifiers", Elsevier Microelectronics Reliability, **46**(2-4), 201-212, 2006.
[200] . H. Friis, "Noise Figures of Radio Receivers", Proc. IRE, **32**(7), 419-422, 1944
[201] . H. Rothe, W. Dahlke, "Theory of Noisy Fourpoles", Proc. IRE, **44**(6), 811-818, 1956
[202] . P. Gaubert, A. Teramoto, S. Sugawa, "Impact of doping concentration on 1/fnoise performances of accumulation-mode Si(100) n-MOSFETs", Japanese Journal of Applied Physics, **55**, 04ED08, 2016
[203] . C. Peng, Y. F. En, Z.-F. Lei, Y. Q. Chen, Y. Liu, B. Li, "Influence of Total Ionizing Dose Irradiation on Low-Frequency Noise Responses in Partially Depleted SOI nMOSFETs", Chinese Physics Letters, **34**, 118501 , 2017
[204] . Z. Çelik-Butler, M. Mahmud, P. Hao, F. Hou, B. Amey S. Pendharkar, "Determination of active oxide trap density and 1/f noise mechanism in RESURF LDMOS transistors", Solid-State Electronics, **111**, 141-146 , 2015
[205] . S. A. Francis, A. Dasgupta, D. M. Fleetwood, "Effects of Total Dose Irradiation on the Gate-Voltage Dependence of the 1/f Noise of nMOS and pMOS Transistors", IEEE Transactions on Electron Devices, **57**, 503-510 , 2010
[206] . Z. Çelik-Butler, S. P. Devireddy, H. H. Tseng, P. Tobin, A. Zlotnicka, "A low-frequency noise model for advanced gate-stack MOSFETs", Microelectronics Reliability, Elsevier BV, **49**, 103-112 , 2009
[207] . W.-C. Hua, M. Lee, P. Chen, S. Maikap, C. Liu, K. Chen, "Ge Outdiffusion Effect on Flicker Noise in Strained-Si nMOSFETs", IEEE Electron Device Letters, **25**(10), 693-695, 2004
[208] . M. Myronov, O. Mironov, S. Durov, T. Whall, E. Parker, T. Hackbarth, G. Hock, H.-J. Herzog, U. Konig, "Reduced 1/f Noise in p-$Si_{0.3}Ge_{0.7}$ Metamorphic Metal-Oxide-Semiconductor Field-Effect Transistor", Appl. Phys. Lett., **84**(4), 610-612, 2004
[209] . O. Marinov, M. J. Deen, "Oxide Trap Density and Correlated Mobility Noise Parameter vs. MOS Transistor Oxide Thickness EOT", Raw Data in Electronics (www.rdie.ca), **1**(3), 76, 2007
[210] . J. Tang, G. Niu, Z. Jin, J. Cressler, S. Zhang, A. Joseph, D. Harame, "Modeling and Characterization of SiGe HBT Low-Frequency Noise Figures-of-Merit for RFIC Applications", IEEE Trans. Microwave Theory Techniques, **50**(11), 2467-2473, 2002
[211] . F. Pascal, C. Chay, M.J. Deen, S. G-Jarrix, C. Delseny, A. Penarier, "Comparison of Low-Frequency Noise in III–V and Si/SiGe HBTs", IEE Proc.-Circuits Devices Syst., **151**(2), 138-147, 2004
[212] . G. Niu, J. Tang, Z. Feng, A. Joseph, D. Harame, "Scaling and Technological Limitations of 1/f Noise and Oscillator Phase Noise in SiGe HBTs", IEEE Trans. Microwave Theory Techniques, **53**(2), 506-514, 2005
[213] . J. Cressler, "On the Potential of SiGe HBTs for Extreme Environment Electronics", Proc. of IEEE, **93**(9), 1559-1582, 2005
[214] . C.-C. Ho, G.-H. Liang, C.-F. Huang, Y.-J. Chan, C.-S. Chang, and C.-P. Chao, "VCO Phase-Noise Improvement by Gate-Finger Configuration of 0.13-μm CMOS Transistors", IEEE Electron Device Letters, **26**(4), 258-260, 2005
[215] . E. Haller, "Germanium: From its Discovery to SiGe Devices", Mater. Sci. Semicond. Process., **9**(4-5), 408-422, 2006
[216] . M. K. Cho, I. Song, Z. E. Fleetwood, J. D. Cressler, "A SiGe-BiCMOS Wideband Active Bidirectional Digital Step Attenuator With Bandwidth Tuning and Equalization", IEEE Transactions on Microwave Theory and Techniques, **66**, 3866-3876 , 2018

[217] . J. D. Chen, S. H. Wang, “A Low-Power, High-Gain, and Low-Noise 802.11a Down-Conversion Mixer in 0.35-upmum SiGe Bi-CMOS Technology”, Journal of Circuits, Systems and Computers, **26**, 1750134 , 2017
[218] . N. Sarmah, J. Grzyb, K. Statnikov, S. Malz, P. R. Vazquez, W. Foerster, B. Heinemann, U. R. Pfeiffer, “A Fully Integrated 240-GHz Direct-Conversion Quadrature Transmitter and Receiver Chipset in SiGe Technology”, IEEE Transactions on Microwave Theory and Techniques, **64**, 562-574, 2016
[219] . A. Luque Rodríguez, M. Bargallo Gonzalez, G. Eneman, C. Claeys, D. Kobayashi, E. Simoen, J. A. Jiménez Tejada, “Impact of Ge Content and Recess Depth on the Leakage Current in Strained Heterojunctions”, IEEE Trans. Electron Devices, **58**(8), 2362-2370, 2011
[220] . T. Ando, P. Hashemi, J. Bruley, J. Rozen, Y. Ogawa, S. Koswatta, K. K. Chan, E. A. Cartier, R. Mo, V. Narayanan, “High Mobility High-Ge-Content SiGe PMOSFETs Using Al_2O_3/HfO_2 Stacks With In-Situ O_3 Treatment”, IEEE Electron Device Letters, *38,* 303-305, 2017
[221] . T. E. Lee, M. Ke, K. Kato, M. Takenaka, S. Takagi, “Metal-oxide-semiconductor interface properties of $TiN/Y_2O_3/Si_{0.62}Ge_{0.38}$ gate stacks with high temperature post-metallization annealing”, Journal of Applied Physics, **127**, 185705, 2020
[222] . I. A. Pindoo, S. K. Sinha, S. Chander, “Improvement of Electrical Characteristics of SiGe Source Based Tunnel FET Device”, Silicon, Springer Science and Business Media LLC, 2020
[223] . T. E. Lee, M. Ke, K. Toprasertpong, M. Takenaka, S. Takagi, “Reduction of MOS Interface Defects in $TiN/Y_2O_3/Si_{0.78}$ $Ge_{0.22}$ Structures by Trimethylaluminum Treatment”, IEEE Transactions on Electron Devices, **67**, 4067-4072, 2020
[224] . J. Tang, G. Niu, Z. Feng, A. Joseph, D. Harame, "Impact of Collector-Base Junction Traps on Low-Frequency Noise in High Breakdown Voltage SiGe HBTs", IEEE Trans. Electron Devices, **51**(9), 1475-1482, 2004
[225] . N. Lukyanchikova, N. Garbar, A. Smolanka, M. Lokshin, S. Hall, O. Buiu, I. Mitrovic, H. Mubarek, P. Ashburn, "The Base Current and Related 1/f Noise for SiGe HBTs Realized by SEG/NSEG Technology on SOI and Bulk Substrates", Mater. Sci. Semicond. Process., **9**(4-5), 727-731, 2006
[226] . O. Marinov, M. J. Deen, “1/f Noise Coefficient Kf vs. Emitter Area of SiGe Heterojunction Bipolar Transistors (HBT, BJT)”, Raw Data in Electronics (www.rdie.ca), **1**(3), 77, 2007
[227] . M. Haartman, M. Sanden, M. Ostling, G. Bosman, "Random Telegraph Signal Noise in SiGe Heterojunction Bipolar Transistors", J. Appl. Phys., **92**(8), 4414-4421, 2002
[228] . M. Seif, F. Pascal, B. Sagnes, S. Haendler, “Characterization and modeling of low frequency noise in 0.13 mu m BiCMOS SiGe: C heterojunction bipolar trasnsistors”, 10th Conference on Ph D Research in Microelectronics and Electronics (PRIME), 2014
[229] . M. Seif, F. Pascal, B. Sagnes, A. Hoffmann, S. Haendler, P. Chevalier, D. Gloria, “Low frequency noise measurements of advanced BiCMOS SiGeC Heterojunction bipolar Transistors used for mm-wave to terahertz applications”, 22nd International Conference on Noise and Fluctuations (ICNF), 1-4 , 2013
[230] . G. Avenier, M. Diop, P. Chevalier,G. Troillard, N. Loubet, J. Bouvier, L. Depoyan, N. Derrier, M. Buczko, C. Leyris, S. Boret, S Montusclat, A. Margain, S. Pruvost, S. Nicolson, K. Yau, N. Revil, D. Gloria, D. Dutartre, S. Voinigescu, A. Chantre, “0.13 µm SiGe BiCMOS Technology Fully Dedicated to mm-Wave Applications”, IEEE Journal of Solid-State Circuits, **44**(9), 2312-2321, 2009
[231] . M. Prest, A. Bacon, D. Fulgoni, T. Grasby, E. Parker, T. Whall, A. Waite, "Low-Frequency Noise Mechanisms in Si and Pseudomorphic SiGe p-Channel Field-Effect Transistors", Appl. Phys. Lett., **85**(24), 6019-6021, 2004
[232] . G. Ghibaudo, J. Chroboczek, "On the Origin of the LF Noise in Si/Ge MOSFETs", Solid-State Electronics, **46**(3), 393-398, 2002
[233] . F. Assaderaghi, D. Sinitsky, S. Parke, J. Bokor, P. Ko, C. Hu, "Dynamic Threshold-Voltage MOSFET (DTMOS) for Ultra-Low Voltage VLSI", IEEE Trans. Electron Devices, **44**(3), 414-422, 1997
[234] . N. Lindert, T. Sugii, S. Tang, C. Hu, "Dynamic Threshold Pass-Transistor Logic for Improved Delay at Lower Power Supply Voltages", IEEE J. Solid-State Circuits, **34**(1), 85-89, 1999
[235] . H. Achigui, C. Fayomi, M. Sawan, "1-V DTMOS-Based Class-AB Operational Amplifier: Implementation and Experimental Results", IEEE J. Solid-State Circuits, **41**(11), 2440-2448, 2006
[236] . J. Rosenfeld, M. Kozak, E. Friedman, "A Bulk-Driven CMOS OTA with 68 dB DC Gain", Proc. ICECS, IEEE, 5-8, 2004
[237] . S. Chatterjee, Y. Tsividis, P. Kinget, "0.5-V Analog Circuit Techniques and their Application in OTA and Filter Design", IEEE J. Solid-State Circuits, **40**(12), 2373-2387, 2005
[238] . M. Maymandi-Nejad, M. Sachdev, "DTMOS Technique for Low-Voltage Analog Circuits", IEEE Trans. VLSI Sys., **14**(10), 1151-1156, 2006

[239] . M. J. Deen, R. Murji, A. Fakhr, N. Jafferali, W. Ngan, "Low-Power CMOS Integrated Circuits for Radio Frequency Applications", IEE Proc.-Circuits Devices Syst., **152**(5), 509-522, 2005
[240] . L. Geynet, E. De Foucauld, P. Vincent, G. Jacquemod, "Fully-Integrated Multi-Standard VCOs with Switched LC Tank and Power Controlled by Body Voltage in 130nm CMOS/SOI", IEEE RFIC, Symp., 4 pages, 2006
[241] . K. Ishibashi, S. Ohbayashi, K. Eikyu, M. Tanizawa, Y. Tsukamoto, K. Osada, M. Miyazaki, M. Yamaoka, "Circuit Technologies for Reducing the Power of SOC and Issues on Transistor Models", IEDM 2006, Technical Digest, 1-4, 2006
[242] . L. Clark, M. Morrow, W. Brown, "Reverse-Body Bias and Supply Collapse for Low Effective Standby Power", IEEE Trans. VLSI Sys., **12**(9), 947-956, 2004
[243] . H. Ananthan, C. Kim, K. Roy, "Larger-than-Vdd Forward Body Bias in Sub-0.5V Nanoscale CMOS", Proc. Low Power Electronics and Design, ISLPED '04, 8-13, 2004
[244] . J. Kao, M. Miyazaki, A. Chandrakasan, "A 175-mV Multiply-Accumulate Unit Using an Adaptive Supply Voltage and Body Bias Architecture", IEEE J. Solid-State Circuits, **37**(11), 1545-1554, 2002
[245] . M. J. Deen, O. Marinov, "Effect of Forward and Reverse Substrate Biasing on Low-Frequency Noise in Silicon PMOSFETs", IEEE Trans. Electron Devices, **49**(3), 409-413, 2002
[246] . K. Akarvardar, B. Dufrene, S. Cristoloveanu, P. Gentil, B. Blalock, M. Mojarradi, "Low-Frequency Noise in SOI Four-Gate Transistors", IEEE Trans. Electron Devices, **53**(4), 829-835, 2006
[247] . A. Asai, J. Sato-Iwanaga, A. Inoue, Y. Hara, Y. Kanzawa, H. Sorada, T. Kawashima, T. Ohnishi, T. Takagi, M. Kubo, "Low-Frequency Noise Characteristics in SiGe Channel Heterostructure Dynamic Threshold pMOSFET (HDTMOS)", IEDM 2002, Technical Digest, 35-38 , 2002
[248] . A. Jimenez, F. De la Hidalga, M. J. Deen, "Modelling of the Dynamic Threshold MOSFET", IEE Proc.-Circuits Devices Syst., **152**(5), 502-508, 2005
[249] . M. Marin, M. J. Deen, M. de Murcia, P. Llinares, J. Vildeuil, "Effects of Body Biasing on the Low Frequency Noise of MOSFETs from a 130 nm CMOS Technology", IEE Proc.-Circuits Devices Syst., **151**(2), 95-101, 2004
[250] . J. Jomaah, F. Balestra, "Low-Frequency Noise in Advanced CMOS/SOI Devices", IEE Proc.-Circuits Devices Syst., **151**(2), 111-117, 2004
[251] . S. Haendler, J. Jomaah, G. Ghibaudo, F. Balestra, "Improved Analysis of Low Frequency Noise in Dynamic Threshold MOS/SOI Transistors", Microelectron. Reliab., **41**(6), 855-860, 2001
[252] . C. Wann, K. Noda, T. Tanaka, M. Yoshida, C. Hu, "A Comparative Study of Advanced MOSFET Concepts", IEEE Trans. Electron Devices, **43**(10), 1742-1753, 1996
[253] . E. Simoen, C. Claeys, N. Lukyanchikova, N. Garbar, A. Smolanka, P. Agopian, J. Martino, "Electron Valence-Band Tunnelling Excess Noise in Twin-Gate Silicon-on-Insulator MOSFETs", Solid-State Electronics, **50**(1), 52-57, 2006
[254] . J. Chen, P. Fang, P. Ko, C. Hu, R. Solomon, T.-Y. Chan, C. Sodini, "Noise Overshoot at Drain Current Kink in SOI MOSFET", IEEE SOS/SOI, Technology Conference, 40-41, 1990
[255] . Y.-C. Tseng, W. Huang, M. Mendicino, D. Monk, P. Welch, J. Woo, "Comprehensive Study on Low-Frequency Noise Characteristics in Surface Channel SOI CMOSFETs and Device Design Optimization for RF ICs", IEEE Trans. Electron Devices, **48**(7), 1428-1437, 2001
[256] . N. Lukyanchikova, N. Garbar, A. Smolanka, E. Simoen, C. Claeys, "Excess Lorentzian Noise in Partially Depleted SOI nMOSFETs Induced by an Accumulation Back-Gate Bias", IEEE Electron Device Letters, **25**(6), 433-435, 2004
[257] . E. Simoen, C. Claeys, N. Lukyanchikova, N. Garbar, A. Smolanka, "Linear-Kink-Noise Suppression in Partially Depleted SOI Using the Twin-Gate MOSFET Configuration", IEEE Electron Device Letters, **26**(7), 510-512, 2005
[258] . W. Jin, P. Chan, S. Fung, P. Ko, "A Physically-Based Low-Frequency Noise Model for NFD SOI MOSFET's", Proc. IEEE SOI, International Conference, 23-24, 1998
[259] . G. Workman, J. Fossum, "Physical Noise Modeling of SOI MOSFET's with Analysis of the Lorentzian Component in the Low-Frequency Noise Spectrum", IEEE Trans. Electron Devices, **47**(6), 1192-1201, 2000
[260] . M. Nakahara, H. Iwasawa, K. Yasutake, "Anomalous Low-Frequency Noise Enhancement Beyond Pinch-Off in Silicon n-Channel MOS Transistors", Proc. of IEEE, **57**(12), 2177-2178, 1969
[261] . S. Zhang, X. Lin, R. Huang, R. Han, M. Chan, "A Self-Aligned, Electrically Separable Double-Gate MOS Transistor Technology for Dynamic Threshold Voltage Application", IEEE Trans. Electron Devices, **50**(11), 2297-2300, 2003
[262] . T. Poiroux, M. Vinet, O. Faynot, J. Widiez, J. Lolivier, B. Previtali, T. Ernst, S. Deleonibus, "Multigate Silicon MOSFETs for 45 nm Node and Beyond", Solid-State Electronics, **50**(1), 18-23, 2006

[263] . J. Saint-Martin, A. Bournel, P. Dollfus, "Comparison of Multiple-Gate MOSFET Architectures Using Monte Carlo Simulation ", Solid-State Electronics, **50**(1), 94-101, 2006
[264] . A. Marchi, E. Gnani, S. Reggiani, M. Rudan, G. Baccarani, "Investigating the Performance Limits of Silicon-Nanowire and Carbon-Nanotube FETs", Solid-State Electronics, **50**(1), 78-85, 2006
[265] . J. A. Jiménez Tejada, A. Luque Rodriguez, A. Godoy, J. A. López Villanueva, F. Gómez-Campos, S. Rodríguez-Bolivar, "A Low-Frequency Noise Model for Four-Gate Field-Effect Transistors", IEEE Trans. Electron. Devices, **55**, 896 -903, 2008
[266] . J. A. Jiménez Tejada, A. Luque Rodríguez, A. Godoy, S. Rodríguez-Bolívar, J. A. López Villanueva, O. Marinov, M. J. Deen, "Effects of Gate Oxide and Junction Nonuniformity on the DC and Low-Frequency Noise Performance of Four-Gate Transistors", IEEE Trans. Electron Devices, ***59***, 459-467, 2012
[267] . P. Collins, M. Fuhrer, A. Zettl, "1/f Noise in Carbon Nanotubes", Appl. Phys. Lett., **76**(7), 894-896, 2000
[268] . S. Reza, Q. Huynh, G. Bosman, J. Sippel-Oakley, A. Rinzler, "Thermally Activated Low Frequency Noise in Carbon Nanotubes", J. Appl. Phys., **99**(11), 114309-1 to 4, 2006
[269] . T. Pedersen, "Variational Approach to Excitons in Carbon Nanotubes", Phys. Rev. B, **67**(7), 073401-1 to 4, 2003
[270] . G. Pennington, N. Goldsman, "Semiclassical Transport and Phonon Scattering of Electrons in Semiconducting Carbon Nanotubes", Phys. Rev. B, **68**(4), 045426-1 to 11, 2003
[271] . S. Reza, G. Bosman, M. Islam, T. Kamins, S. Sharma, R. Williams, "Noise in Silicon Nanowires", IEEE Trans. Nanotechnology, **5**(5), 523-529, 2006
[272] . R. Tarkiainen, L. Roschier, M. Ahlskog, M. Paalanen, P. Hakonen, "Low-Frequency Current Noise and Resistance Fluctuations in Multiwalled Carbon Nanotubes ", Physica E, **28**(1), 57-65, 2005
[273] . I. Lee, X. Liu, C. Zhou, B. Kosko, "Noise-Enhanced Detection of Subthreshold Signals with Carbon Nanotubes", IEEE Trans. Nanotechnology, **6**(5), 613-627, 2006
[274] . F. Liu, K. Wang, D. Zhang, C. Zhou, "Noise in Carbon Nanotube Field Effect Transistor", Appl. Phys. Lett., **89**(6), 63116-1 to 3, 2006
[275] . M. Briman, K. Bradley, G. Gruner, "Source of 1/ f Noise in Carbon Nanotube Devices", J. Appl. Phys., **100**(1), 013505-1 to 5, 2006
[276] . F. V. Gasparyan, A. Poghossian, S. A. Vitusevich, M. V. Petrychuk, V. A. Sydoruk, J: R. Siqueira, O. N. Oliveira, A. Offenhausser, M. J. Schoning, "Low-Frequency Noise in Field-Effect Devices Functionalized With Dendrimer/Carbon- Nanotube Multilayers", IEEE Sensors Journal, **11**, 142-149, 2011
[277] . A. V. de Oliveira, A. Veloso, C. Claeys, N. Horiguchi, E. Simoen, "Frequency Noise in Vertically Stacked Si - Channel Nanosheet FETs", IEEE Electron Device Letters, **41**, 317-320, 2020
[278] . G. Hellings, H. Mertens, A. Subirats, E. Simoen, T. Schram, L.-A. Ragnarsson, M. Simicic, S.-H. Chen, B. Parvais, D. Boudier, B. Cretu, J. Machillot, V. Pena, S. Sun, N. Yoshida, N. Kim, A. Mocuta, D. Linten, N. Horiguchi, "Si/SiGe superlattice I/O finFETs in a vertically-stacked Gate-All-Around horizontal Nanowire", Technology 2018 IEEE Symposium on VLSI Technology, IEEE, 2018
[279] . E. Simoen, M. G. C. de Andrade, M. Aoulaiche, N. Collaert, C. Claeys, "Low-Frequency-Noise Investigation of n-Channel Bulk FinFETs Developed for One-Transistor Memory Cells", IEEE Transactions on Electron Devices, **59**, 1272-127, 2012
[280] . D. Jang, J. W. Lee, C.-W. Lee, J.-P. Colinge, L. Montès, J. I. Lee, G. T. Kim, G. Ghibaudo, "Low-frequency noise in junctionless multigate transistors", Applied Physics Letters, **98**, 133502 , 2011
[281] . E. Simoen, A. Veloso, P. Matagne, N. Collaert, C. Claeys, "Junctionless Versus Inversion-Mode Gate-All-Around Nanowire Transistors From a Low-Frequency Noise Perspective", IEEE Transactions on Electron Devices, **65**, 1487-1492, 2018
[282] . P. Puczkarski, Q. Wu, H. Sadeghi, S. Hou, A. Karimi, Y. Sheng, J. H. Warner, C. J. Lambert, G. A. D. Briggs, J. A. Mol, "Low-Frequency Noise in Graphene Tunnel Junctions", ACS Nano, **12**, 9451-9460, 2018
[283] . M. Tretjak, S. Pralgauskaite, J. Matukas, J. Macutkevic, J. Banys, "Low frequency noise and resistivity characteristics of hybrid composites with onion-like carbon and multi-walled carbon nanotubes", Fluctuation and Noise Letters, **18**, 1940009, 2019
[284] . M. Troudi, Y. Bergaoui, P. Bondavalli, N. Sghaier, "Comparison of purity in single walled carbon nanotube gas detectors with Pd and Ti electrodes using low frequency noise techniques", Materials Research Bulletin, **99**, 292-297, 2018
[285] . S. Pralgauskaite, J. Matukas, M. Tretjak, J. Macutkevic, J. Banys, A. Selskis, A. Cataldo, F. Micciulla, S. Bellucci, V. Fierro, A. Celzard, "Resistivity and low-frequency noise characteristics of epoxy-carbon composites", J. Appl. Phys., **121**, 114303, 2017

[286] . G. Liu, S. Rumyantsev, M. A. Bloodgood, T. T. Salguero, M. Shur, A. A. Balandin, "Low-Frequency Electronic Noise in Quasi-1D $TaSe_3$ van der Waals Nanowires", Nano Letters, **17**, 377-383, 2016
[287] . C. Barone, G. Landi, C. Mauro, S. Pagano, H. Neitzert, "Low-frequency electric noise spectroscopy in different polymer/carbon nanotubes composites", Diamond and Related Materials, **65**, 32-36, 2016
[288] . Y.-M. Lin, J. C. Tsang, M. Freitag, P. Avouris, "Impact of oxide substrate on electrical and optical properties of carbon nanotube devices", Nanotechnology, **18**, 295202, 2007
[289] . T. Tanaka, E. Sano, "Low-frequency noise in carbon nanotube network thin-film transistors", Japanese Journal of Applied Physics, **53**, 090302, 2014
[290] . V. A. Sydoruk, K. Goß, C. Meyer, M. V. Petrychuk, B. A. Danilchenko, P. Weber, C. Stampfer, J. Li, S. A. Vitusevich, "Low-frequency noise in individual carbon nanotube field-effect transistors with top, side and back gate configurations: effect of gamma irradiation", Nanotechnology, 035703, 2013
[291] . M.-K. Joo, M. Mouis, D.-Y. Jeon, G.-T. Kim, U. J. Kim, G. Ghibaudo, "Static and low frequency noise characterization of N-type random network of carbon nanotubes thin film transistors", Journal of Applied Physics, **114**, 154503, 2013
[292] . D. Talukdar, P. Yotprayoonsak, O. Herranen, M. Ahlskog, "Linear current fluctuations in the power-law region of metallic carbon nanotubes", Physical Review B, **88**, 2013
[293] . X. Xie, D. Sarkar, W Liu, J. Kang, O. Marinov, M. J. Deen, K. Banerjee, "Low-Frequency Noise in Bilayer MoS_2 Transistor", ACS Nano, **8**(6), 5633-5640, 2014.
[294] . C. Enz, E. Vittoz, "MOS Transistor Modeling for Low-Voltage and Low-Power Analog IC Design", Microelectronic Engineering, **39**(1-4), 59-76, 1997
[295] . Linear Technology, "LT1007/LT1037 Low Noise, High Speed Precision Operational Amplifiers", Datasheet, www.linear.com, 1985
[296] . M. Sanden, M. J. Deen, "Low-Frequency Noise in Advanced Si-Based Bipolar Transistors and Circuits", in "Noise and Fluctuations Control in Electronic Devices", ed. A. Balandin, American Scientific Publishers, 235-247, 2002
[297] . R. Murji, M. J. Deen, "Noise Contributors in a 7.2 GHz Low-Power VCO with Automatic Amplitude Control", IEEE RFIC, Symp., 407-410, 2005
[298] . D. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proc. of IEEE, **54**(2), 329-330, 1966
[299] . H. Chen, A. van der Ziel, K. Amberiadis, "Oscillator with Odd- Symmetrical Characteristics Eliminates Low-Frequency Noise Sidebands", IEEE Trans. Circuits Systems, **31**(9), 807-809, 1984
[300] . A. A. L. de Souza, E. Dupouy, J. C. Nallatamby, M. Prigent, J. Obregon, "Experimental characterization of the cyclostationary low-frequency noise of microwave semiconductor devices under large signal operation", International Journal of Microwave and Wireless Technologies, **2**, 225-233, 2010
[301] . S. D. Vamvakos, V. Stojanovic, B. Nikolic, "Discrete-Time, Linear Periodically Time-Variant Phase-Locked Loop Model for Jitter Analysis", IEEE Transactions on Circuits and Systems I: Regular Papers, **58**, 1211-1224, 2011
[302] . S. Sancho, S. Hernandez, A. Suarez, "Noise Analysis of Super-Regenerative Oscillators in Linear and Nonlinear Modes", IEEE Transactions on Microwave Theory and Techniques, **67**, 4955-4965, 2019
[303] . C. Dragone, "Analysis of Thermal and Shot noise in Pumped Resistive Diodes", Bell Syst. Tech. J., **47**, 1883-1902, 1968
[304] . R. Poore, "Accurate Simulation of Mixer Noise and Oscillator Phase Noise in Large RFICs", Proc. APMC (Asia-Pacific Microwave Conference), **1**, 357-360, 1997
[305] . R. Poore, "Oscillator Phase noise Simulation Using ADS, Including Flicker Noise Conversion", presented at the IEEE MTT-S Int. Microwave Symp. Workshop, 2003
[306] . A. Demir, "Phase Noise in Oscillators: DAEs and Colored Noise Sources", IEEE-ICCAD (Internat. Conf. Computer-Aided Design), Digest Tech. Papers, 170-177, 1998
[307] . A. Demir, A. Mehrotra, J. Roychowdhury, "Phase Noise in Oscillators: A Unifying Theory and Numerical Methods for Characterization", IEEE Trans. Circuits Systems, **47**(5), 655-674, 2000
[308] . A. Demir, "Phase Noise and Timing Jitter in Oscillators with Colored-Noise Sources", IEEE Trans. Circuits Systems, **49**(12), 1782-1791, 2002
[309] . A. Demir, "Computing Timing Jitter From Phase Noise Spectra for Oscillators and Phase-Locked Loops With White and 1/f Noise", IEEE Trans. Circuits Systems, **53**(9), 1869-1884, 2006
[310] . C. Samori, A. Lacaita, A. Zanchi, F. Pizzolato, "Experimental Verification of the Link Between Timing Jitter and Phase Noise", Electronics Letters, **34**(21), 2024-2025, 1998
[311] . D. Allan, "Statistics of Atomic Frequency Standards", Proc. of IEEE, **54**(2), 221-230, 1966

[312] . A. Barnes, "Atomic Timekeeping and the Statistics of Precision Signal Generators", Proc. of IEEE, **54**(2), 207-220, 1966
[313] . F. Vernotte, J. Groslambert, J. Gagnepain, "A New Method of Measurement of the Different Types of Noise Altering the Output Signal of Oscillators", IEEE Trans. Instrum. Meas., **42**(6), 968-915, 1993
[314] . A. Chorti, M. Brookes, "A Spectral Model for RF Oscillators with Power-Law Phase Noise", IEEE Trans. Circuits Systems, **53**(9), 1989-1999, 2006
[315] . A. Hajimiri, T. Lee, "A General Theory of Phase Noise in Electrical Oscillators", IEEE J. Solid-State Circuits, **33**(2), 179-194, 1998; see also: -, "Corrections to "A General Theory ..."", -, **33**(6), 928, 1998
[316] . C. Samori, A. Lacaita, F. Villa, F. Zappa, "Spectrum Folding and Phase Noise in LC Tuned Oscillators", IEEE Trans. Circuits Systems, **45**(7), 781-790, 1998
[317] . A. Hajimiri, S. Limotyrakis, T. Lee, "Jitter and Phase Noise in Ring Oscillators", IEEE J. Solid-State Circuits, **34**(6), 790-804, 1999
[318] . L. Dai, R. Harjani, "Design of Low-Phase-Noise CMOS Ring Oscillators", IEEE Trans. Circuits Systems, **49**(5), 328-338, 2002
[319] . M. Grozing, M. Berroth, "Derivation of Single-Ended CMOS Inverter Ring Oscillator Close-In Phase Noise from Basic Circuit and Device Properties", IEEE RFIC, Symp., 277-280, 2004
[320] . R. Navid, T. Lee, R. Dutton, "Minimum Achievable Phase Noise of RC Oscillators", IEEE J. Solid-State Circuits, **40**(3), 630-637, 2005
[321] . J. Post, I. Linscott, M. Oslick, "Waveform Symmetry Properties and Phase Noise in Oscillators", Electronics Letters, **34**(16), 1547-1548, 1998
[322] . A. Hajimiri, T. Lee, "Phase Noise in CMOS Differential LC Oscillators", IEEE VLSI Circuits, Symp., 48-51, 1998
[323] . M. Margarit, J. Tham, R. Meyer, M. Deen, "A Low-Noise, Low-Power VCO with Automatic Amplitude Control for Wireless Applications", IEEE J. Solid-State Circuits, **34**(6), 761-771, 1999
[324] . M. Lax, "Classical Noise. V. Noise in Self-Sustained Oscillators", Phys. Rev., **160**(2), 290-307, 1967
[325] . A. Abidi, R. Meyer, "Noise in Relaxation Oscillators", IEEE J. Solid-State Circuits, **18**(6), 794-802, 1983
[326] . V. Vannicola, P. Varshney, "Spectral Dispersion of Modulated Signals due to Oscillator Phase Instability: White and Random Walk Phase Model", IEEE Trans. Communications, **31**(7), 886-895, 1983
[327] . F. Kartner, "Analysis of White and 1/f noise in Oscillators", International Journal of Circuit Theory and Applications, **18**, 485-519, 1990
[328] . T. Weigandt, B. Kim, P. Gray, "Analysis of Timing Jitter in CMOS Ring Oscillators", IEEE ISCAS, **4**, 27-30, 1994
[329] . J. McNeill, "Jitter in Ring Oscillators", IEEE ISCAS, **6**, 201-204, 1994
[330] . J. McNeill, "Jitter in Ring Oscillators", IEEE J. Solid-State Circuits, **32**(6), 870-879, 1997
[331] . B. Razavi, "A Study of Phase Noise in CMOS Oscillators", IEEE J. Solid-State Circuits, **31**(3), 331-343, 1996
[332] . K. Takagi, S. Serikawa, T. Kurita, "Phase Noise and a Low-Frequency Noise Reduction Method in Bipolar Transistors", IEEE Trans. Electron Devices, **44**(7), 1180-1181 , 1997
[333] . K. Takagi, "Correlation Between Low Frequency Phase and Amplitude Fluctuations in Transmission Line Model of a Bipolar Transistor and the Noise Reduction Method", Solid-State Electronics, **49**(8), 1449-1451, 2005
[334] . F. Herzel, "An Analytical Model for the Power Spectral Density of a Voltage-Controlled Oscillator and its Analogy to the Laser Linewidth Theory", IEEE Trans. Circuits Systems, **45**(9), 904-908, 1998
[335] . G. Coram, "A Simple 2-D Oscillator to Determine the Correct Decomposition of Perturbations into Amplitude and Phase Noise", IEEE Trans. Circuits Systems, **48**(7), 896-898, 2001
[336] . P. Vanassche, G. Gielen, W. Sansen, "On the Difference Between Two Widely Publicized Methods for Analyzing Oscillator Phase Behavior", IEEE-ICCAD (Int. Conf. Computer-Aided Design), Digest Tech. Papers, 229-233, 2002
[337] . P. Vanassche, G. Gielen, W. Sansen, "A Generalized Method for Computing Oscillator Phase Noise Spectra", IEEE-ICCAD (Internat. Conf. Computer-Aided Design), Digest Tech. Papers, 247-250, 2003
[338] . Y. Zhao, Z. G. Wang, “20-GHz Differential Colpitts VCO in 0.35-upmum BiCMOS”, Journal of Infrared, Millimeter, and Terahertz Waves, 30, 250-258, 2009
[339] . W. Liu, L. Huang, “A low power CMOS VCO using inductive-biasing with high performance FoM”, Journal of Semiconductors, 37, 045001, 2016
[340] . K. Siddiq, M. K. Hobden, S. R. Pennock, R. J. Watson, “Phase Noise in FMCW Radar Systems”, IEEE Transactions on Aerospace and Electronic Systems, 55, 70-81, 2019
[341] . A Fakhr, M. J. Deen, H. deBruin, "Low-Voltage, Low-Power and Low Phase Noise 2.4 GHz VCO for Medical Wireless Telemetry", Canadian Conf. Electrical and Computer Engineering, **CCECE 2004(3)**, 1321-1324, 2004

[342] . S. Kim, T. Rim, K. Kim, U. Lee, E. Baek, H. Lee, C.-K. Baek, M. Meyyappan, M. J. Deen, J.-S. Lee, "Silicon Nanowire Ion Sensitive Field Effect Transistor with Integrated Ag/AgCl Electrode: pH Sensing and Noise Characteristics", Royal Soc. Chem. Analyst, **136**(23), 5012-5016, 2011.
[343] . M. J. Deen, M. Shinwari, J. Ranuarez, "Noise Considerations in Field-Effect Biosensors", J. Appl. Phys., **100**(7), 074703, 2006.
[344] . R. McIntyre, "Multiplication Noise in Uniform Avalanche Diodes", IEEE Trans. Electron Devices, **13**(1), 164-168, 1966
[345] . R. McIntyre, "The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory", IEEE Trans. Electron Devices, **19**(6), 703-713, 1972
[346] . X. Zhao, M. J. Deen, L. Tarof, "Low Frequency Noise in Separate Absorption, Grading, Charge and Multiplication (SAGCM) Avalanche Photodiodes", Electronics Letters, **32**(3), 250-252, 1996
[347] . S. An, M. J. Deen, "Low-Frequency Noise in Single Growth Planar Separate Absorption, Grading, Charge, and Multiplication Avalanche Photodiodes", IEEE Trans. Electron Devices, **47**(3), 537-543, 2000
[348] . S. An, M. J. Deen, A. Vetter, W. Clark, J.-P. Noel, F. Shepherd, "Effect of Mesa Overgrowth on Low-Frequency Noise in Planar Separate Absorption, Grading, Charge, and Multiplication Avalanche Photodiodes", IEEE J. Quantum Electronics, **35**(8), 1196-1202, 1999
[349] . N. Faramarzpour, M. J. Deen, S. Shirani, "An Approach to Improve the Signal-to-Noise Ratio of Active Pixel Sensor for Low-Light-Level Applications", IEEE Trans. Electron Devices, **53**(9), 2384-2391, 2006.
[350] . H. Wey, W. Guggenbuhl, "Noise Transier Characteristics of a Correlated Double Sampling Circuit", IEEE Trans. Circuits and Systems, **33**(10), 1028-1030, 1986.
[351] . X. Wang, P. Rao, A. Mierop, A. Theuwissen, "Random Telegraph Signal in CMOS Image Sensor Pixels", IEDM 2006, Technical Digest, 15-118, 2006
[352] . M. J. Deen, S. Majumder, O. Marinov, M. El-Desouki, "Random Telegraph Signal Noise in CMOS Active Pixel Sensors", 21st International Conference on Noise and Fluctuations (ICNF), 208-211, 2011
[353] . A. van der Wel, E. Klumperink, J. Kolhatkar, E. Hoekstra, M. Snoeij, C. Salm, H. Wallinga, B. Nauta, "Low-Frequency Noise Phenomena in Switched MOSFETs", IEEE J. Solid-State Circuits, **42**(3), 540-550, 2007
[354] . S. Kawahito, N. Kawai, "Column Parallel Signal Processing Techniques for Reducing Thermal and RTS Noises in CMOS Image Sensors", Int. Image Sensor Workshop, **IISW 2007**, 226-229, 2007
[355] . W. Schottky, "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern", Annalen der Physik, **362**(23), 541-567, 1918
[356] . W. Schottky, "Small-Shot Effect and Flicker Effect", Phys. Rev., **28**(1), 74-103, 1926
[357] . M. Surdin, "Fluctuations de Courant Thermionique et le 'flicker effect'", J. de Physique et le Radium, Ser. 7, **10**(4), 188-189, 1939
[358] . F. Du Pre, "A Suggestion Regarding the Spectral Density of Flicker Noise", Phys. Rev., **78**(5), 615-615, 1950
[359] . R. Jones, "Noise in Radiation Detectors", Proc. IRE, **47**(9), 1481-1486, 1959
[360] . N. Mantena, R. Lucas, "Experimental Study of Flicker Noise in M.I.S. Field-Effect Transistors", Electronics Letters, **5**(24), 607-608, 1969
[361] . S. Hsu, "Surface State Related 1/f Noise in MOS Transistors", Solid-State Electronics, **13**(11), 1451-1459, 1970
[362] . A. Boornard, E. Herrmann, S. Hsu, "Low-Noise Integrated Silicon-Gate FET Amplifier", IEEE J. Solid-State Circuits, **10**(6), 542-544, 1975
[363] . M. Das, "FET Noise Sources and Their Effects on Amplifier Performance at Low Frequencies", IEEE Trans. Electron Devices, **19**(3), 338-348, 1972
[364] . J. Johnson, "The Schottky Effect in Low Frequency Circuits", Phys. Rev., **26**(1), 71-85, 1925
[365] . C. Mueller, J. Pankove, "A P-N-P Triode Alloy-Junction Transistor for Radio-Frequency Amplification", Proc. IRE, **42**(2), 386-391, 1954
[366] . P. Bargellini, M. Herscher, "Investigations of Noise in Audio Frequency Amplifiers Using Junction Transistors", Proc. IRE, **43**(2), 217-226, 1955
[367] . J. Gibbons, "Low-Frequency Noise Figure and its Application to the Measurement of Certain Transistor Parameters", IEEE Trans. Electron Devices, **9**(3), 308-315, 1962
[368] . M. Tutt, D. Pavlidis, H.-F. Chau, "1/f Noise Characteristics of InP/InGaAs Heterojunction Bipolar Transistors", Proc. Int. Conf. on Indium Phosphide and Related Materials, 364-367, 1992
[369] . L. Vempati, J. Cressler, J. Babcock, R. Jaeger, D. Harame, "Low-Frequency Noise in UHV/CVD Epitaxial Si and SiGe Bipolar Transistors", IEEE J. Solid-State Circuits, **31**(10), 1458-1467, 1996
[370] . T.-P. Lee, C. Burrus, "Noise in the Detected Output of Small-Area Light-Emitting Diodes", IEEE J. Quantum Electronics, **8**(3), 370-373, 1972

[371] . E. Seevinck, F. List, J. Lohstroh, "Static-Noise Margin Analysis of MOS SRAM Cells", IEEE J. Solid-State Circuits, **22**(5), 748-754, 1987
[372] . K. Samsudin, B. Cheng, A. Brown, S. Roy, A. Asenov, "Integrating Intrinsic Parameter Fluctuation Description into BSIMSOI to Forecast sub-15 nm UTB SOI Based 6T SRAM Operation", Solid-State Electronics, **50**(1), 86-93, 2006
[373] . C. E. Calosso, A. C. C. Olaya, E. Rubiola, "Phase-Noise and Amplitude-Noise Measurement of DACs and DDSs", IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, N, 431-439, 2020
[374] . B. Sánta, Z. Balogh, A. Gubicza, L. Pósa, D. Krisztián, G. Mihály, M. Csontos, A. Halbritter, "Universal 1/f type current noise of Ag filaments in redox-based memristive nanojunctions", Nanoscale, 11, 4719-4725, 2019
[375] . G. H. Shin, C. K. Kim, G. S. Bang, J. Y. Kim, B. C. Jang, B. J. Koo, M. H. Woo, Y. K. Choi, S.Y. Choi, "Multilevel resistive switching nonvolatile memory based on MoS_2 nanosheet-embedded graphene oxide", 2D Materials, **3**, 034002, 2016
[376] . N.Mahmud; A. J . Narasimham, J. R. Lloyd" 1/f Noise Analysis of Hafnium Oxide based ReRAM devices using ac plus dc measurement technique", 2016 IEEE International Integrated Reliability Workshop (IIRW), 77-79, 2016
[377] . S. Ambrogio, S. Balatti, V. McCaffrey, D. C. Wang, D. Ielmini, "Noise-Induced Resistance Broadening in Resistive Switching MemorytextemdashPart I: Intrinsic Cell Behavior", IEEE Trans. Electron Devices, **62**, 3805-3811, 2015
[378] . R. Brederlow, W. Weber, D. Schmitt-Landsiedel, R. Thewes, "Fluctuations of the Low Frequency Noise of MOS Transistors and their Modeling in Analog and RF-Circuits", IEDM 1999, Technical Digest, 159-162, 1999
[379] . K. (C.) van Vliet, "Noise in Semiconductors and Photoconductors", Proc. IRE, **46**(6), 1004-1018, 1958
[380] . A. Ziel, A. Becking, "Theory of Junction Diode and Junction Transistor Noise", Proc. IRE, **46**(3), 589-594, 1958
[381] . P. Handel, "Quantum Approach to 1/f Noise", Phys. Rev. A, **22**(2), 745-757, 1980
[382] . A. Ziel, "Unified Presentation of 1/f Noise in Electron Devices: Fundamental 1/f Noise Sources", Proc. of IEEE, **76**(3), 233-258, 1988
[383] . C. Pfeiffer, "Semiconductor 1/f Noise from Dynamic Coupling of Charge Carriers and Lattice", J. Appl. Phys., **90**(7), 3653-3655, 2001
[384] . F. Sthal, M. Devel, S. Ghosh, J. Imbaud, G. Cibiel, R. Bourquin, "Volume dependence in Handel's model of quartz crystal resonator noise", IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, **60**, 1971-1977, 2013
[385] . R. de Sousa, "Dangling-bond spin relaxation and magnetic 1/*f* noise from the amorphous-semiconductor/oxide interface: Theory", Physical Review B, **76**, *24,* 2007
[386] . J. Brews, W. Fichtner, E. Nicollian, S. Sze, "Generalized Guide for MOSFET Miniaturization", IEEE Electron Device Letters, **1**(1), 2-4, 1980
[387] . S. Springer, S. Lee, N. Lu, E. Nowak, J.-O. Plouchart, J. Watts, R. Williams, N. Zamdmer, "Modeling of Variation in Submicrometer CMOS ULSI Technologies", IEEE Trans. Electron Devices, **53**(9), 2168-2178, 2006
[388] . C. Alexander, A. Brown, J. Watling, A. Asenov, "Impact of Single Charge Trapping in Nano-MOSFETs - Electrostatics versus Transport Effects", IEEE Trans. Nanotechnology, **4**(3), 339-344, 2005
[389] . C. Alexander, A. Brown, J. Watling, A. Asenov, "Impact of Scattering in 'Atomistic' Device Simulations", Solid-State Electronics, **49**(5), 733-739, 2005
[390] . C. Chen, Q. Huang, J. Zhu, Y. Zhao, L. Guo, R. Huang, "New Understanding of Random Telegraph Noise Amplitude in Tunnel FETs", IEEE Transactions on Electron Devices, **64**, 3324-3330, 2017
[391] . K. Maekawa, H. Makiyama, Y. Yamamoto, T. Hasegawa, S. Okanishi, K. Sonoda, H. Shinkawata, T. Yamashita, S. Kamohara, Y. Yamaguchi, "Comprehensive analysis of low-frequency noise variability components in bulk and fully depleted silicon-on-insulator metal-oxide-semiconductor field-effect transistor", Japanese Journal of Applied Physics, **57**, 04FD19, 2018
[392] . M. Toita, L. Vandamme, S. Sugawa, A. Teramoto, T. Ohmi, "Geometry and Bias Dependence of Low-Frequency Random Telegraph Signal and 1/F Noise Levels in MOSFETs", Fluctuation and Noise Lett., **5**(4), L539-L548, 2005
[393] . H. Davis, R. Jones, "Estimation of the Innovation Variance of a Stationary Time Series", J. American Statistical Assoc., **63**(321), 141-149, 1968
[394] . T. Pukkila, H. Nyquist, "On the Frequency Domain Estimation of the Innovation Variance of a Stationary Univariate Time Series", Biometrika, **72**(2), 317-323, 1985
[395] . A. Walden, "Multitaper Estimation of the Innovation Variance of a Stationary Time Series", IEEE Trans. Signal Processing, **43**(1), 181-187, 1995
[396] . N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications", First published during the II World War as a classified report to Section D 2, National Defense Research Committee; several editions after, e.g. MIT Press, Cambridge, Mass., 1970

[397] . L. Froeb, "Log Spectral Analysis: Components of Asset Prices", Chapter 13 in "Computational Economics and Finance: Modeling and Analysis with Mathematica", ed. H. Varian, Springer, 305 ff., 1996
[398] . W. Kinsner, "Characterizing Chaos Through Lyapunov Metrics", IEEE Trans. Sys. Man Cybernetics, **36**(2), 141-151, 2006
[399] . Y. Leng, T. Wang, Y. Guo, Y. Xu, S. Fan, "Engineering Signal Processing Based on Bistable Stochastic Resonance", Mech. Sys. Signal Process., **21**, 138-150, 2007
[400] . Y. Kutovyi, I. Madrid, I. Zadorozhnyi, N. Boichuk, S. H. Kim, T. Fujii, L. Jalabert, A. Offenhaeusser, S. Vitusevich, N. Clément, "Noise suppression beyond the thermal limit with nanotransistor biosensors", Scientific Reports, **10**, 2020
[401] . V. P. Koverda, V. N. Skokov, "Dynamical Chaos in a Nonlinear System with 1/f Spectrum", Technical Physics Letters, **45**, 1159-1162, 2019
[402] . A. Setiadi, H. Fujii, S. Kasai, K. i. Yamashita, T. Ogawa, T. Ikuta, Y. Kanai, K. Matsumoto, Y. Kuwahara, M. Akai-Kasaya, "Room-temperature discrete-charge-fluctuation dynamics of a single molecule adsorbed on a carbon nanotube", Nanoscale, **9**, 10674-10683, 2017
[403] . K. Chen, N. H. Chan, C. Y. Yau, "Bartlett correction of frequency domain empirical likelihood for time series with unknown innovation variance", Annals of the Institute of Statistical Mathematics, **72**, 1159-1173, 2019
[404] . Z. Rongyun, G. Changfu, S. Peicheng, Z. Linfeng, Z. Changsheng, W. Chen, "The Permanent Magnet Synchronous Motor Sensorless Control of Electric Power Steering Based on Iterative Fifth-Order Cubature Kalman Filter", Journal of Dynamic Systems, Measurement, and Control, **142**, 2020
[405] . J. Cooley, J. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series", Mathematics of Computation, **19**(90), 297-301, 1965
[406] . M. Mihaila, "Low-Frequency Noise in Nanomaterials and Nanostructures", in "Noise and Fluctuations Control in Electronic Devices", ed. A. Balandin, American Scientific Publishers, 367-385, 2002
[407] . R. Feynman, "There's Plenty of Room at the Bottom", Transcript of a talk given in 1959 from J. Microelectromech. Sys., **1**(1), 60-66, 1992

Table 1. Main types of noise in electronic devices, in terms of PSD of noise current. The noise in real devices is combination of these types, and the low-frequency noise is with 1/f and/or Lorentzian spectra

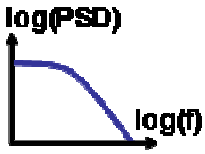
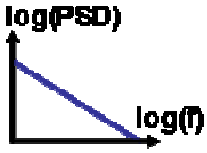
Type	Cause	Power Spectral Density	
Thermal (Johnson, Nyquist)	Thermal motion of carriers	$4kT/R$ (white)	
Shot	Discrete carriers crossing a barrier	$2qI$ (white)	
Generation-recombination and RTS	Trapping and detrapping of carriers	$\propto \frac{\tau}{1+(2\pi f\tau)^2}$ (Lorentzian)	
Flicker (1/f, excess)	Several possible origins	$\propto 1/f$ (pink)	

Table 2. Parameters related to tunneling attenuation distance λ in insulators. $\lambda = h/\left(4\pi\sqrt{2m_o m^* \Phi}\right)$, according to Wentzel-Kramer-Brillouin (WKB) approximation, where $h=6.63\times10^{-34}$ Js is Planck constant, $m_o=9.11\times10^{-34}$ kg is electron mass, $m^*=m/m_o$ is effective mass and Φ is the energy offset (barrier) between conduction bands for electrons and valence bands for holes. The band gap for Si is taken from 1.1eV to 1.12 eV. Some discrepancies exist between the values reported in the literature.

Interface	Content proportion	Barrier Φ		Insulator band gap	Effective mass m*		Attenuation distance λ		Reference
		electrons	holes		electrons	holes	electrons	holes	
	%	eV	eV	eV	m_e/m_o	m_h/m_o	nm	nm	
SiO_2-Si		3.50	4.50	9.00	0.36	0.50	0.087	0.065	[45]
SiO_2-Si		3.05	4.46	8.65	0.38	0.51	0.091	0.065	[45]
$SiON_x$-Si	x=1.8%	2.99	4.28	8.42	0.40	0.52	0.089	0.066	[45]
$SiON_x$-Si	x=3.0%	2.94	4.15	8.23	0.41	0.53	0.089	0.066	[45]
$SiON_x$-Si	x=4.0%	2.87	4.03	8.06	0.43	0.56	0.088	0.065	[45]
$SiON_x$-Si	x=4.5%	2.80	3.81	7.74	0.48	0.58	0.085	0.066	[45]
SiO_2-Si		3.25	4.63	9.00					[146]
SiO_2-Si		3.50	4.40	9.02					[44]
Si_3N_4-Si		2.40	1.80	5.32					[44]
TaO_5-Si		0.30	3.00	4.42					[44]
$SrTiO_3$-Si		-0.1	2.30	3.32					[44]
$BaZrO_3$-Si		0.80	3.40	5.32					[44]
ZrO_2-Si		1.40	3.30	5.82					[44]
HfO_2-Si		1.50	3.40	6.02					[44]
Al_2O_3-Si		2.80	4.90	8.82					[44]
Y_2O_3-Si		2.30	2.60	6.02					[44]
La_2O_3-Si		2.30	2.60	6.02					[44]
$ZrSiO_3$-Si		1.50	3.40	6.02					[44]
$HfSiO_4$-Si		1.50	3.40	6.02					[44]
$LaAlO_3$-Si		1.90	3.20	6.22					[44]
HfO_2-Si		1.13			0.18		0.217		[50]
HfO_2-Si		1.13			0.18		0.210		[47]
Al_2O_3-Si		2.80			0.28		0.110		[47]
$HfAlO_x$-Si							0.145		[47]
SiO_2-Si							0.100		[47]

Table 3. Device parameters and legend

(Note: Table 3 is embedded in
Figure 31)

Table 4. Typical assumptions for 1/f noise in Analog, RF and Digital applications of MOS transistors

(Note: Table 4 is embedded in Figure 68)

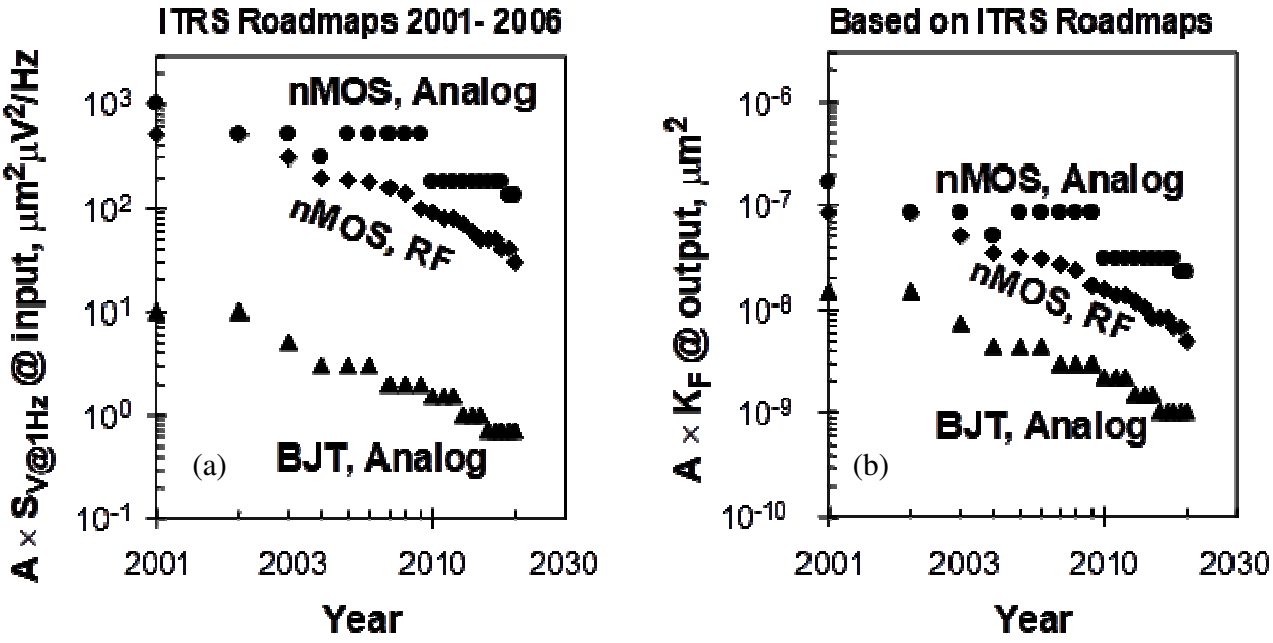


Figure 1. Level of 1/f noise in silicon transistors: a) prediction in ITRS [3] for the input referred voltage noise in npn BJT (▲), RF nMOS (◆) and Analog nMOS (●) transistors; b) corresponding values for the output current noise in terms of the simple SPICE parameter $K_F=f\times S_I/I_{DC}^2$.

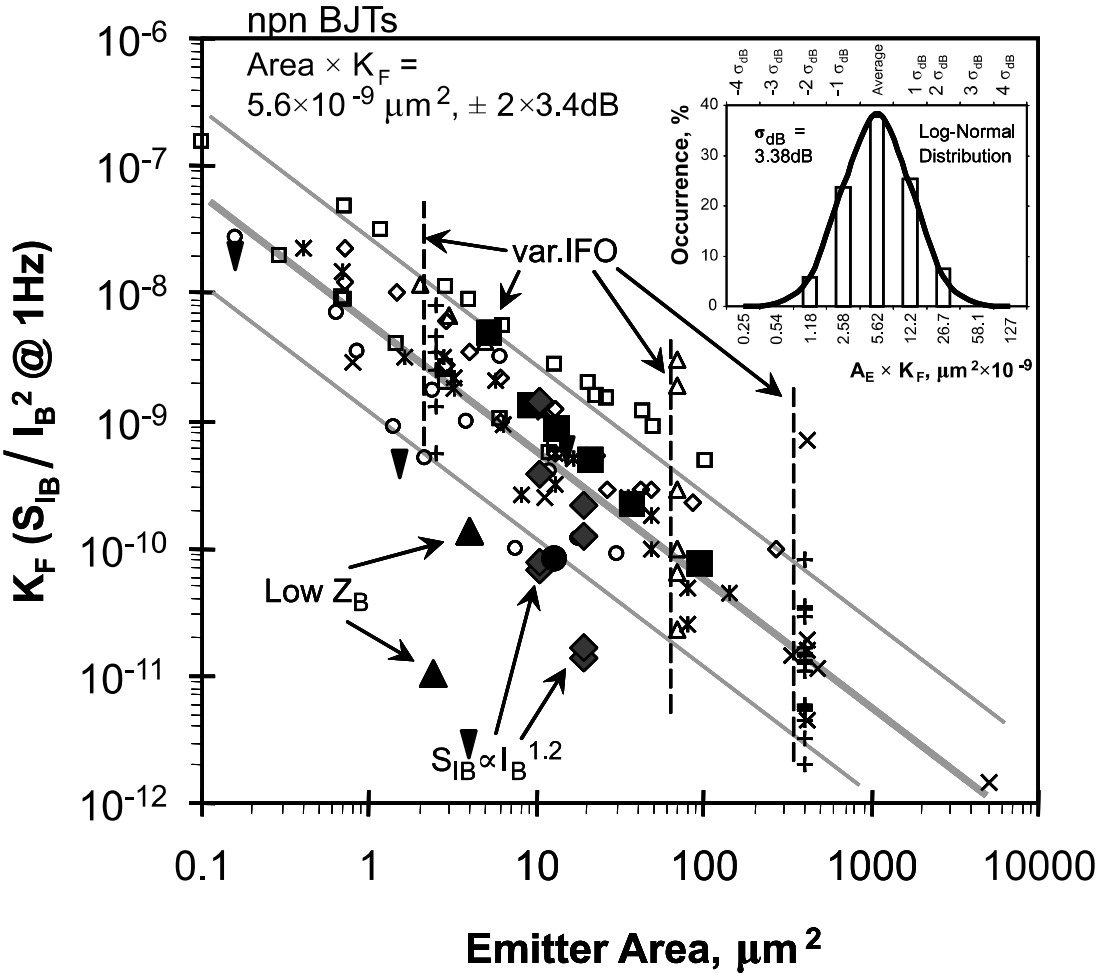
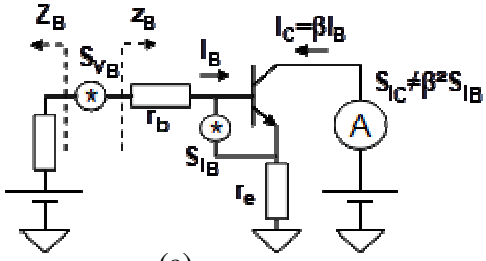


Figure 2. Evolution of normalized noise in silicon npn BJTs with polysilicon emitter. Crosses and asterisks are from earlier publications before year 2000, (× [10], ✳ [11, 12, 13], + [14]). Open symbols are for the period from 2000 to 2004 (□ [15], ○ [16, 17], ◇ [18], △ [19]). Solid symbols are from publications in the period 2005-2020 (■ [20], ◆ [21], ● [22], ▲ [23], ▼ [24, 25, 26, 27]). Geometric averaging of the product $A_E \times K_F$ is used to evaluate the trend in the dependence of K_F on the emitter area A_E (solid gray line) and the variation limits $\pm 2\sigma_{dB}$ (thin grey lines). The insert shows the distribution of $A_E \times K_F$.

$Z_B << z_B \rightarrow S_{V_B}$ dominates $\rightarrow S_{I_C}$= min
$Z_B >> z_B \rightarrow S_{I_B}$ dominates $\rightarrow S_{I_C}$= max
S_{I_C} variation with Z_B: 2 and more decades



(a)

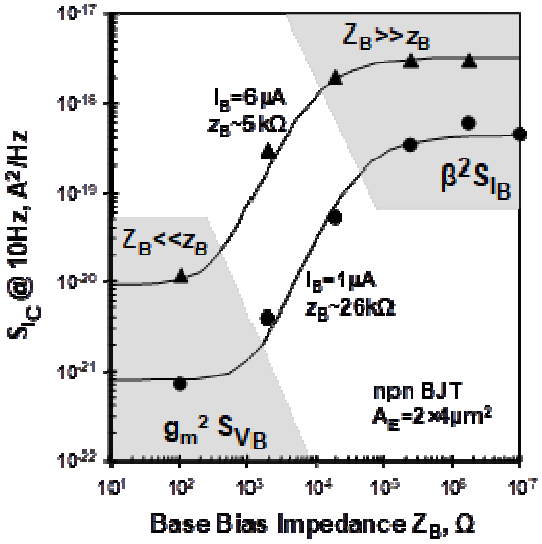


Figure 3. Noise equivalent circuit and evolution of low-frequency noise in BJT with the impedance Z_B of the bias circuit in the base terminal at a constant base DC current [12]

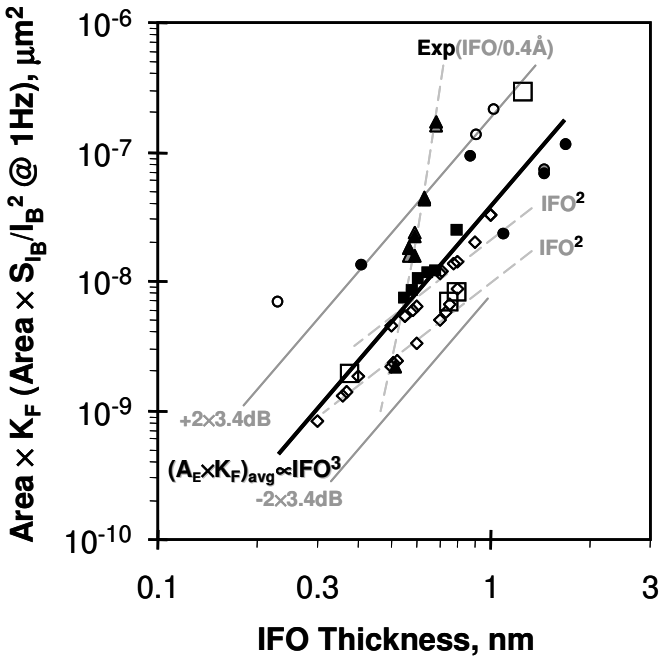


Figure 4. Evolution of the 1/f noise level with the thickness t_{IFO} of the interfacial oxide (IFO) between poly and monosilicon layers in the emitter of BJT. Data for: npn Si (□)[10], (◇) [14], (○) [19], (■) [31]; pnp Si (●)[32]; npn and pnp SiGe BJT (▲) [33, 34].

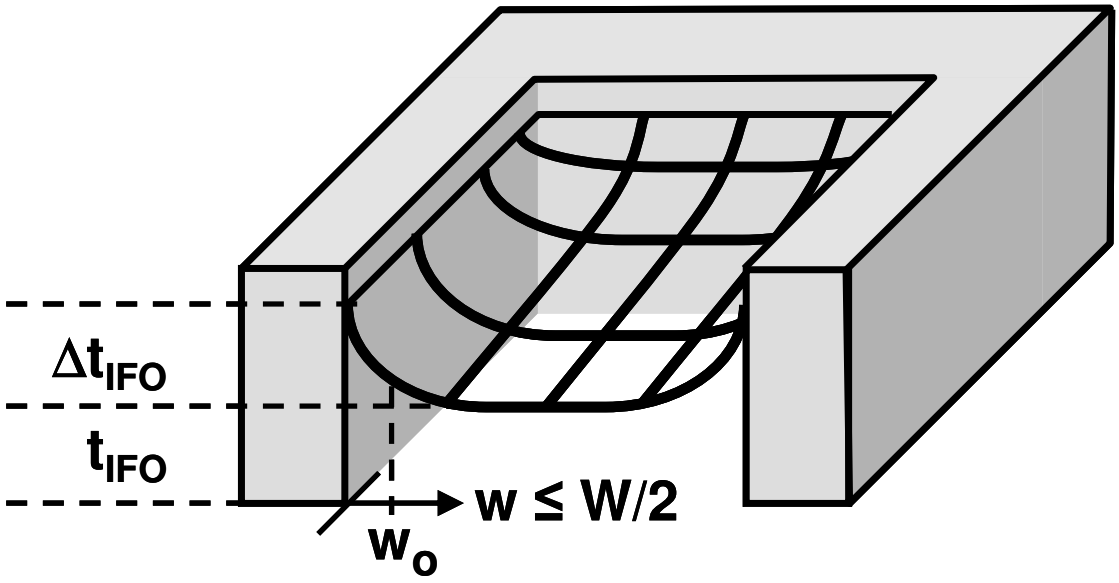


Figure 5. Non-uniform IFO, which is thicker at the periphery of the emitter. The average thickness of IFO increases when the width W of the emitter decreases, since w_o can be assumed constant.

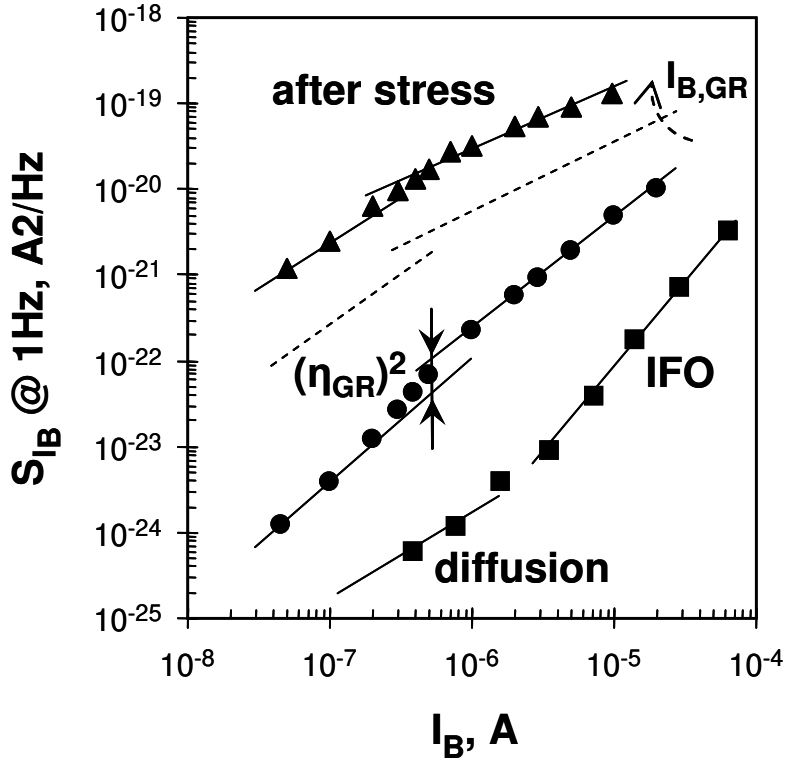


Figure 6. Crossover in the bias dependence of the noise in the base current of BJTs. (■) $A_E=10^4$ μm^2 [10], showing crossover between intrinsic diffusion noise and coupled from IFO noise. (●)$A_E=10$ μm^2 with superficial base doping (SBD) [21], showing crossover between generation-recombination and diffusion base DC current. (▲) Same SBD device after 2 minutes electrical stress, showing off-leveling of noise, caused by high generation-recombination base DC current.

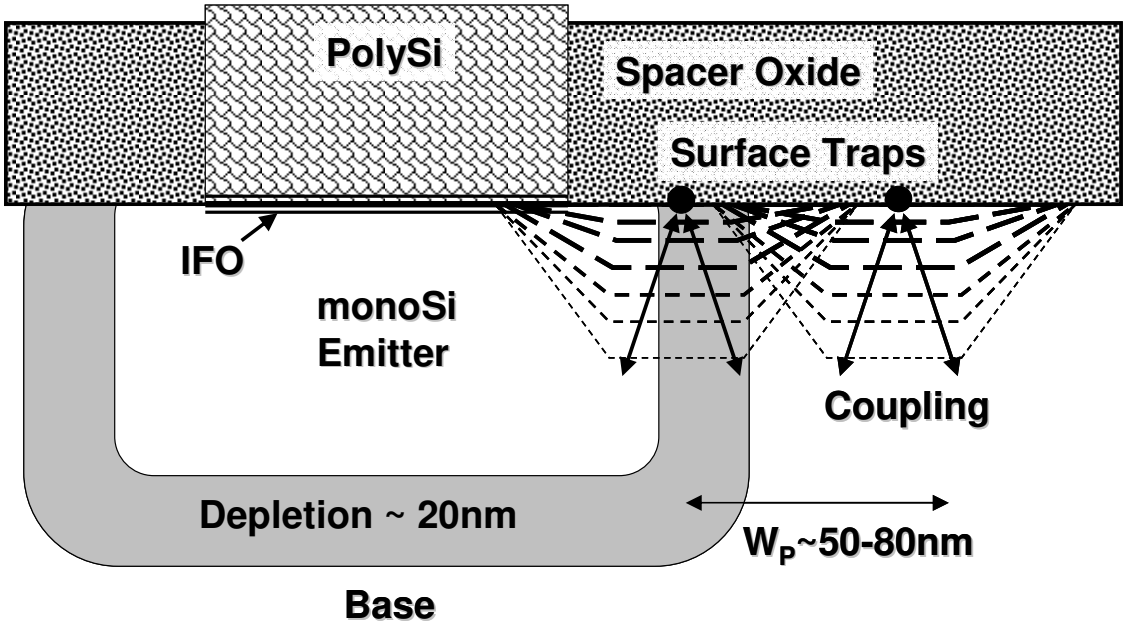


Figure 7. Coupling of surface traps to the base-emitter junction.

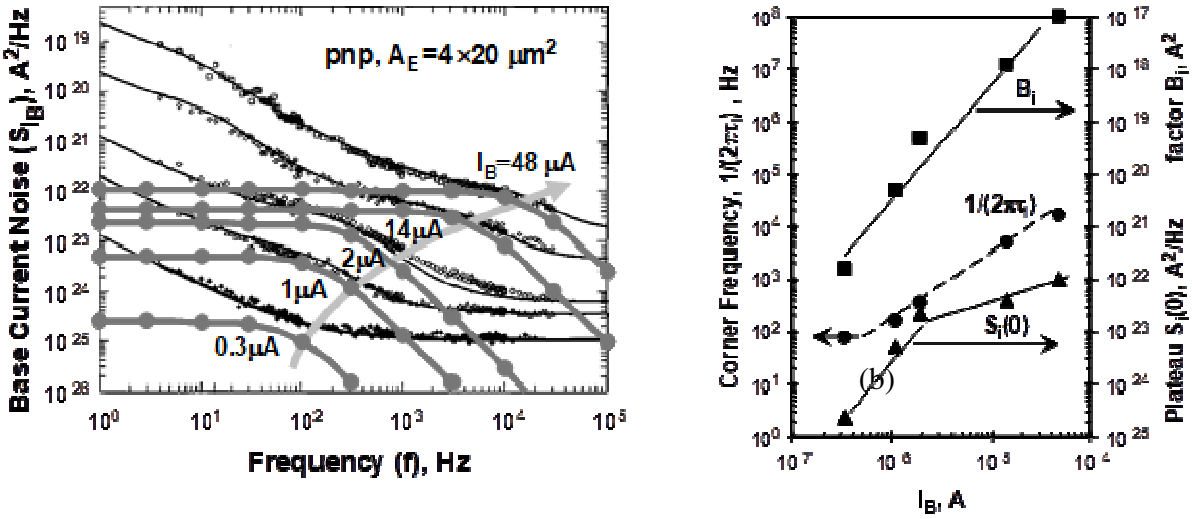


Figure 8. Bias dependence of Lorentzian noise, when coupled from single generation-recombination center in BJT. Data from [12]. a) Evolution of power spectrum density. Solid circles represent the evolution of one Lorentzian component. b) Evolution of parameters of Lorentzian noise.

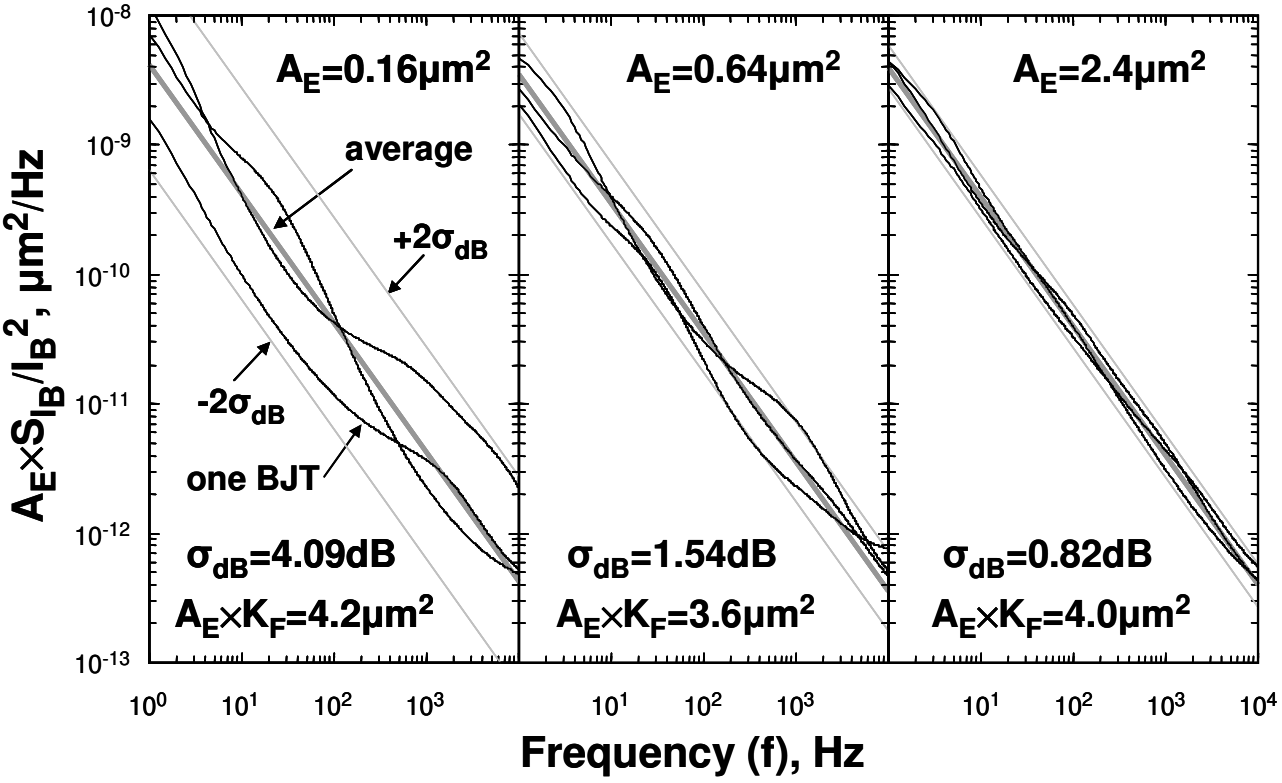


Figure 9. Crossover from Lorentzian noise in BJTs with small emitter area to 1/f noise in BJTs with large emitter area [17]. The measured noise in individual BJT is shown with thin black lines. The average noise among several devices is shown with thick gray line. The range for variation of the noise in individual devices is within $\pm 2\sigma_{dB}$, and it is shown with thin gray lines.

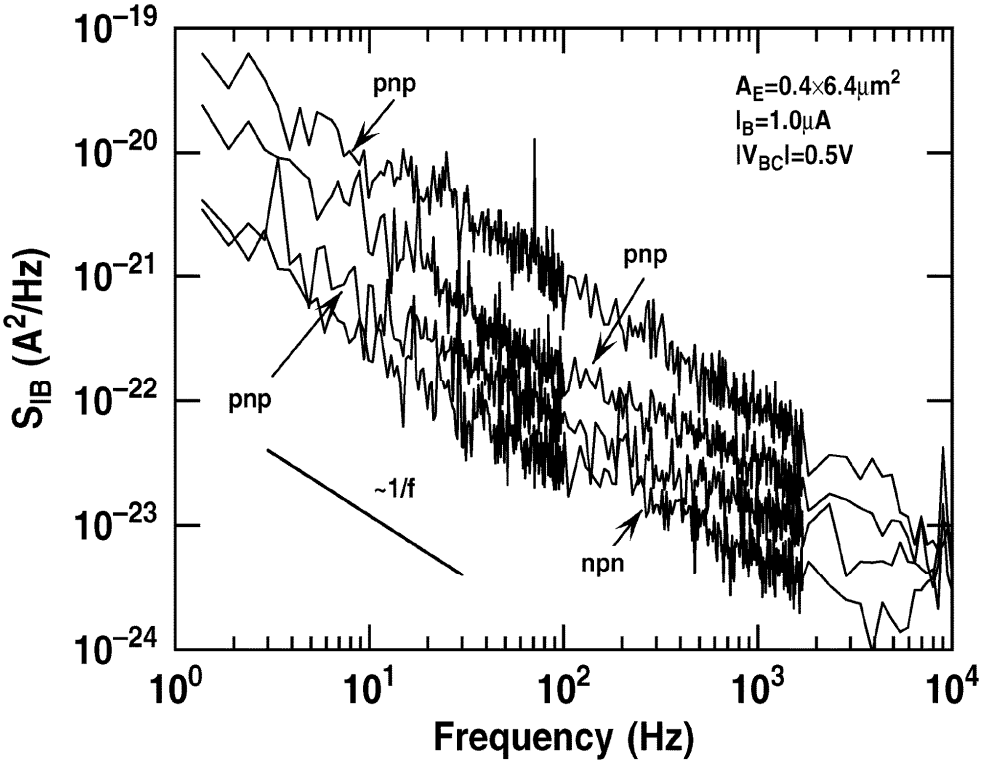


Figure 10. Scattering of measured noise spectrum around its mean, when small numbers (e.g. 10) of captured records are averaged. The figure is from a journal publication, but the reference is omitted to avoid misinterpretation of our intentions.

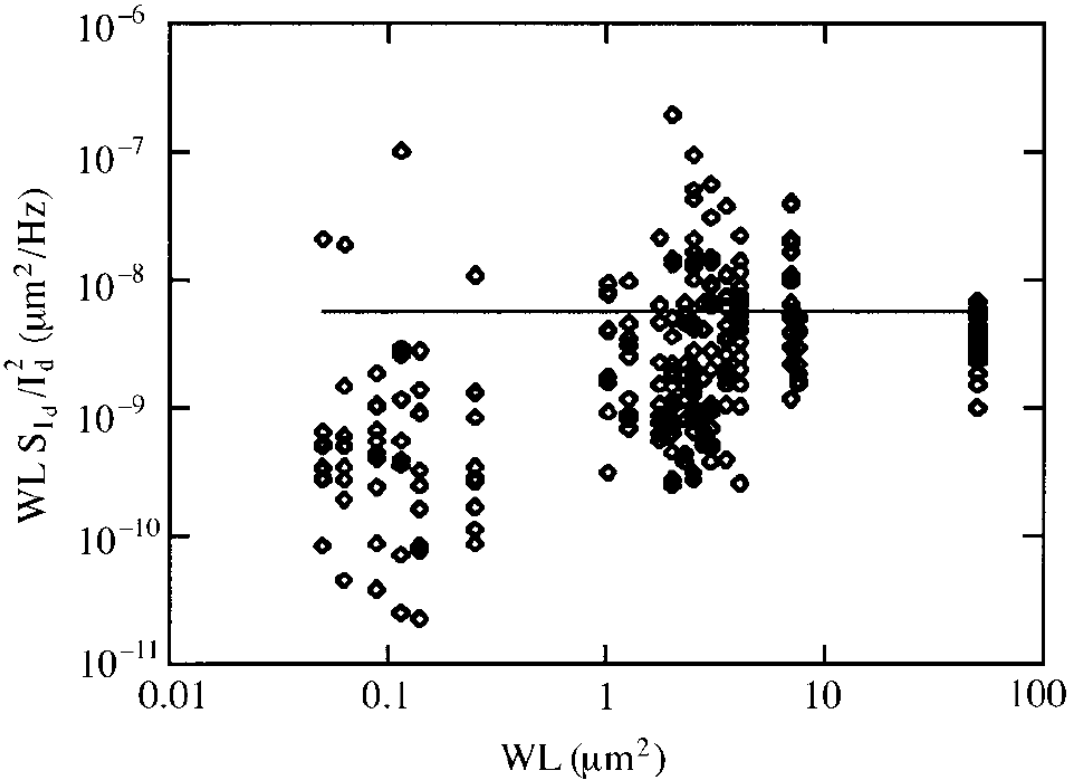


Figure 11. Increase of noise variation in small-area MOS transistors from 0.18μm technology node, as obtained by applying the approach of choosing a frequency of interest from the low-frequency noise spectrum, in this case f=10Hz. Biasing conditions are DC drain currents $I_D \times L/W=0.4\mu A$, DC drain voltage $V_{DS}=50mV$. The figure is from [74].

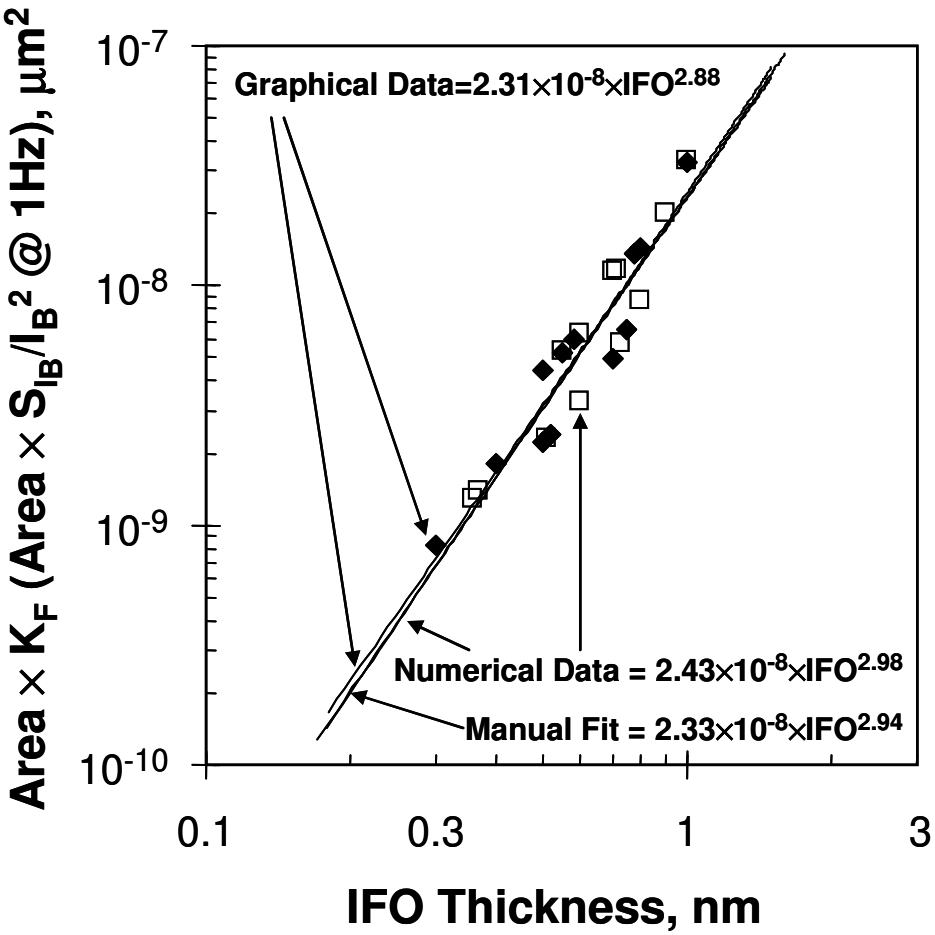


Figure 12. Variation of values in manually processed data from noise measurements. Data are sub-set from Figure 2. Original data from [14]. Numerical (□) and graphical data (◆) are fitted using least mean square method. Manual fit is made directly in the plot with all data, but before they have been fitted by least mean square method.

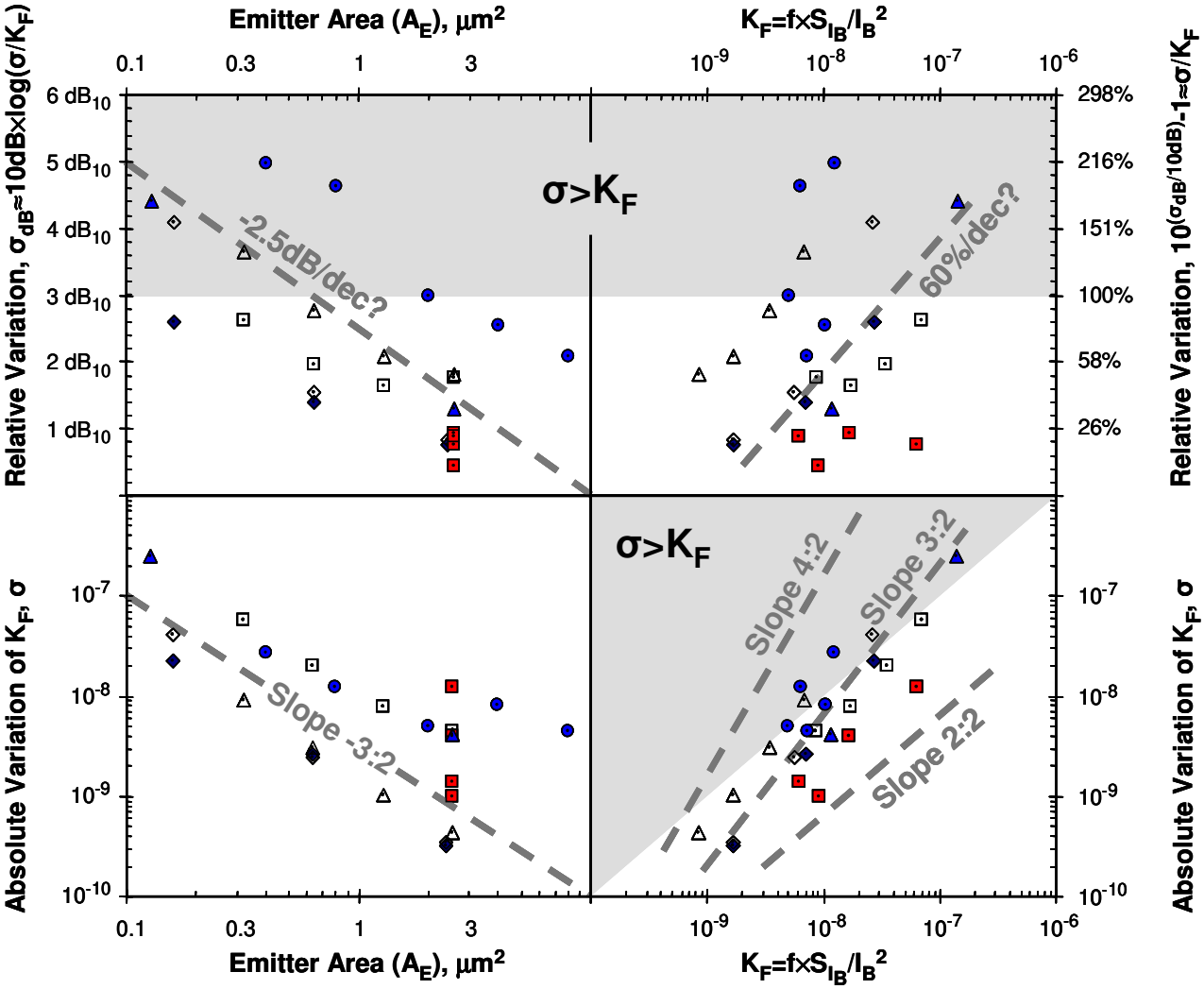


Figure 13. Variation of noise around its average in BJTs. Silicon npn BJTs (f_T=25GHz) from [17, 66, 67] using arithmetic mean (◆) and geometric mean (◇) – see Figure 9 for the later. SiGe npn HBTs with f_T=20GHz (△) from [34], f_T=70GHz (●) from [36], f_T=120GHz (▲) from [81]. SiGe pnp HBTs with different IFO thicknesses (■) from [33] and f_T=20GHz with different emitter areas A_E (□) from [34]. Shaded areas in the figure represent situations when the standard deviation σ of noise is larger than the average level of the noise. The average noise is given by K_F=geometric-mean($f \times S_{I_B}/I_B^2$), calculated according to eq. (83).

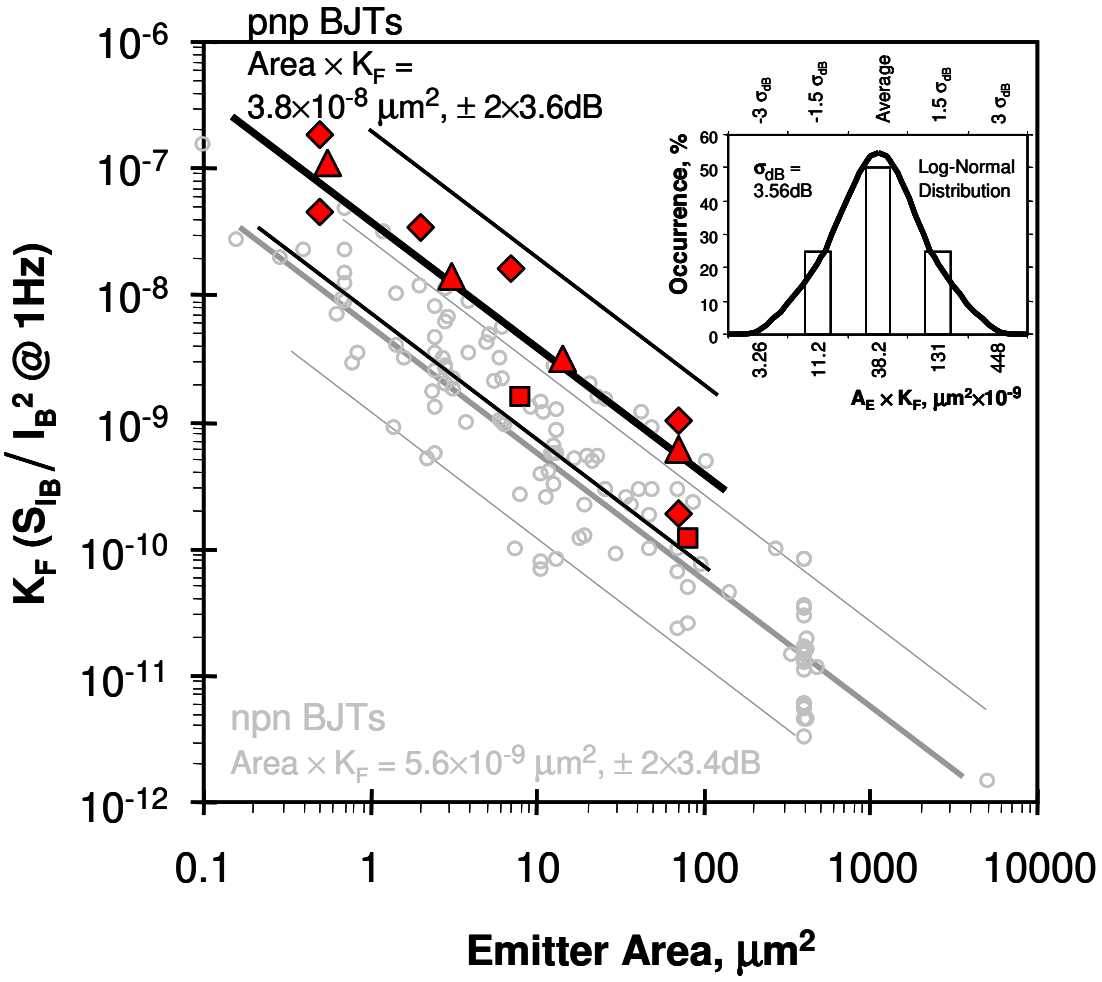


Figure 14. Evolution of normalized noise in silicon pnp BJTs with polysilicon emitter. (■) from [12] for the period before year 2000, (▲) from [83] for the period 2000-2004, (◆) from [32] for the period 2005-2006. Geometric averaging of the product $A_E \times K_F$ is used to evaluate the trend in the dependence of K_F on the emitter area A_E (solid black line) and the variation limits $\pm 2\sigma_{dB}$ (thin black lines). The insert shows the distribution of $A_E \times K_F$ in pnp BJTs. Gray circles (○) and grey lines are from Figure 2 and represent npn BJTs.

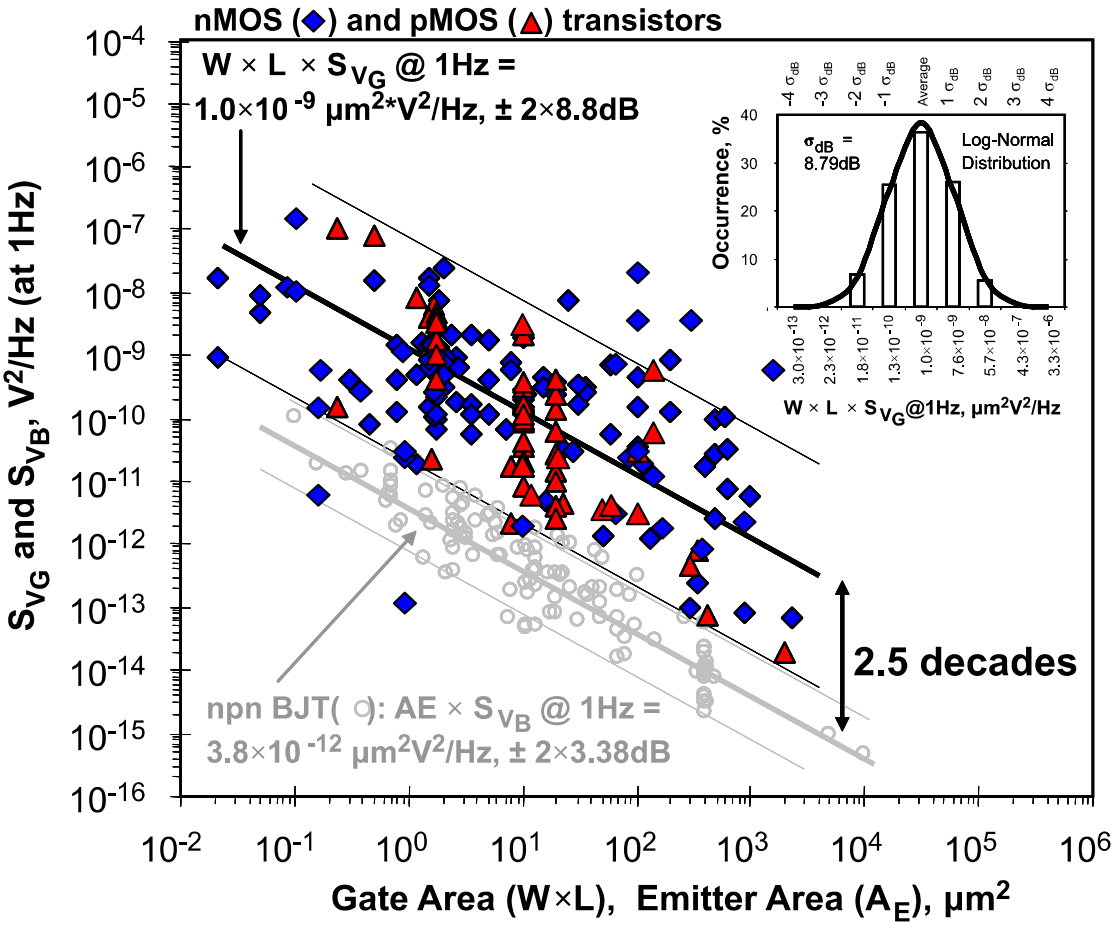


Figure 15. Input-referred voltage noise S_{V_G} (gate voltage 1/f noise at 1Hz) in silicon MOS transistors compared to the corresponding S_{V_B} in silicon npn BJTs. (◆) for nMOS transistors from [22, 47, 48, 49, 50, 51, 72, 82, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123], (▲) for pMOS transistors from [47, 52, 86, 88, 95, 96, 97, 100, 102, 104, 105, 109, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132]. Geometric averaging of the product $W \times L \times S_{V_G}$ is used to evaluate the trend in the dependence of S_{V_G} on the gate area W×L (solid black line) and the variation limits $\pm 2\sigma_{dB}$ (thin black lines). The insert shows the distribution of $W \times L \times S_{V_G}$ in MOS transistors. Gray circles (○) and gray lines represent data from Figure 2 for npn BJTs, using $S_{V_B}@1Hz = K_F \times (\varphi_t)^2/1Hz$, according to eq. (3).

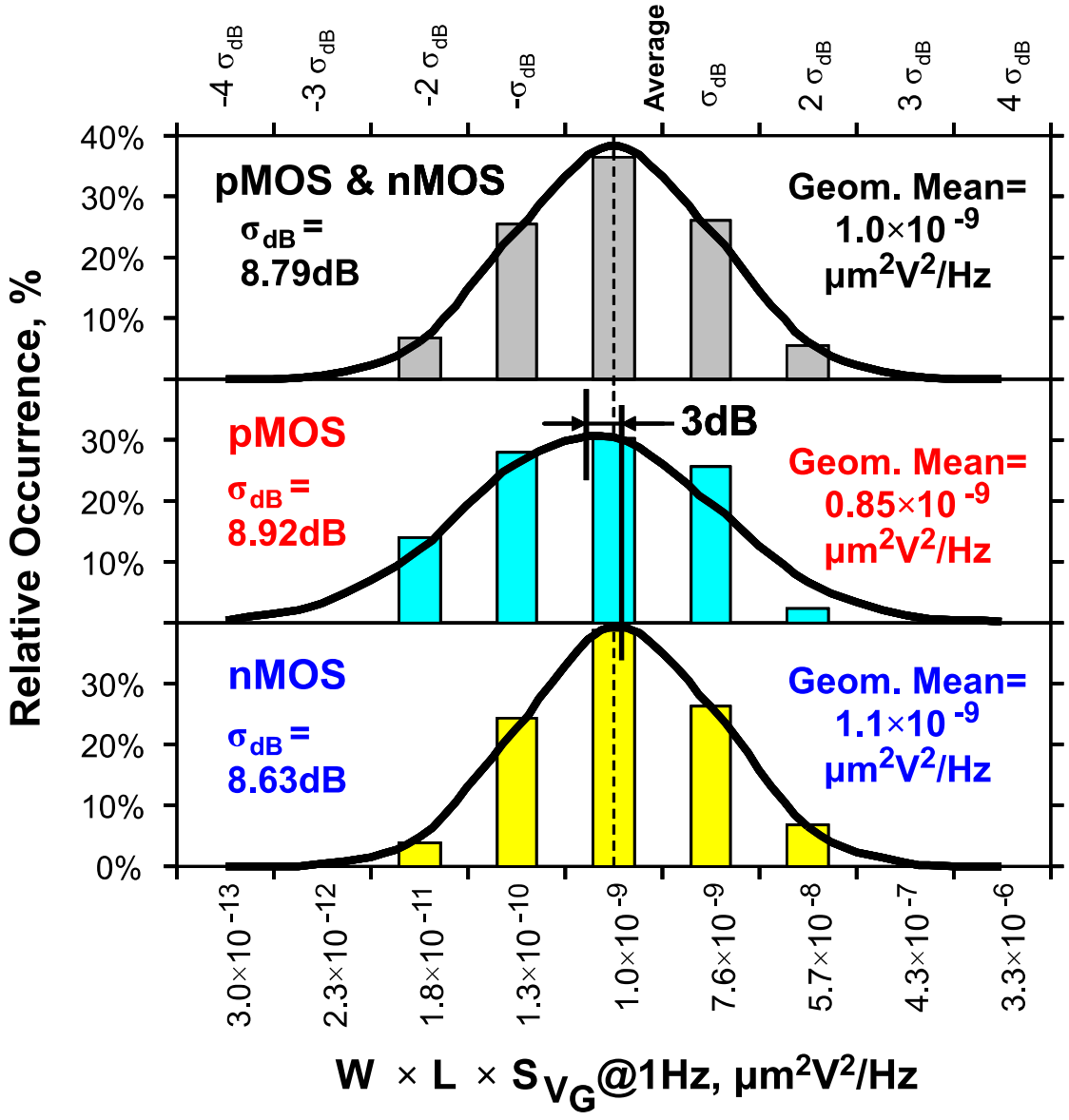


Figure 16. Distributions of the product $W \times L \times S_{V_G}$ for the input-referred voltage noise S_{V_G} (gate voltage 1/f noise at 1Hz) in silicon MOS transistors, as deduced from the data in Figure 15.

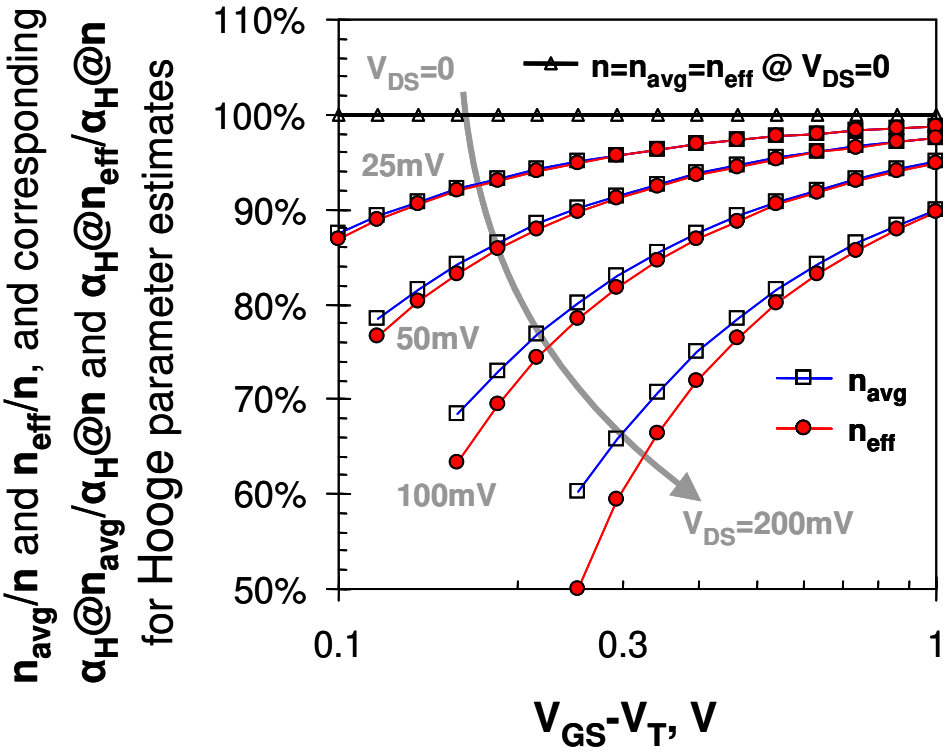


Figure 17. Possible variations in estimating Hooge parameter α_H for the intrinsic (mobility) noise in MOS transistors operating in linear (ohmic) mode ($V_{DS} \le V_{GS}-V_T-50mV$), when different models and approximations for the charge carriers in the channel are used. (△) simplest approximation for the number of charges $n=WLC_{ox}(V_{GS}-V_T)/q$ uniformly distributed from source to drain. (□) approximation with average of number of charges at source and drain sides $n_{avg}=(n_{source}+n_{drain})/2$. (●) Gradual (linear) charge sheet approximation $n_{eff} \le n_{avg} \le n$.

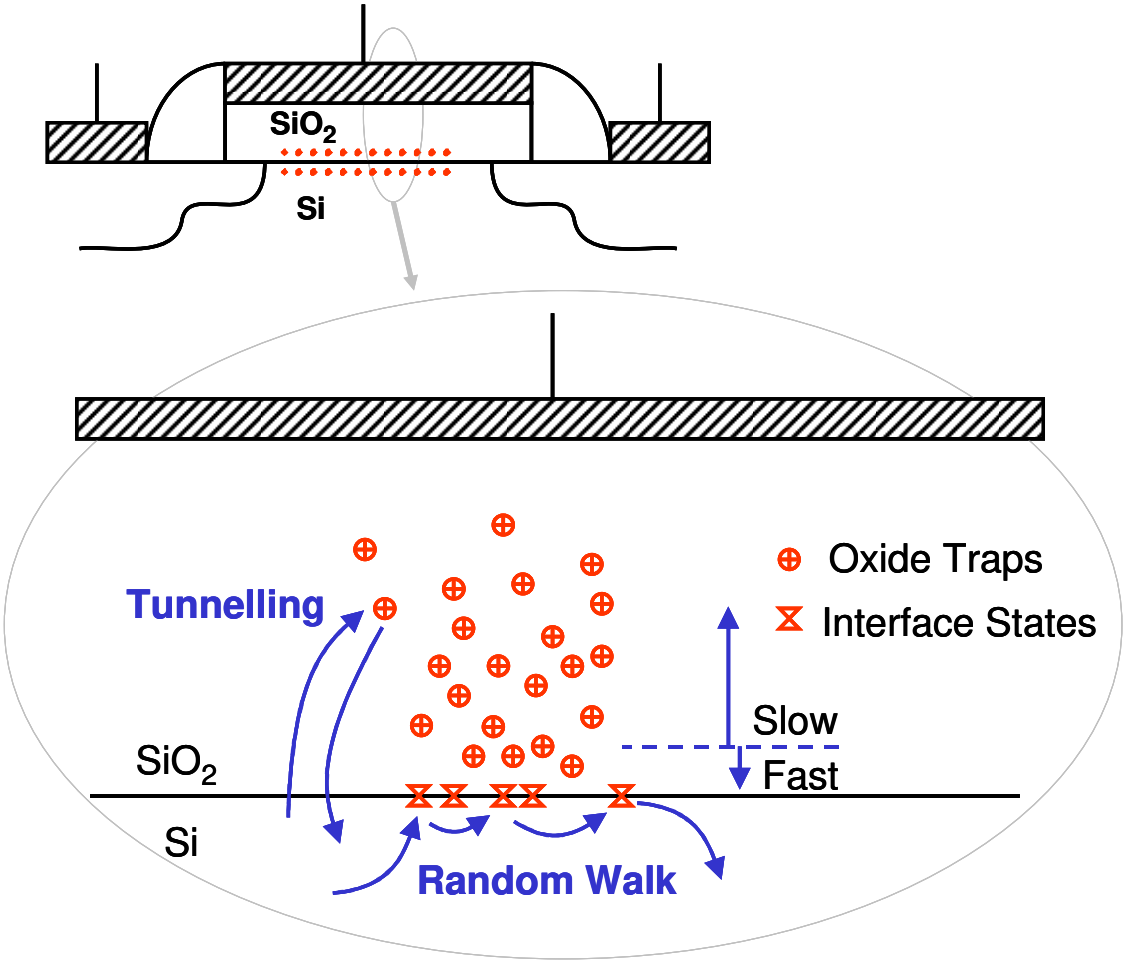


Figure 18. Possible trapping mechanisms, which couple low-frequency noise in MOS transistors and result in gate referred 1/f noise voltage S_{V_G}

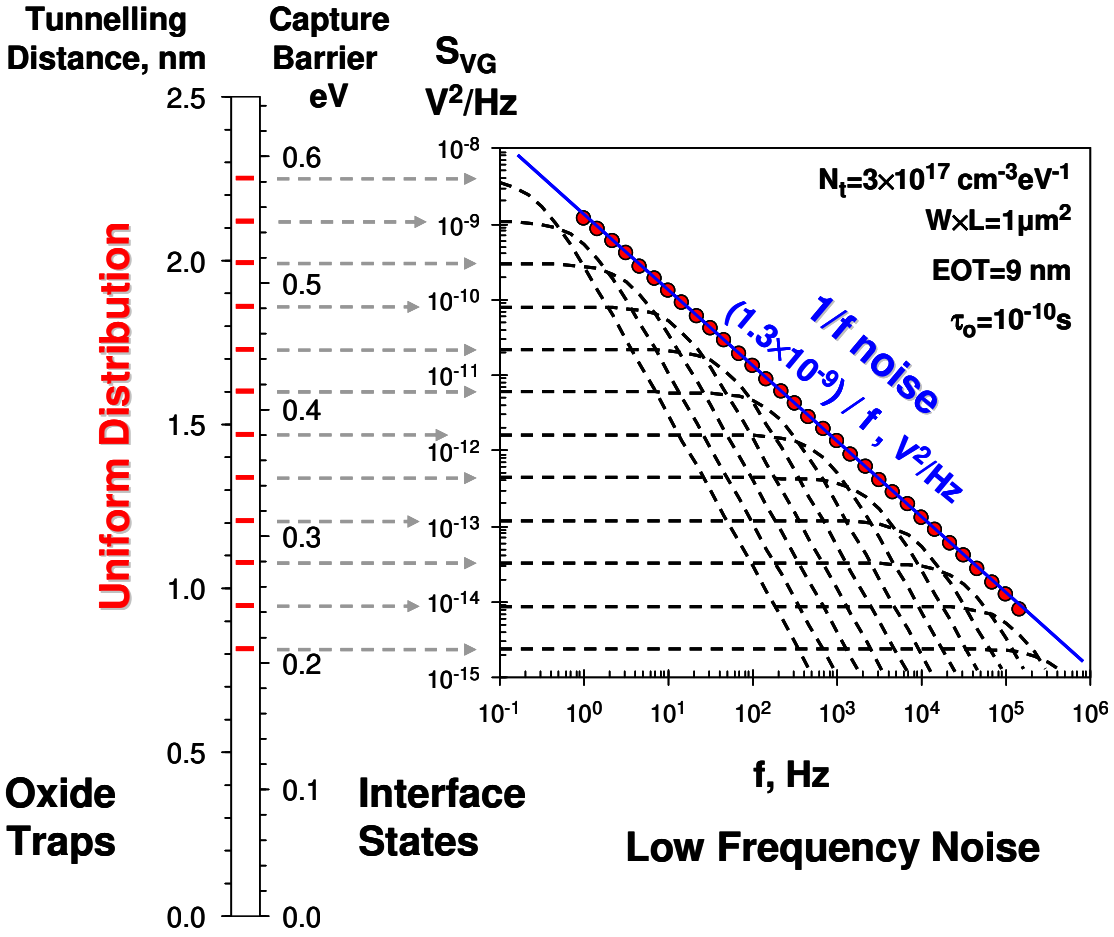


Figure 19. Uniform distribution in capture barrier energy ΔE_B=0.4eV of traps (interface states) and tunneling distance to oxide trap Δx_{ti}=1.5 nm results in 1/f noise after superposition of the individual Lorentzian spectra

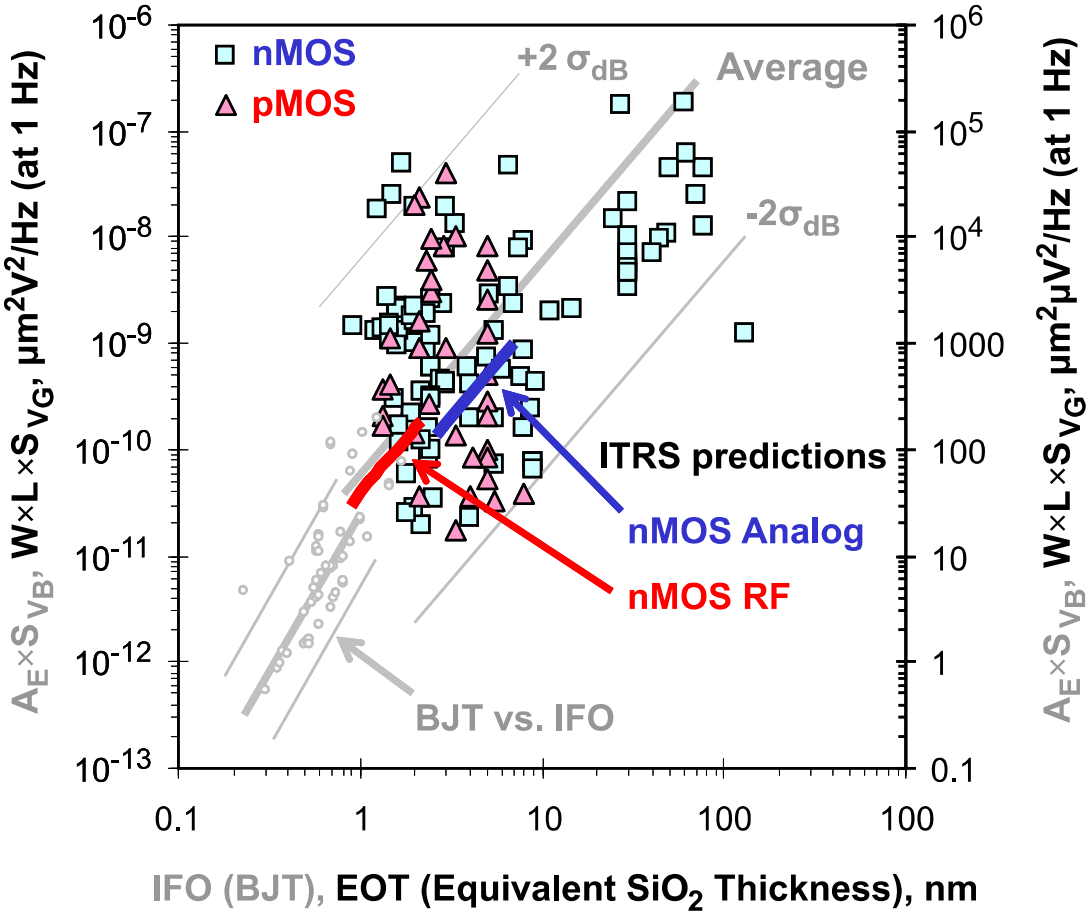
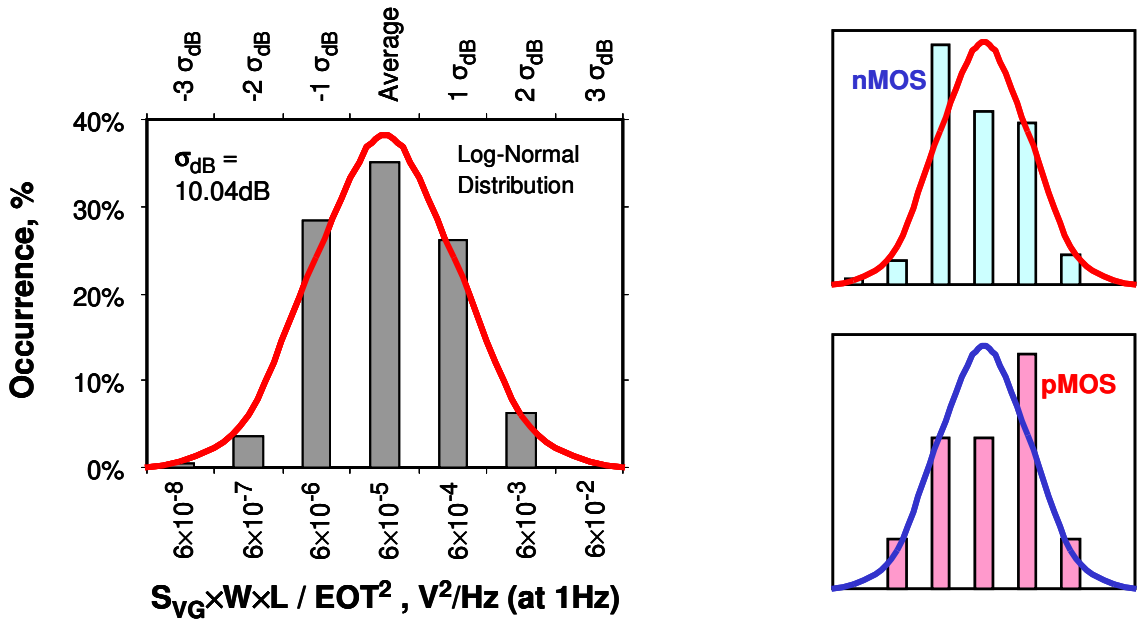


Figure 20. Input referred 1/f noise voltage in MOS transistors of unit gate area W×L=1µm^2 at 1Hz ($FOM_{S_{VG}}$) for nMOS (■) and pMOS (▲) transistors, according to eq. (106). Supplementary data from Figure 4 for input referred 1/f noise voltage in BJT (O) of unit emitter area A_E=1µm^2 ($FOM_{S_{VB}}$) vs. thickness of interfacial oxide IFO, where $FOM_{S_{VB}}$ is according to eq. (3), rewritten as $FOM_{S_{VB}}=A_E\times S_{V_B}(1Hz)=A_E\times K_F\times(\varphi_t)^2/1Hz$. See Figure 21 for values of Average and σ_{dB}.



a) b)

Figure 21. Distributions of input referred 1/f noise voltage in MOS transistors of unit gate area W×L=1μm^2 at 1Hz, corresponding to data in Figure 20, when a) all data are analyzed and b) when data for nMOS and pMOS transistors are analyzed separately. The labels (other than pMOS and nMOS), scales and ranges are the same in all plots, and are omitted in sub-figure b) for clarity.

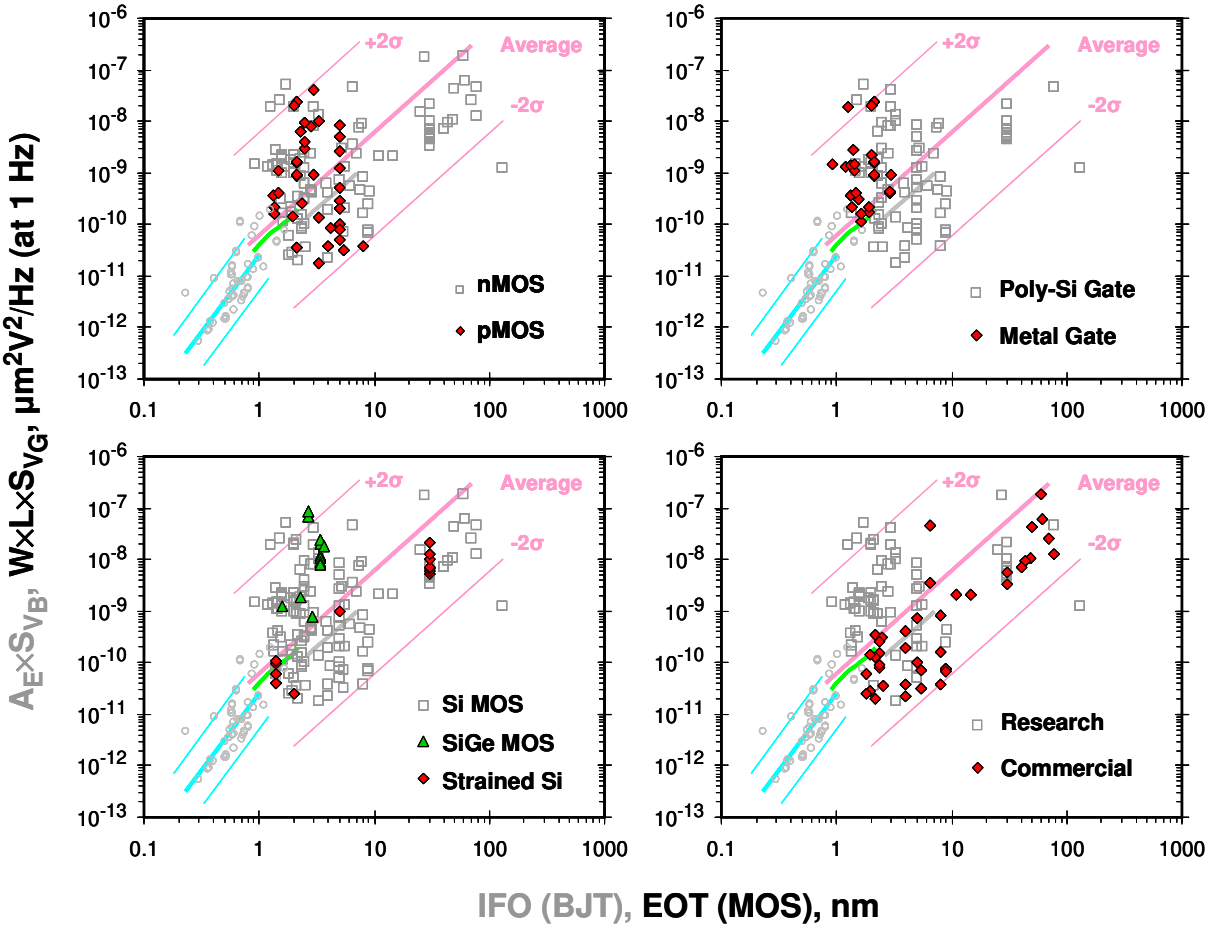
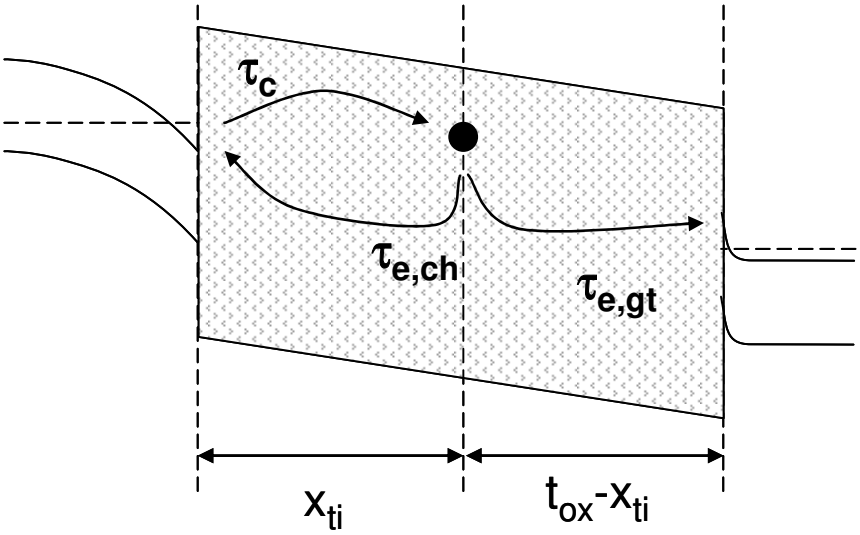
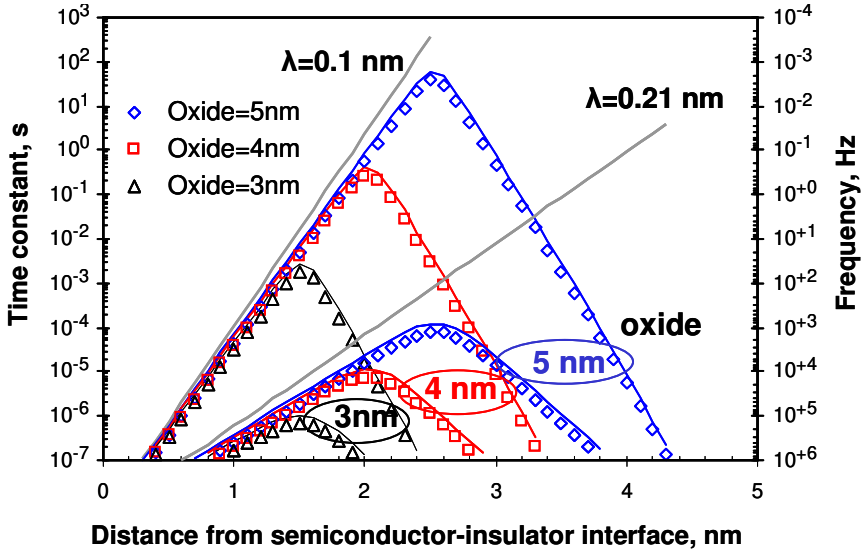


Figure 22. Top-down case study for input referred 1/f noise voltage in MOS transistors of unit gate area. The scales, trend lines and data for BJT (O) are as in Figure 20. Top-left plot: pMOS (◆) vs. nMOS (□) – data scatter similarly. Top-right plot: metal (◆) [50, 52, 57, 94, 99, 100, 109, 126, 128] vs. poly silicon gates (□) – no indication that metal gate reduces the noise. Bottom-left plot: strained lattice (◆) [82, 155] and SiGe (▲) [52, 57, 109] vs. Si (□) – no indication that strained lattice increases the noise and Ge decreases the noise. Bottom-right plot: commercial (◆) vs. research (□) MOS transistors – conservative maturity in commercial fabricators and risk for innovation is at research centers.



a)



b)

Figure 23. Problem in tunneling model for obtaining large time constants in thin oxides with larger attenuation distance. a) emission time constant is set by the shorter of distances to semiconductor and gate. b) the maximum time constant of tunneling can be low (<ms) in high-k dielectrics of thickness less than 5 nm. Straight lines are for the case when the tunneling to the gate is neglected.

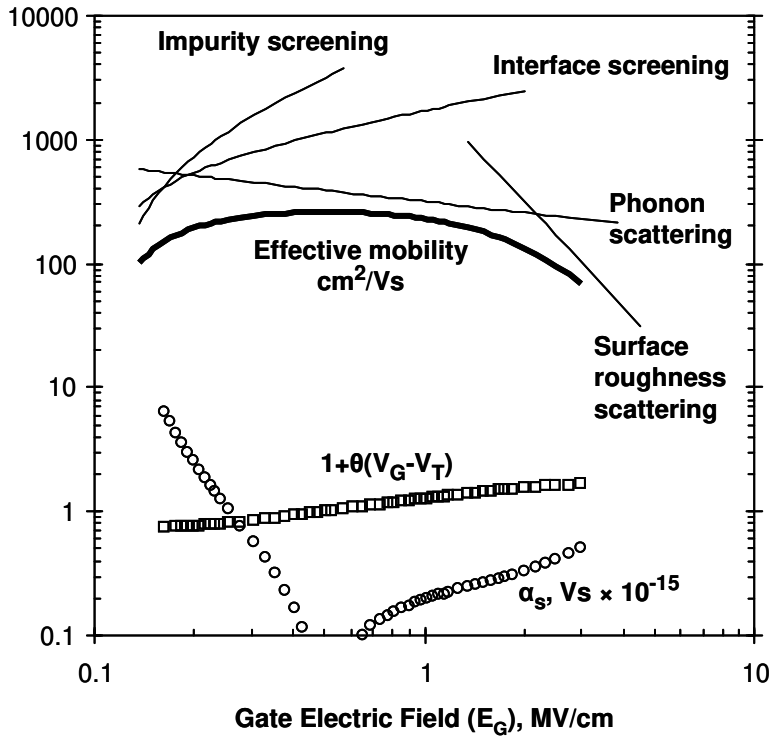
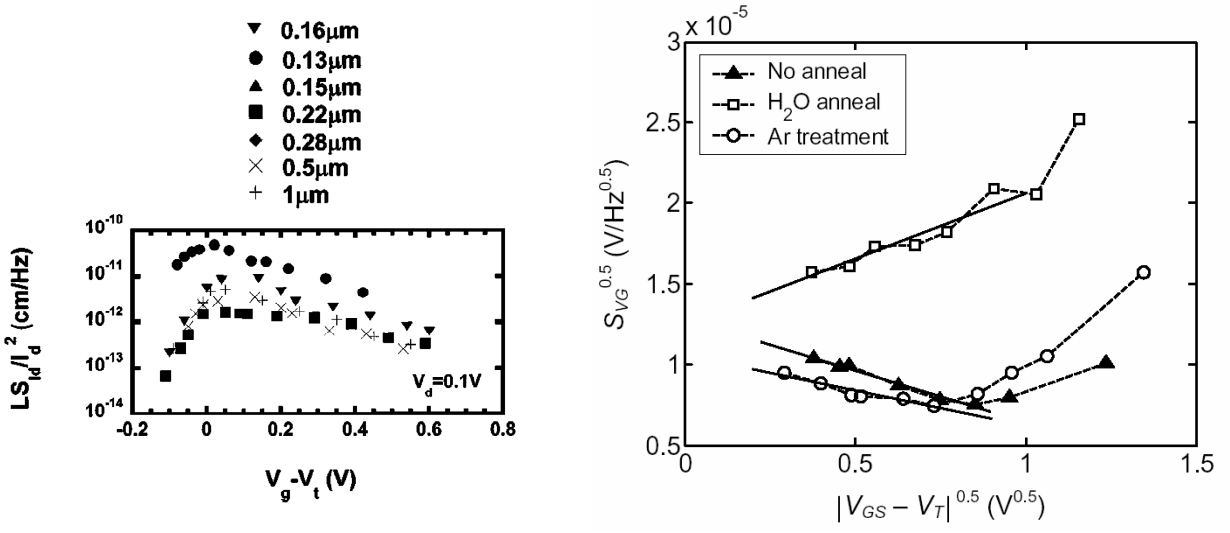


Figure 24. Variation of mobility μ (lines, in unit cm²/Vs), scattering parameter α_s (circles, in unit $Vs\times10^{-15}$) and the term for correlated mobility $[1+\theta(V_G-V_T)]$ (squares, dimensionless numbers) with gate electric field E_G



a) b)

Figure 25. Variation of noise at crossover between different scattering mechanisms, when the gate overdrive voltage (V_G-V_T) is increasing. a) Peaking of noise around threshold voltage due to dominance of Coulomb scattering, which ceases when the transistor is well above threshold [48]. b) Crossover between ceasing Coulomb scattering and rising phonon or roughness scattering [57].

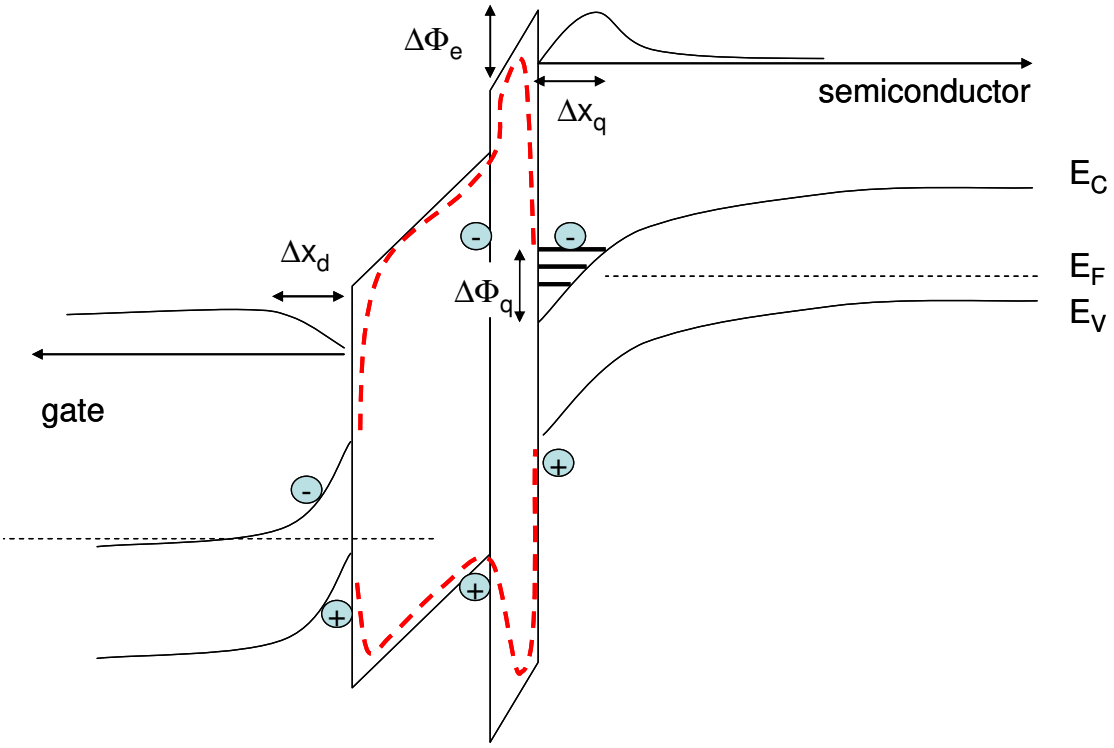


Figure 26. Depletion, quantum and barrier lowering effects at high electric field and thin gate insulator stacks

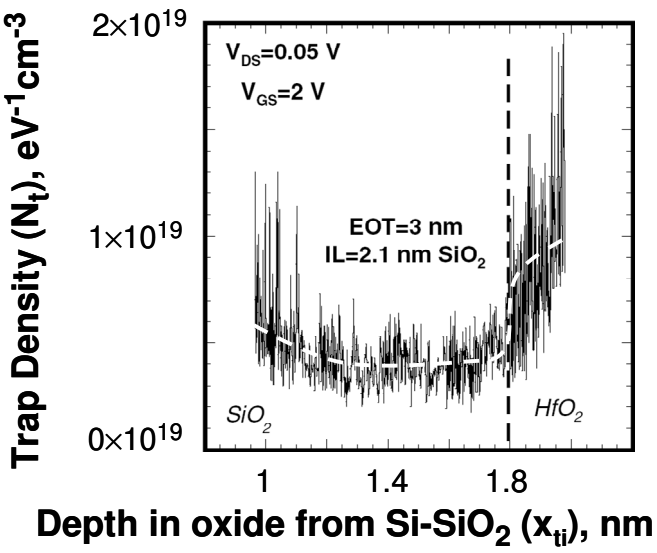


Figure 27. Qualitative oxide trap profiling of a MOS transistor with gate stack of 2.1nm SiO_2 interfacial layer at semiconductor interface and 5nm HfO_2 on top of it [140], according to eqs. (155) and (156), at assumptions that at given frequency f_i, the contribution of traps with time constant $\tau_i=1/(2\pi f_i)$ is dominant in the 1/f noise, and the tunneling attenuation distances $\lambda_{SiO_2}\approx\lambda_{HfO_2}$ of the materials in the gate insulator stacks are similar

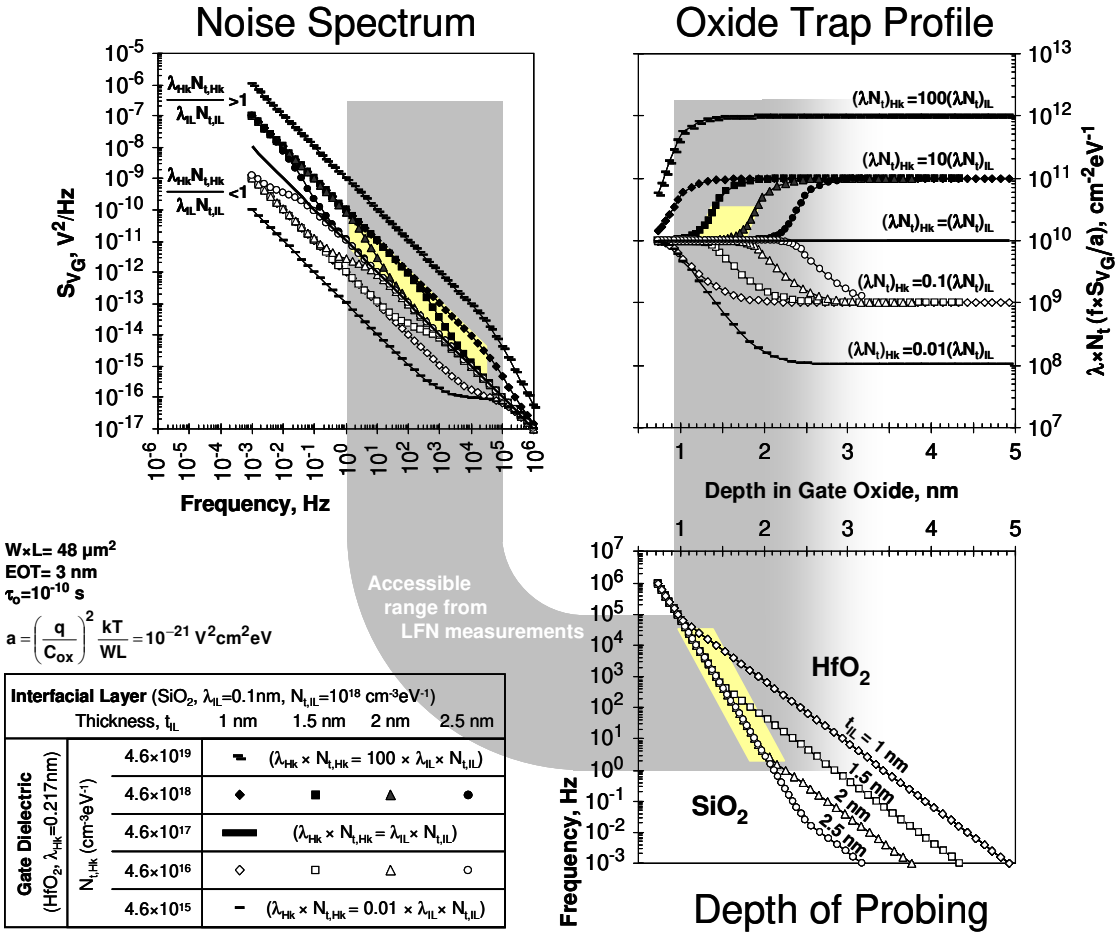


Figure 28. Oxide trap profiling of gate insulators with abrupt interfaces. Although λN_t has a step change at depth t_{IL}, the noise spectra and trap profiles exhibit transition regions. Shaded area corresponds to ranges accessible by low-frequency noise measurements. The solid triangles in highlighted regions of the shaded area correspond to measurement in Figure 27, and show two slopes in noise spectrum (top-left plot), probing only within the depth of interfacial SiO_2 layer (steep slope in bottom-right plot), and observation only of the tail in the transition region for $\lambda_{Hk}N_{t,Hk}$ below the abrupt interface at t_{IL}=2.1nm (top-right plot)

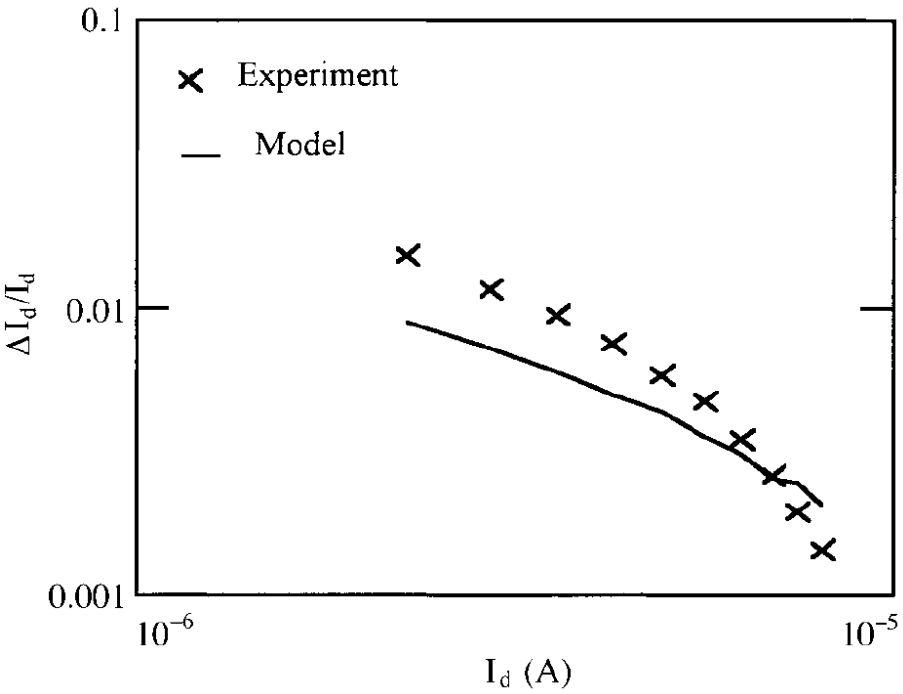


Figure 29. Some discrepancy between measured and predicted with the simplest model of eq. (180) values for the RTS amplitudes in nMOS transistor. W=0.2μm, L=0.25μm, EOT=4.5nm, V_d=50mV, from [74]

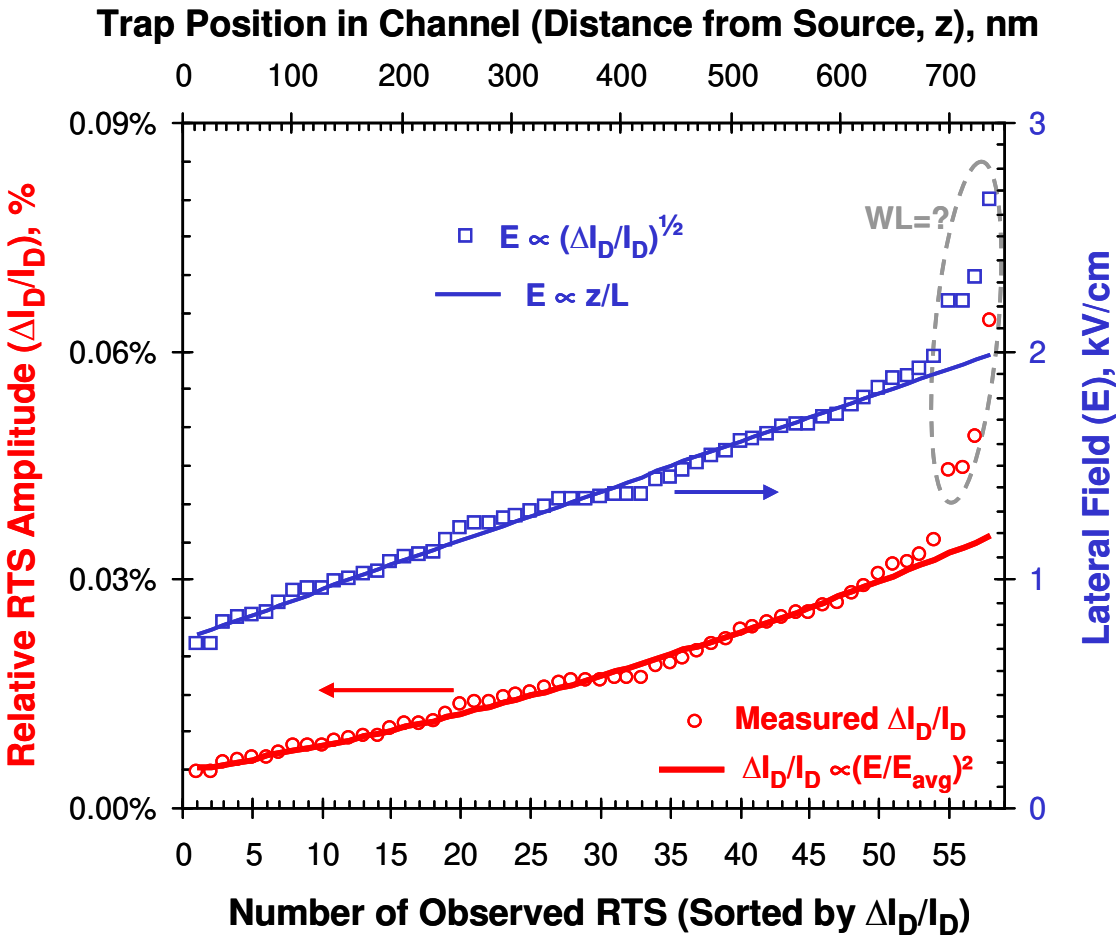


Figure 30. Variation of amplitudes of different RTS in identical MOS transistors at fixed experimental condition – see the text for values of device parameters and test conditions. Measured data for relative RTS amplitudes $\Delta I_D/I_D$ (circles) are from [68]. The different values E(z) of lateral electric field at different trap positions (z) along MOS transistor channel can explain the variation, according to [167] - see text. It is assumed that z∝Number, since the original data (circles) are given in scatter plot sorted by $\Delta I_D/I_D$. The data points for lateral electric field (squares) are obtained from $E/E_{avg} \propto \sqrt{\Delta I_D/I_D}$, according to eqs. (182) and (183). The linear fit E∝z/L (thin line through the squares) is expected for ohmic regime of operation. This linear fit for E is used to calculate the RTS amplitudes (thick line through circles), according to eq. (183).

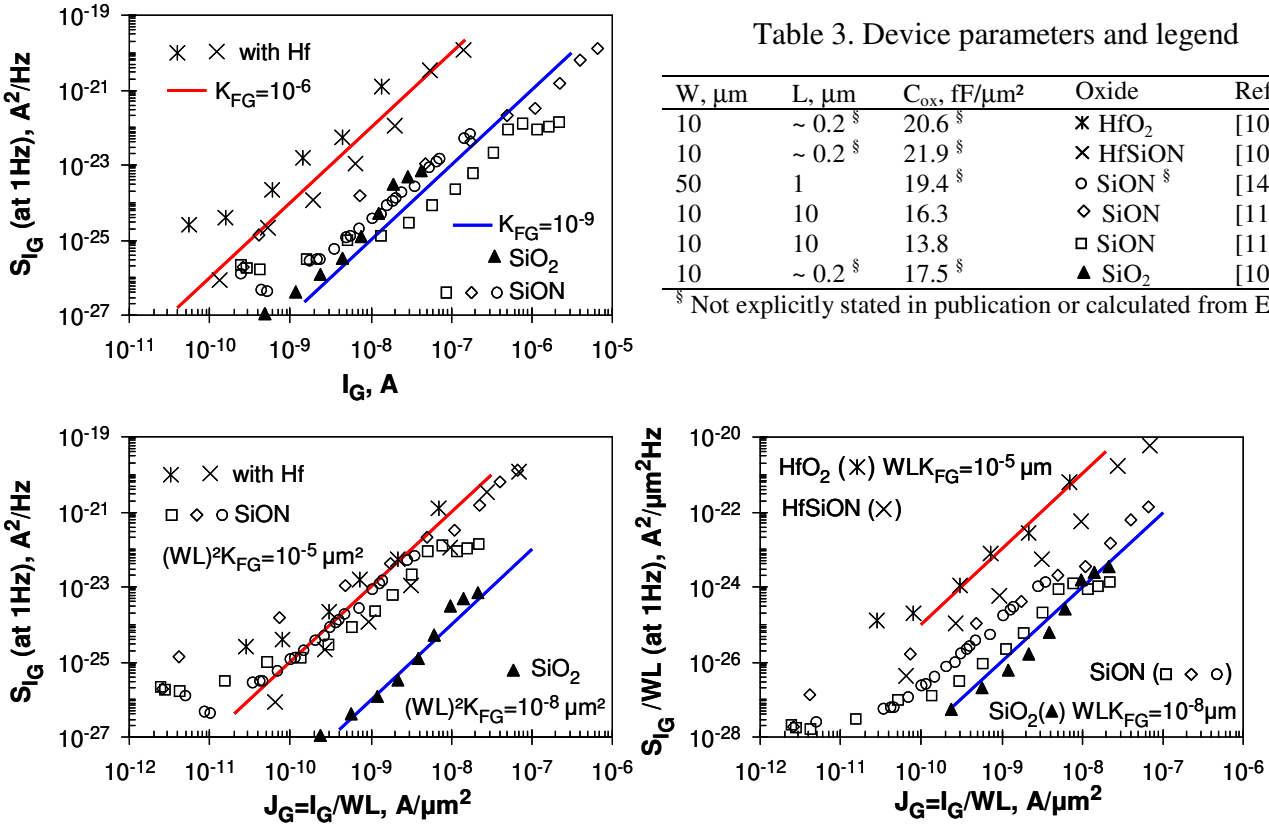


Table 3. Device parameters and legend

W, µm	L, µm	C_{ox}, fF/µm²	Oxide	Ref.
10	~ 0.2 §	20.6 §	✳ HfO_2	[101]
10	~ 0.2 §	21.9 §	× HfSiON	[101]
50	1	19.4 §	○ SiON §	[149]
10	10	16.3	◇ SiON	[110]
10	10	13.8	□ SiON	[110]
10	~ 0.2 §	17.5 §	▲ SiO_2	[101]

§ Not explicitly stated in publication or calculated from EOT

Figure 31. Flicker noise S_{I_G} in the leakage gate current I_G for MOS transistors with thin oxides analyzed by different figures of merit (FOM). Top-left: Physical quantities according to the table on right. Bottom-left: Using scaling rule for gate current ($J_G=I_G/WL$). Bottom-right: Using the regular scaling rule from eq. (14) for (reciprocal) areal dependence of noise.

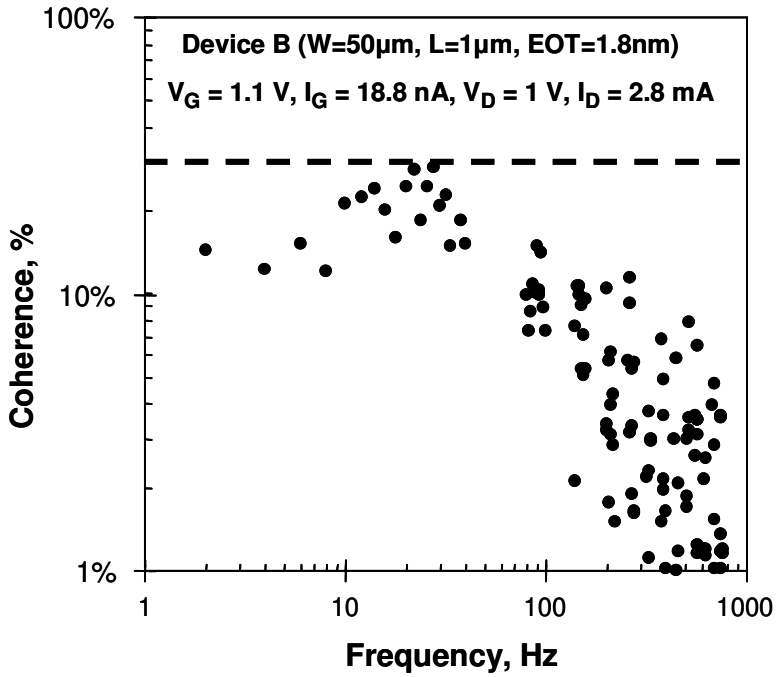


Figure 32. Weak correlation in the flicker noise between gate and drain currents obtained in [149] by measurement of coherence – see eq. (27). Coherence less than 30% indicates independent noise sources. The measurement with maximum value for the coherence is shown among the different biasing conditions that correspond to the data denoted with circles in Figure 31.

Sample: nMOS transistor

Channel Width W=50µm

Channel Length L=1µm

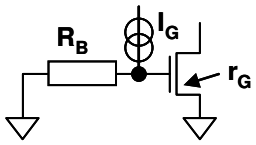
Equivalent Oxide Thickness EOT=1.8nm

Bias conditions

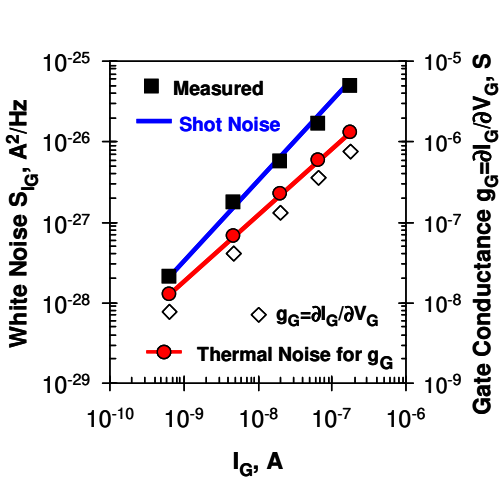
Source and Body $V_S=V_B=0V$

Drain $V_D=0.05V$

Gate $V_G=0.3V$, 0.5V, 0.7V, 0.9V and 1.1V



$$\frac{1}{r_{AC}} = g_G + \frac{1}{R_B}$$



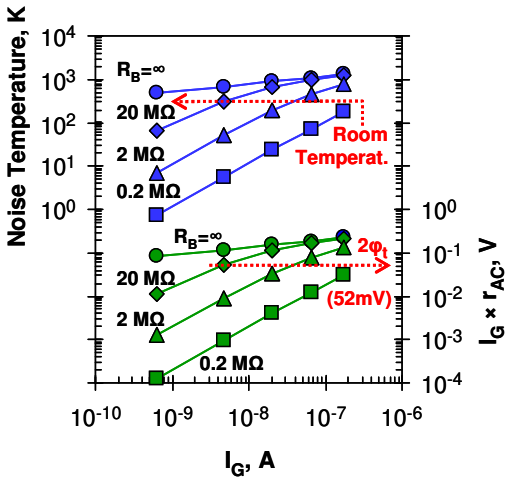
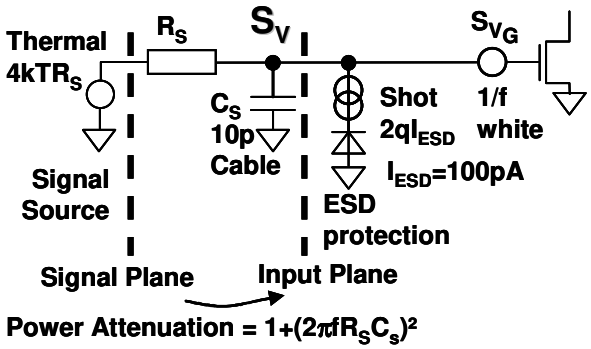


Figure 33. Evaluation of noise temperature of shot noise in the gate leakage current of nMOS transistor with thin oxide. Left-hand plot – the measured white noise S_{I_G} (filled squares) follows the relation for shot noise $S_{I,SH}=2qI_G$ (line through squares), and S_{I_G} does not follow the relation for thermal noise $S_{Ith}=4kT\cdot g_G$ (line through circles), where $g_G=\partial I_G/\partial V_G$ is the dynamic (AC) conductance of the gate insulator (open diamonds). Right-hand plot – the noise temperature also depends on the resistance of circuit, since $r_{AC}=R_B\|r_G$ is a parallel connection of device $r_G=1/g_G$ and bias R_B resistances, and by reducing R_B, the circuit noise temperature decreases, and the shot noise becomes "cold" once $I_G\times R_B<2\varphi_t$, which can be seen in the figure by comparing upper and bottom plots for noise temperature and "voltage drop" $I_G\times r_{AC}$, respectively.



Parameters of nMOS transistor, in linear (ohmic) regime

Parameter	Notation	Value
Channel Width	W	100 μm
Channel Length	L	1 μm
Equivalent Oxide Thickness	EOT	4 nm
Effective Mobility	μ	300 cm²/Vs
Oxide Trap Density	N_t	10^{17} $eV^{-1}cm^{-3}$
Tunneling Attenuation Distance	λ	0.1 nm
Gate Overdrive (Bias) Voltage	V_G-V_T	1 V
Drain Bias Voltage	V_D	0.1 V
Source and Body Bias Voltages	V_S=V_B	0 V
Drain (Channel) DC Current	I_D	≈ 2.5 mA
Gate to Drain Transconductance	g_m	≈2.6 mS
Drain to Source (Channel) Conductance	g_d	≈ 23mS

Figure 34. Circuit and nMOS transistor parameters used in the calculation of the noise levels and noise figure contours in Figure 35. The transistor is assumed operating in linear (ohmic) regime.

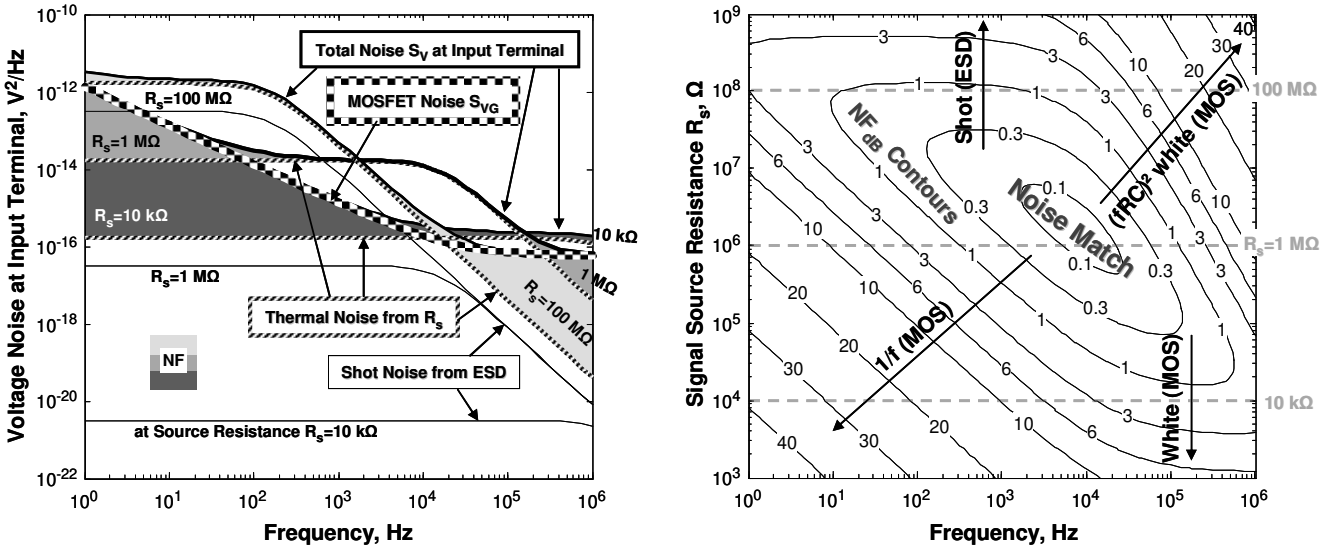


Figure 35. Voltage noise spectra (left-hand figure) and noise figure contours (right-hand figure) at the input plane of the circuit with nMOS transistor operating in ohmic regime shown in Figure 34. The lines and shaded areas are explained in the text.

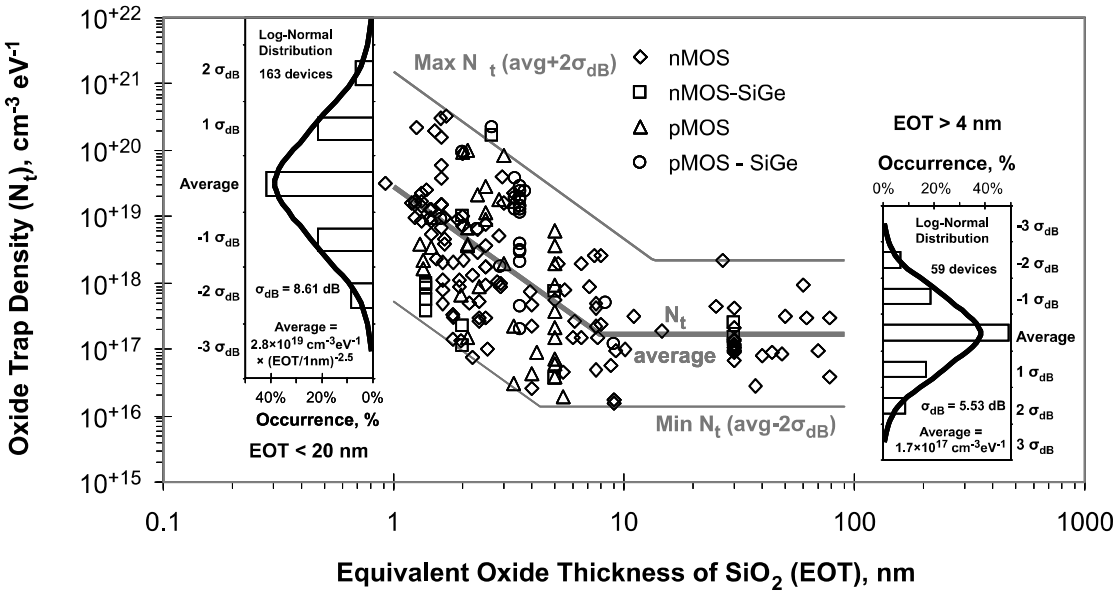


Figure 36. The variation of the values for equivalent oxide trap density N_t with equivalent oxide thickness suggests a crossover at EOT~4-5nm, and N_t increases in MOS transistors with C_{ox}>500nF/cm². The insets show the distribution around the average values for EOT<20nm (left-hand histogram) and for EOT>4nm (right-hand histogram). The data for 191 devices are from [22, 47, 48, 49, 50, 51, 72, 82, 88, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 149, 145, 202, 203, 204, 205, 206] for nMOS transistors (◇), from [109, 155, 207] for Ge on insulator (GOI) and strained on SiGe layer and control nMOS transistors (□), from [47, 52, 88, 94, 95, 96, 97, 100, 102, 104, 105, 109, 124, 125, 126, 128] for pMOS transistors (△), and from [52, 57, 79, 109, 208] for SiGe and SiGeC pMOS transistors (○).

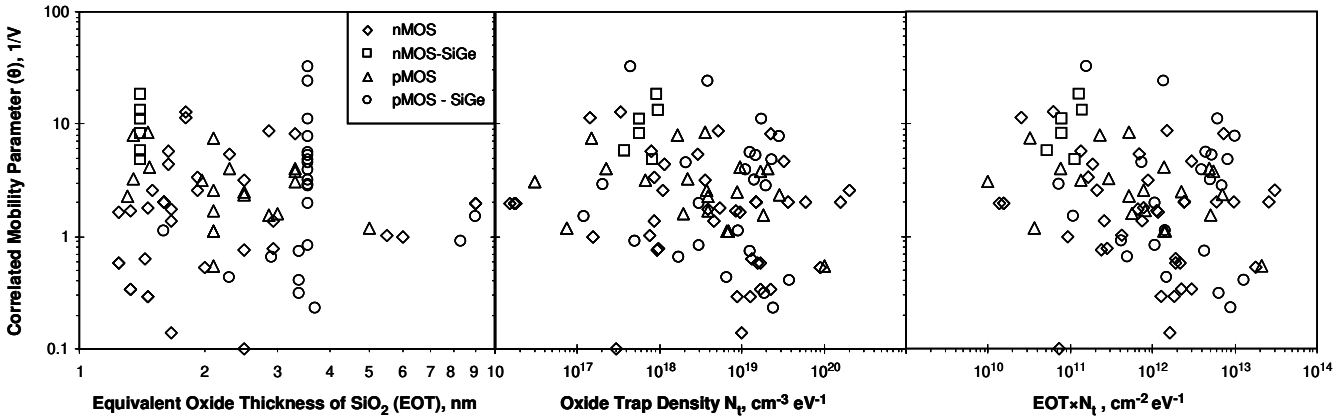


Figure 37. Random variation of the parameter θ for correlated mobility fluctuation in MOS transistors at high gate bias $(V_G-V_T)>0.3V$. According to eq. (243) $\theta \propto \mu C_{ox} \propto (N_t EOT)^{-1}$, but no proportionality is observed in the published data in [47, 48, 49, 50, 51, 95, 100, 106, 149] for nMOS transistors (◇), in [155] for strained on SiGe layer and control nMOS transistors (□), in [47, 52, 88, 94, 95, 100, 105, 124, 126, 128] for pMOS transistors (△), and in [52, 57, 79, 208] for SiGe and SiGeC pMOS transistors (○).

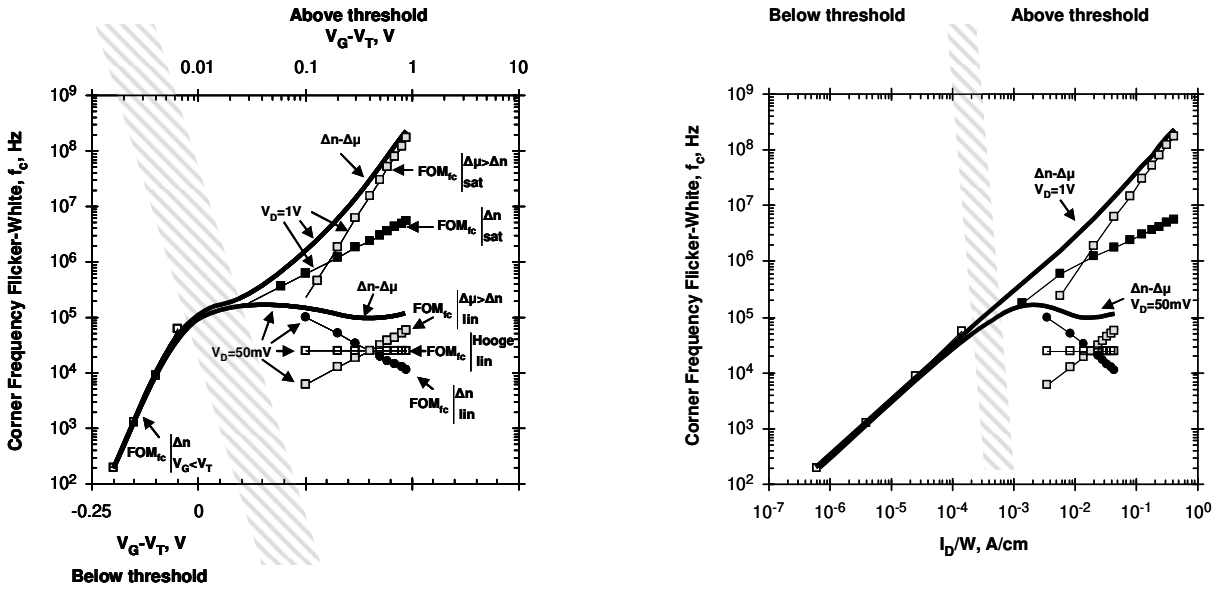


Figure 38. Evolution of corner frequency $f_c \equiv FOM_{fc}$ between flicker and white noise as function of gate bias (left-hand plot) and drain current (right-hand plot). Symbols represent contributions of different flicker noise sources, according to eqs. (255), (256), (257), (258), (259) and (260), as labeled in the left-hand plot. Thick lines represent the combined contribution of number and correlated mobility fluctuations. For saturation regime, V_D=1V, and for linear regime V_D=50mV. Diagonal patterned areas are the gate bias regions around transistor threshold voltage. MOS transistor parameters relevant to the figure are: Gate oxide capacitance per unit area $C_{ox}=10^{-6}$F/cm²; Mobility μ=100cm²/Vs; Correlated mobility parameter $\theta=2.5V^{-1}$; Oxide trap density $N_t=10^{18}cm^{-3}eV^{-1}$; Tunneling attenuation distance λ=0.1nm; Hooge parameter $\alpha_H=10^{-4}$; Gate length L=1μm. For devices with other parameters, one can use the scaling rules given in the text.

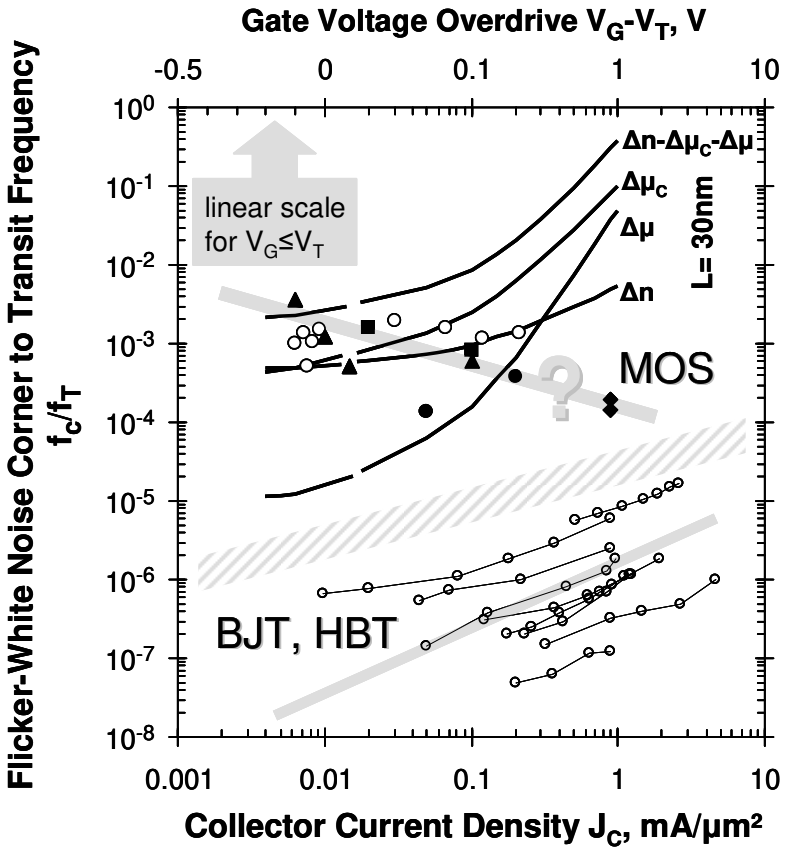


Figure 39. Ratio $f_c/f_T \equiv FOM_{f_c/f_T}$ of flicker-white noise corner frequency f_c to transit frequency f_T. The diagonal patterned area separates data for MOS transistors from data for BJT and HBT. Data for BJT and HBT originate to several publications and were reported earlier in [29]. Since no publication for MOS transistors reporting f_c and f_T simultaneously from measurements was found, then the missing portion of data for MOS transistor was recalculated using long channel approximation [169] applied to available data for (○) nMOS transistors from 0.8μm node [136], (●) for nMOS and pMOS transistors from 0.35μm SOI node [86], (◆) for RF nMOS transistors with drawn channel length L=0.13μm, width W=72 and different number of fingers [214], (▲) for nMOS and pMOS transistors from 0.13μm node [102], and (■) for nMOS transistor from 0.09μm node [105]. The lines are calculated according to eq. (270) for a virtual MOS transistor of channel length L=30nm, mobility μ=300cm²/Vs, C_{ox}=3μF/cm² (EOT~1.15nm), saturation velocity $v_{sat}=10^7$cm/s, with oxide trap density $N_t=3\times10^{17}cm^{-3}eV^{-1}$ and tunneling attenuation distance λ=0.1nm for the number fluctuation Δn, with Coulomb screening parameter $\mu_{C0}=3\times10^8$cm/Vs for mobility fluctuation $\Delta\mu_C$, with $\theta=3V^{-1}$ for phonon or roughness scattering ($\alpha_s=q\theta/\mu C_{ox}=5.3\times10^{-16}$Vs) for mobility fluctuation Δμ, and then all the components were combined together $\Delta n - \Delta\mu_C - \Delta\mu$. Note that a linear scale is used for (V_G-V_T) in sub-threshold, when $V_G<V_T$. Agreement between prediction by eq. (270) and experiments exists at low gate bias, but a discrepancy is apparent at high gate overdrive.

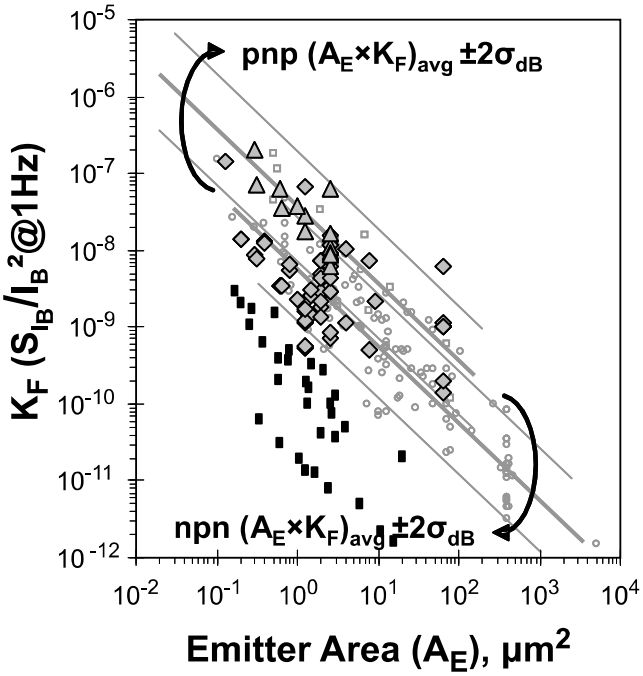


Figure 40. Normalized noise $K_F=f\times S_{I_B}/I_B^2$ in SiGe HBTs (heterojunction bipolar transistors) is within the range for polysilicon emitter bipolar junction transistors (BJTs), when plotting versus emitter area A_E. Data for npn HBT (◆) are from [34, 35, 36, 37, 60, 81, 212, 224, 225]. Data for pnp HBT (▲) are from [33, 34, 60]. Small circles (○) are data for npn BJTs from Figure 2 with trend npn$(A_E\times K_F)_{avg}$= 5.6×10^{-9}μm² and σ_{dB}=3.38dB. Small squares (□) are data for pnp BJTs from Figure 14 with trend pnp$(A_E\times K_F)_{avg}$= 3.8×10^{-8}μm² and σ_{dB}=3.56dB. Data for SiGeC HBTs (■) were treated separatedly as they show a clear improvement over the rest, with $(A_E\times K_F)_{avg}$= 1.25×10^{-10}μm², 16.5dB below the range for npn HBT. Data for SiGeC are from [40, 64, 65, 228, 229, 230].

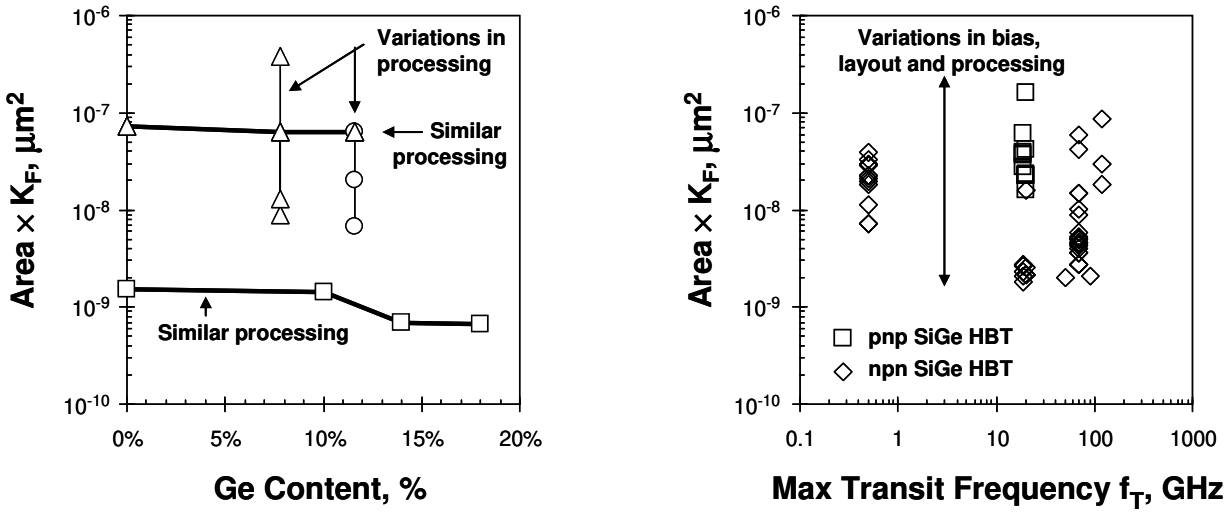


Figure 41. The flicker noise in SiGe HBTs is nearly independent of Ge content in the base (left-hand figure) and transit frequency f_T (right-hand figure). Other factors, but not Ge, affect the 1/f noise in SiGe HBT-see the text. In left-hand figure, the data are for npn SiGe HBTs from [224] processed in similar conditions (□), and (○) from [35] and (△) from [225] processed at variety of conditions, which resulted in scattering of data. In the right-hand figure, the data are from [33, 34, 36, 37, 60, 81, 212, 224] from technology nodes with different maximum transit frequency, and the vertical scattering of data is due to layout, size and bias dependences of 1/f noise in these SiGe HBTs.

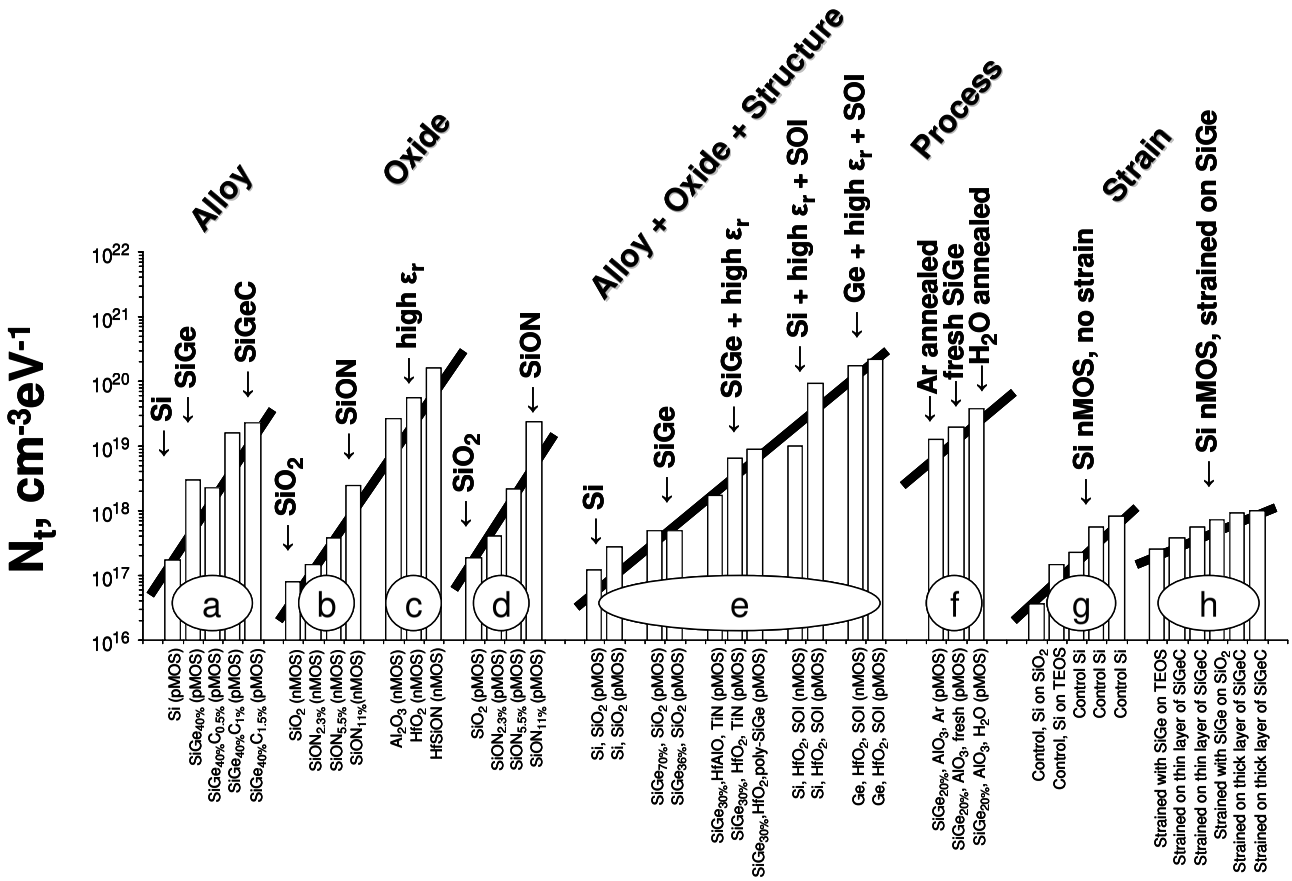


Figure 42. Flicker noise penalty in MOS transistors when using composite materials either in the channel or in the gate dielectric. Oxide trap density N_t is proportional to the level of the input referred 1/f noise voltage S_{V_G}, and it is according to eqs. (104) and (241). Data (a) are from [79] for pMOS transistors with channel of ternary SiGeC alloy, and N_t increases with increasing the content of carbon C in the alloy. Data (b) and (d) are from [139] for silicon nMOS and pMOS transistors, respectively, in which an increase of N_t is observed when using nitridation of silicon oxide dielectric in order to increase the gate breakdown voltage. Data (c) are from [48] where the use of high-k dielectrics in Si nMOS transistors was accompanied also with increase of N_t. Data (e) are from several publications, which reported variations of N_t when increasing complexity of the MOS structures, changing the MOS channel alloy from Si to SiGe [208, 231], then using high-k dielectrics in the gate stack of SiGe MOS transistors [52], followed by semiconductor-on-insulator (SOI) structures based on Si and on Ge, and with high-k gate dielectric [109]. Data (f) are from [57] reporting variation of N_t by post-annealing after complete fabrication of SiGe pMOS transistor. Data (g) and (h) are from [155, 207] for nMOS transistors without and with strained Si lattice on the top of SiGe layer for channel.

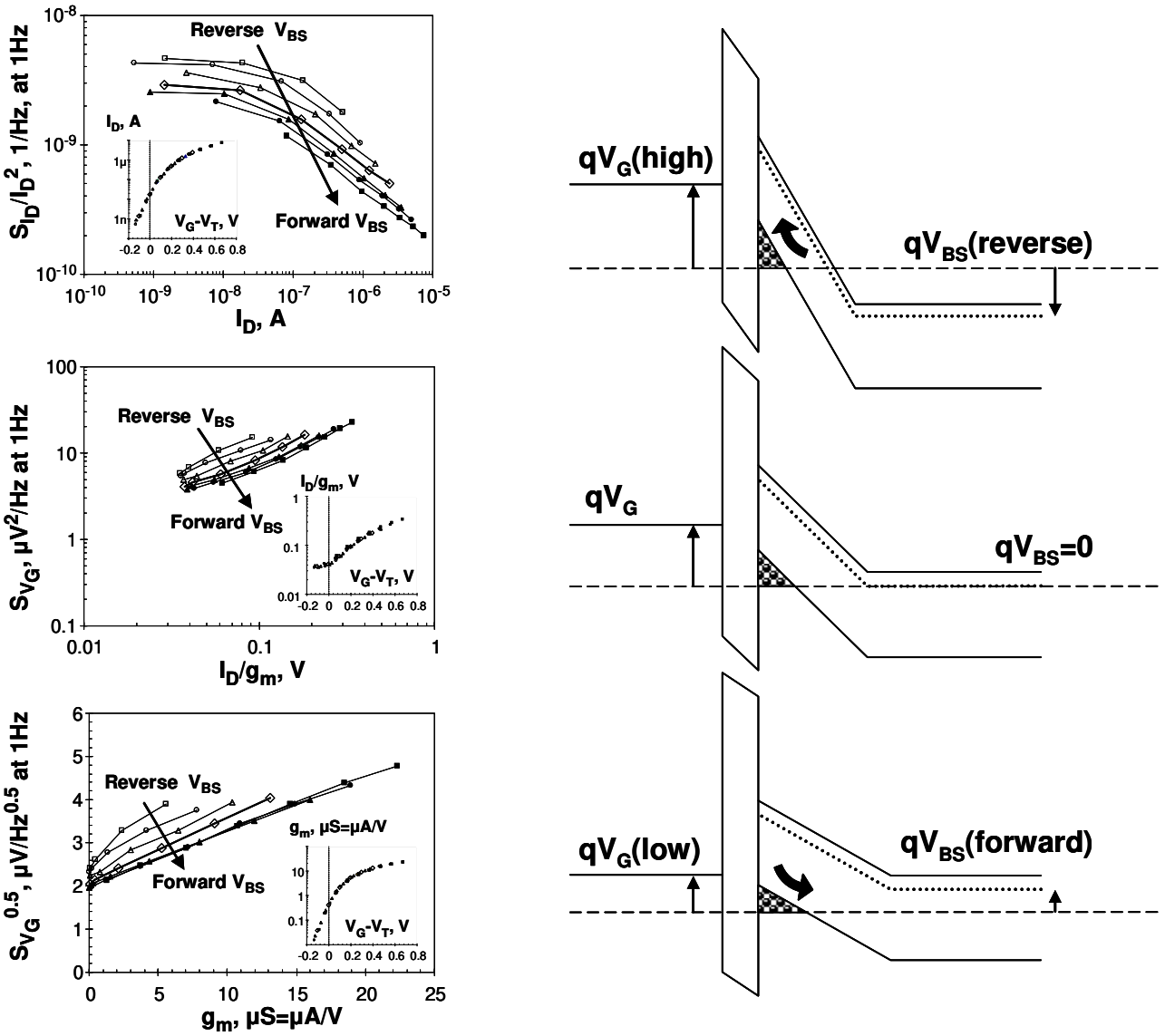


Figure 43. Variation of flicker noise with body bias V_{BS} in pMOS transistor 3×3μm² (left-hand plots) vs. DC drain current I_D, small signal transconductance g_m and their ratio, which are independent of V_{BS}, as depicted in the insets [245]. The different V_{BS} are from −0.6V (reverse, open symbols) to +0.6V (forward, filled symbols) in steps of 0.2V. Reverse or forward body bias is in respect to the direction for conduction of the body−source p−n junction. At given values for the quantities in the horizontal axes, and therefore constant carrier density in the channel, the noise decreases by a transition from high gate bias and reverse body bias toward low gate bias and forward body bias. This transition indicates a decrease of noise when the current flow is moved from the semiconductor-dielectric interface (surface channel at reverse body bias and high gate bias) toward the bulk of the semiconductor (buried channel conduction at forward body bias and low gate bias), as illustrated in the right-hand figures.

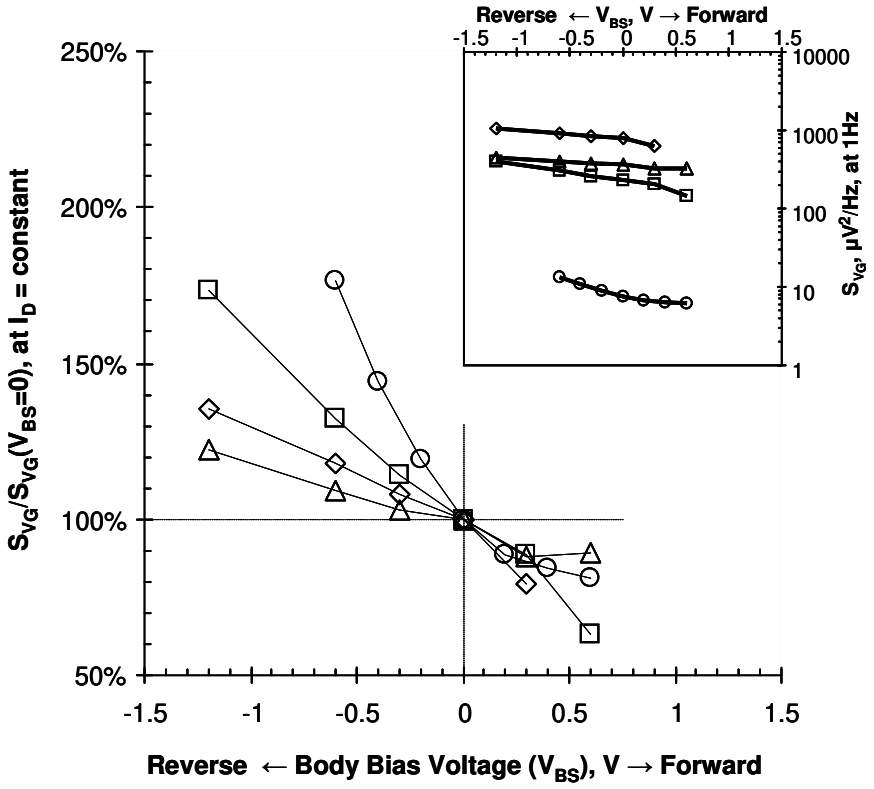


Figure 44. The relative variation of flicker noise with body bias V_{BS} in pMOS transistors (at constant drain current I_D, adjusted by gate bias voltage (V_G) implies that the noise increases at reverse body bias, while the forward body bias can be favorable for low noise circuits. Data for Si pMOS transistor 3×3μm² (○) are from [245] at V_D=0.6V and geometric average of S_{V_G} at I_D≈0.1μA, 0.3μA and 1μA. Data for SiGe pMOS transistors W×L=10×1μm² are from [52] at V_D=0.05V for device with TiN gate and HfAlO gate dielectric (□) by geometric average of S_{V_G} at I_D≈2μA and 17μA, for device with TiN gate and HfO_2 gate dielectric (△) at I_D≈41μA, and for device with poly-Si gate and HfO_2 gate dielectric (◇) by geometric average of S_{V_G} at I_D≈4.2μA and 41μA. The inset shows the data for the gate referred noise S_{V_G} in absolute values.

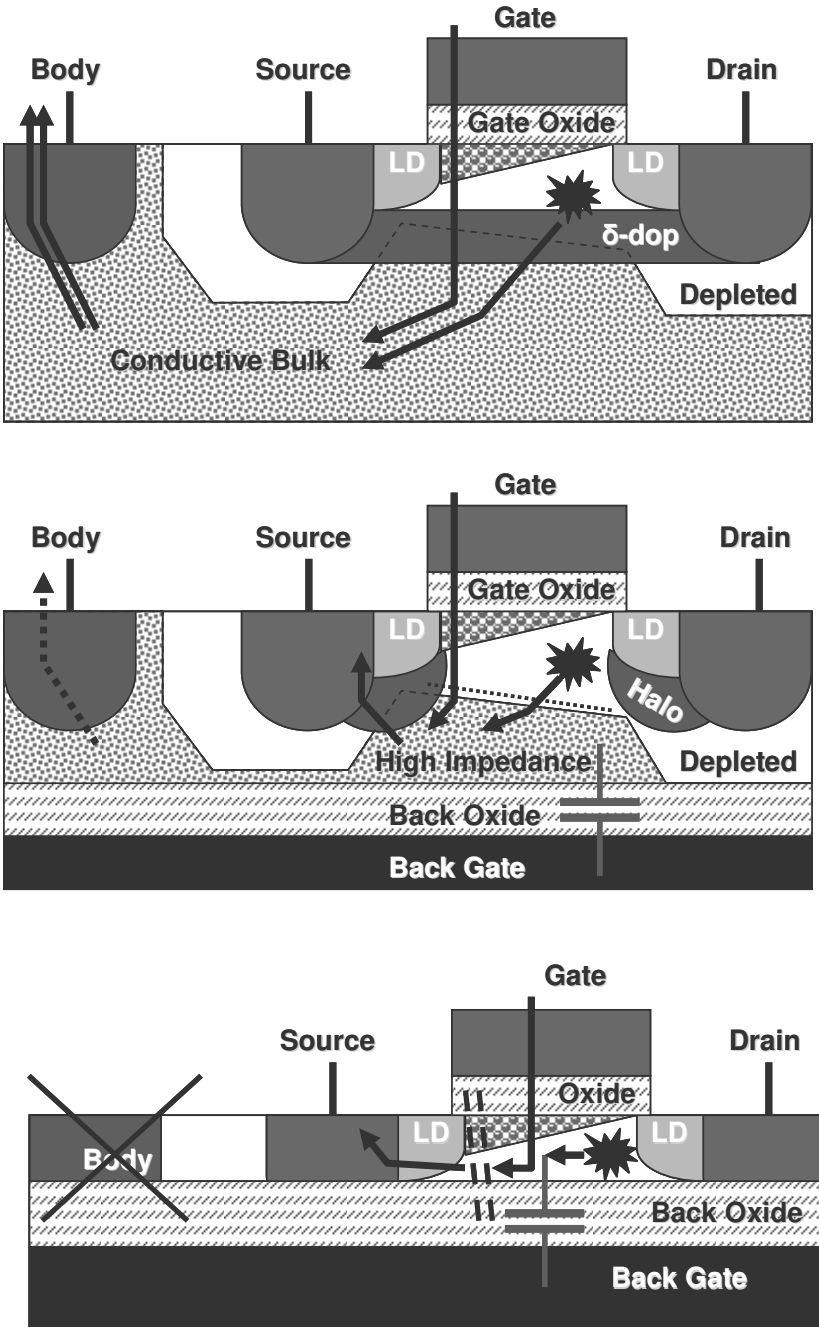
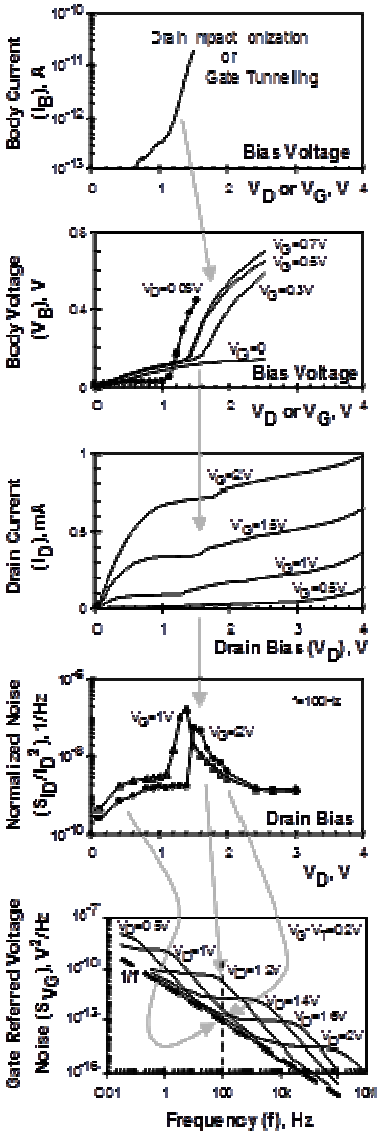


Figure 45. Evolution of MOS transistor structure from bulk device a), toward partially depleted b), and fully depleted c) SOI structures. LD stands for low-doped extensions of drain and source regions. The arrows show the paths of the non-channel charge carries (e.g. holes in nMOS transistors), generated by impact ionization or due to gate leakage owing to tunneling (valence-band tunneling, for example in nMOS). The dashed lines illustrate the noise sources which are associated with the non-channel carries.



a)

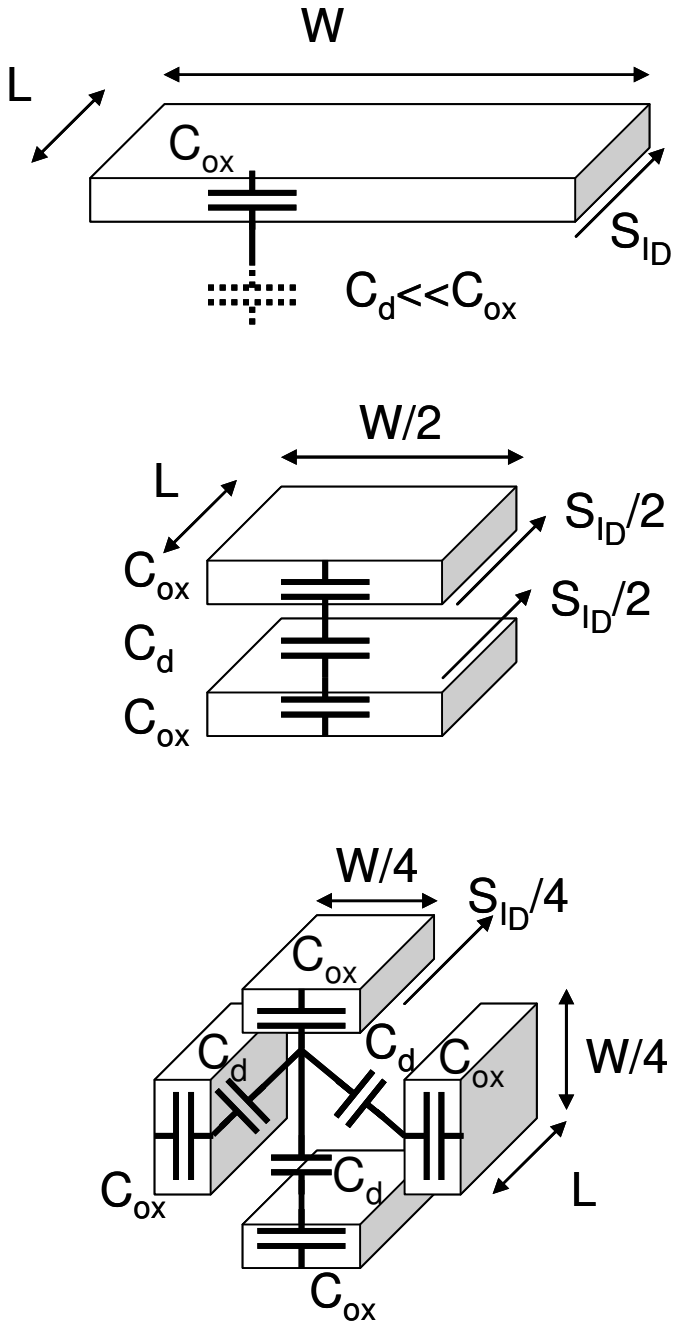
b)

c)

d)

e)

Figure 46. Effects related to the current I_B flowing in the body of partially depleted (PD) SOI MOS transistors with unconnected body terminal. The gray arrows depict the relation between the different effects. a) The increase of the bias voltages increases the body current I_B, illustrated with data for gate leakage current form [250]. b) Due to the increase of I_B, the body voltage V_B increases with the drain bias (lines [252]) and gate bias (circles [253]), biasing in forward the body-source junction in the PD SOI MOS transistor. c) The increase of V_B causes a decrease of the threshold voltage of the transistor, resulting in a higher (and dependent on I_B) overdrive (V_G–V_T) at constant gate bias V_G=constant, and a kink in the output characteristics of PD SOI MOS transistor emerges [254]. d) At the onset of the kink, a "sudden" increase of the noise occurs at given low frequency, as originally reported in [254] as noise "overshot" for f=100Hz. e) The noise "overshot" at fixed frequency is apparent owing to evolution of filtered white noise associated with the impedance of the body and shot noise in I_B [255].

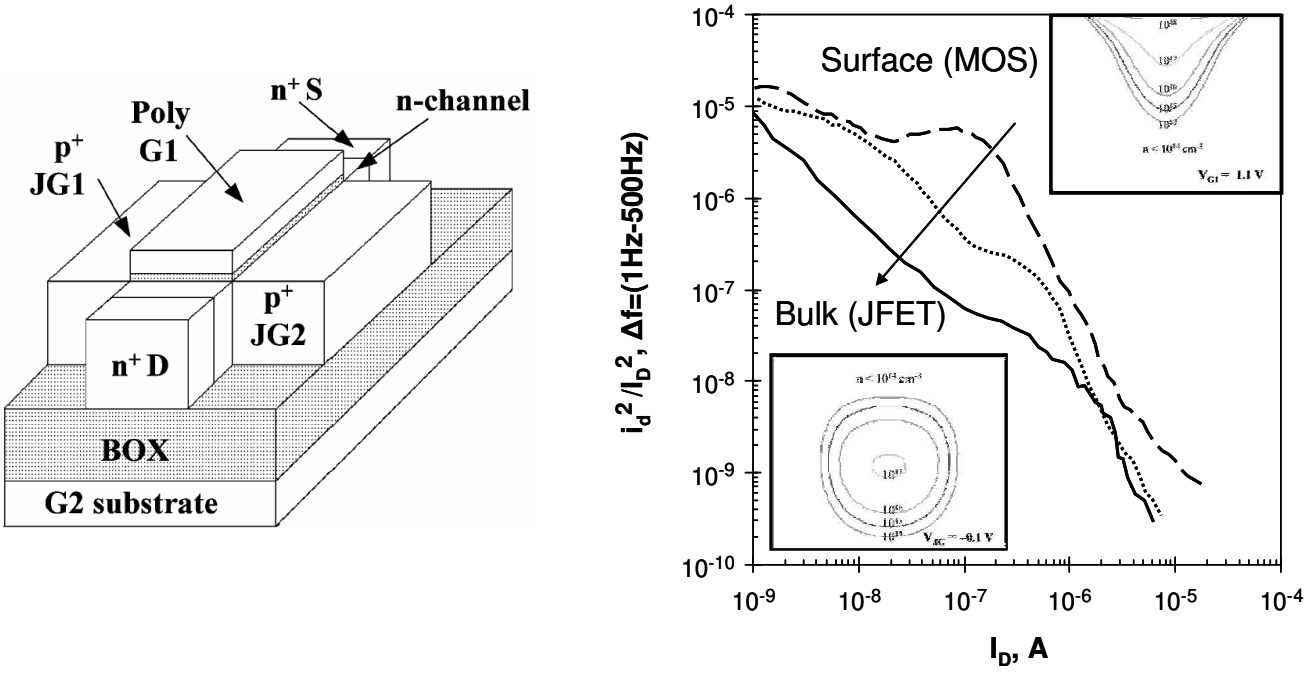


Singe gate (e.g. bulk) MOS; N=1

Two-gate (e.g. FD SOI) MOS; N=2

Four-gate (e.g. Fin FET) MOS; N=4

Figure 47. Capacitive coupling paths, which reduce the noise in multiple-gate MOS transistors. From top to bottom: single gate (number of gates N=1), two-gate (N=2) and four-gate (gate all around, N=4) MOS transistors. In the bottom figure for N=4, the capacitive paths only for the top gate are shown. The semiconductor channel is under the top gate and between the gates. C_{ox} and C_d are capacitances per unit area of gate dielectric and depleted semiconductor, respectively.



a) b)

Figure 48. Crossover between surface (MOS) and bulk (JFET) noise in depletion mode field-effect transistor [246]. a) Structure of four-gate MOS-JFET with top and bottom MOS gates (G1, G2) and left and right junction gates (JG1, JG2). b) The noise decreases when the conduction is moved from surface (dashed line, surface in accumulation) into the bulk (dotted line, surface in depletion) by exchanging the role of the MOS and junction gates. Further reduction of noise is achieved by inverting the surface (solid line). The insets show channel carrier distributions for surface and bulk conduction.

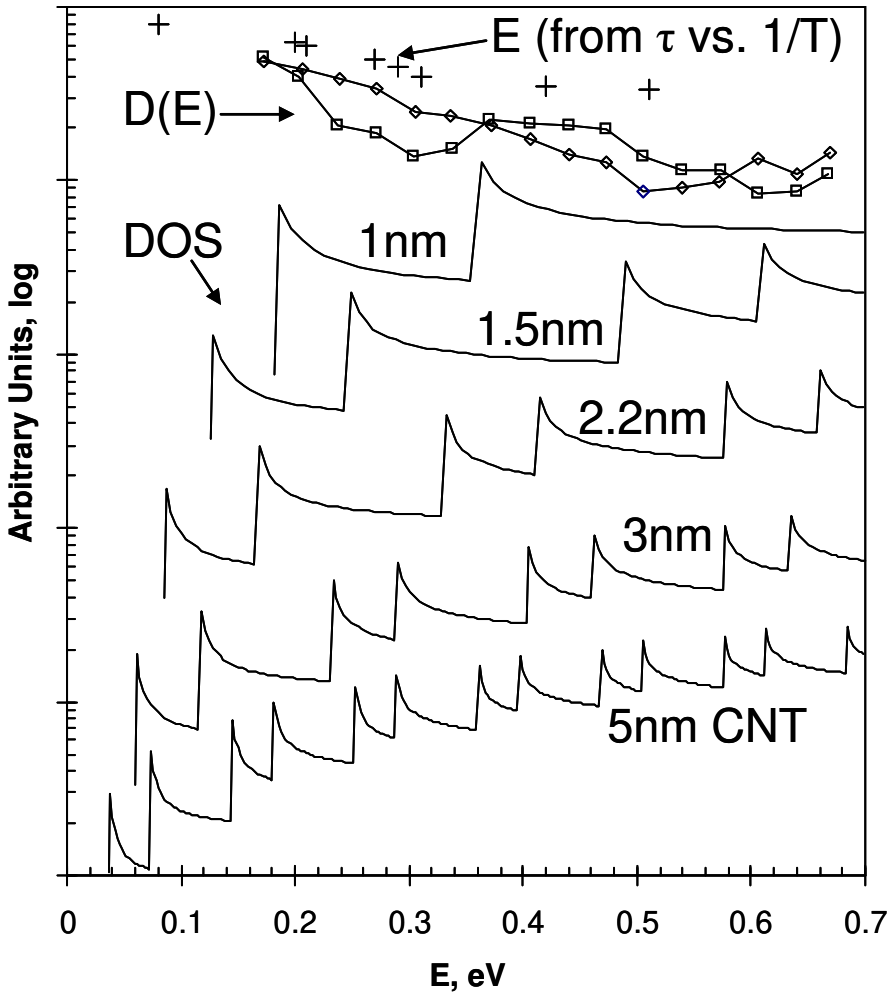


Figure 49. Distributions vs. energy in carbon nanotubes (CNT). Crosses are activation energies obtained from variation of the corner frequency of Lorentzian noise vs. temperature [268]. Lines with symbols are distributions of trap density D(E) vs. activation energy, obtained from the variation of slope of 1/f noise vs. temperature [268]. Lines without symbols are distributions of density of states (DOS) in CNT with different diameters, calculated in respect to Fermi level, according to [264]. Different lines are plotted with vertical offsets for clarity.

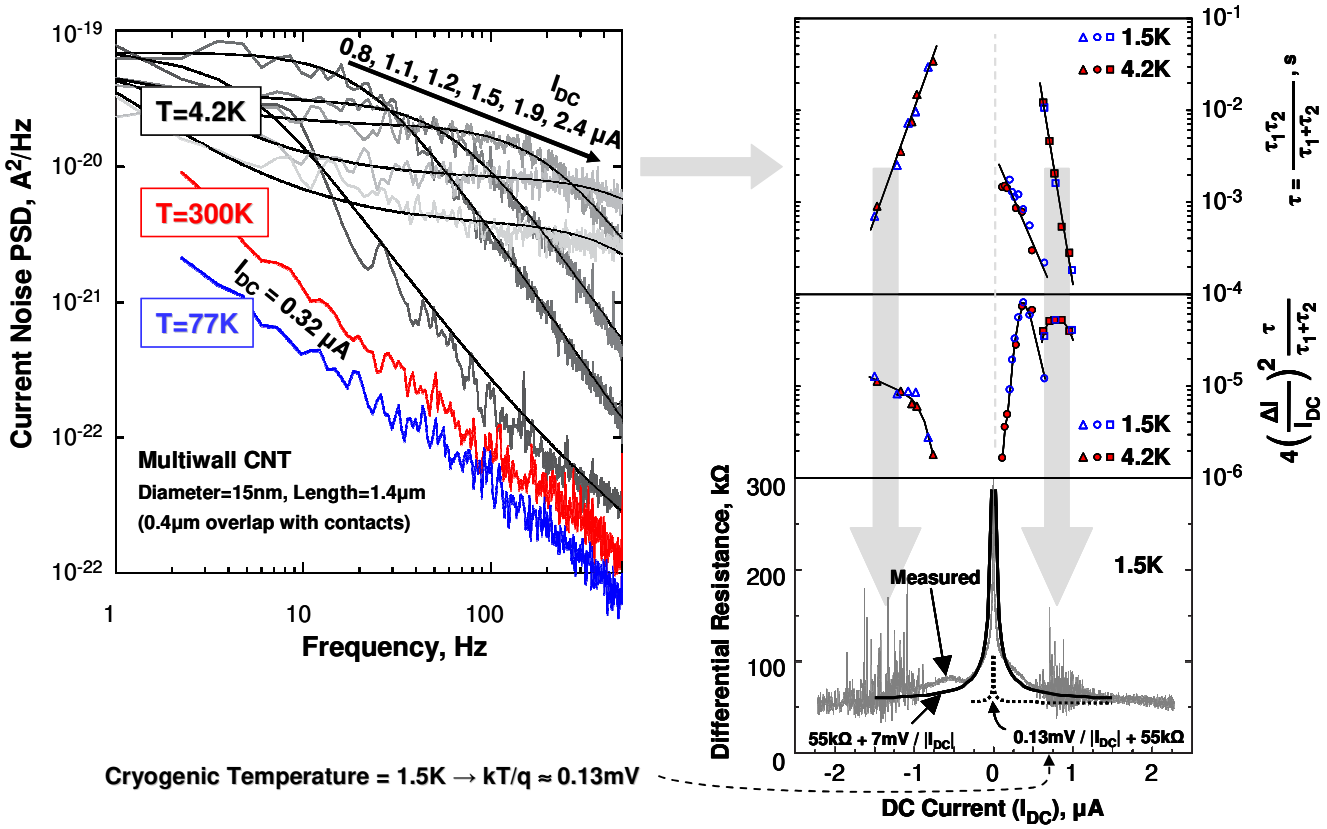


Figure 50. Evidences for tunneling junctions at the contact between gold pads and multiwell carbon nanotube placed on the pads [272]. Left figure: The noise at 77K and above is 1/f and it scales obviously as I_{DC}^2. 1/f noise at low current 0.32μA is shown to avoid overlap with spectra at low temperatures. The noise at 4.2K and below is with Lorentzian spectrum, which corner frequency $(2\pi\tau)^{-1}$ increases with bias, while the low-frequency plateau S_0 decreases. Right figure, from top to bottom: Lorentzian (RTS) time constants τ (bias dependent, but temperature independent), magnitude prefactor S_0/τ and differential resistance (measured with a lock-in amplifier at 30Hz, and proportional to $7mV/I_{DC}$, instead to $0.13mV/I_{DC}$, as the thermal voltage is $\varphi_t=kT/q\approx0.13mV$ at T=1.5K). The asymmetric characteristics as function of bias polarity suggest junctions at contact between nanotube and metal pads.

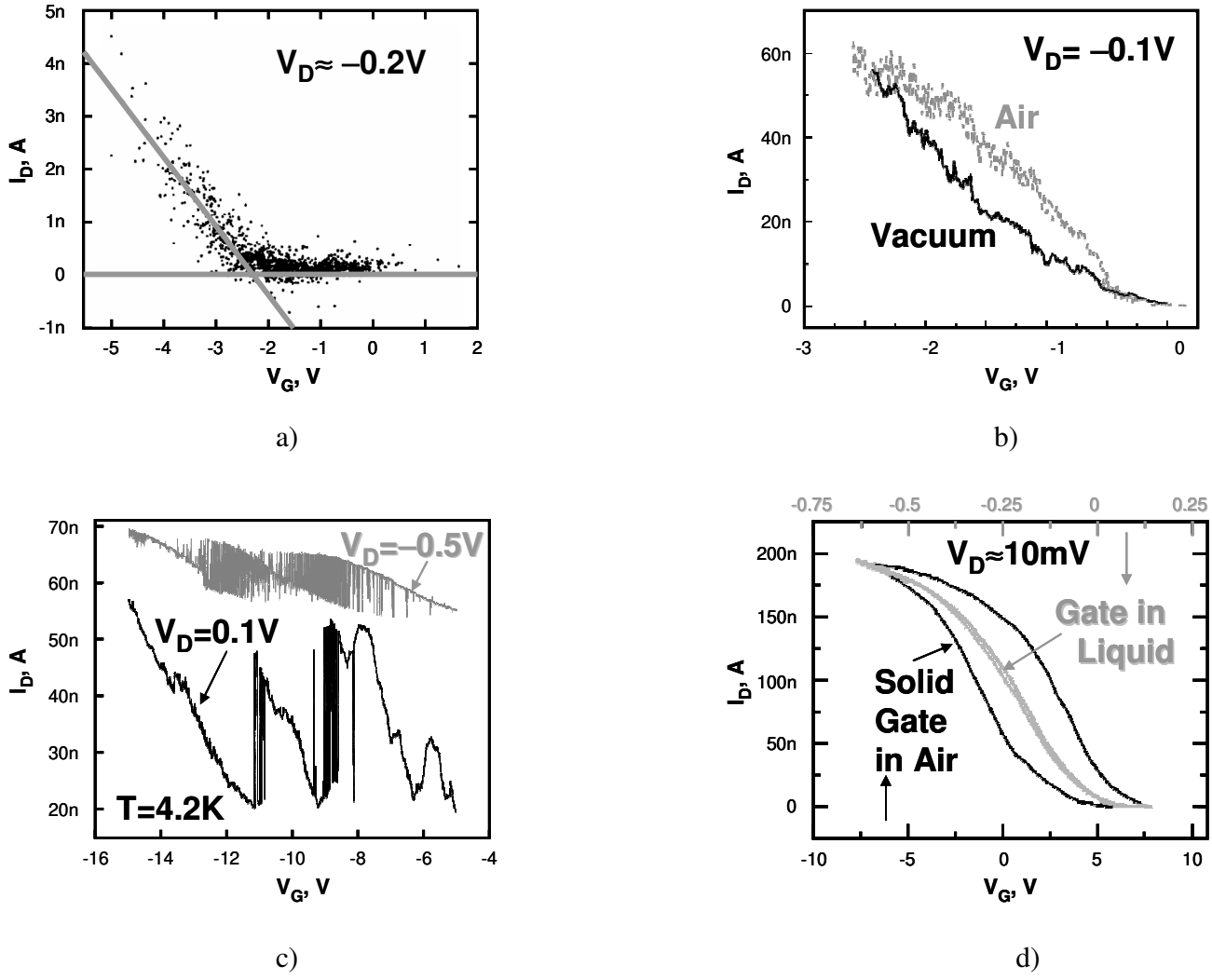
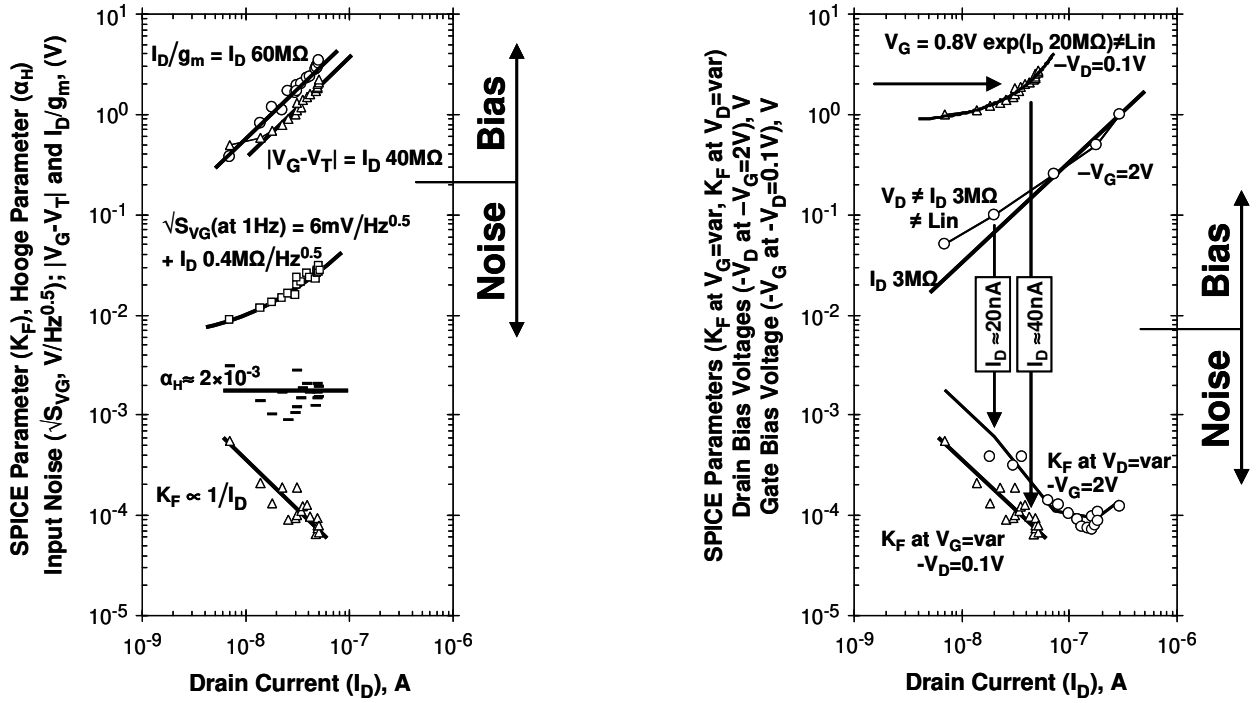


Figure 51. Spiky, stochastic transfer "I–V curves" of field-effect transistor structures based on carbon nanotubes (CNT FET). At given voltage, the current varies randomly around a value, owing to large noise in these devices, and no stable DC value is present in single CNT FET. a) The current and the threshold voltage (cross point of two lines) are analyzed in terms of stochastic resonance in [273]. Details for device and measurement are not reported. b) FET based on single CNT with diameter 1-3 nm, bridging distance 4μm between Ti/Au electrodes patterned on the top of CNT [274]. The currents in air are larger and scatter more than the currents measured when the device is in vacuum (18mTorr). The measurements are at room temperature. c) Similar FET with single CNT (diameter 1-3 nm and distance 4μm between Ti/Au electrodes) measured at cryogenic temperature in vacuum and opposite directions of the channel current [171]. Several (three) giant and bias dependent RTS are observed with amplitudes 30% to 60% of "DC" current. d) Smooth DC characteristics are measured at room temperature for a FET based on a random network of single-wall CNTs [275]. The gate voltage bias is with triangular waveform (0.5Hz). The Si substrate serves as solid-state gate, while electrochemical gate (reference and working electrodes) was used as liquid gate via solution with pH 7.4. The hysteresis is evident when the device was operating in air (200-500nm SiO_2 gate oxide), while it was small for liquid gate (gate insulator capacitance ~10-20μF/cm² was estimated from quasistatic CV measurement, 56nF gate capacitance, and the count of nanotubes in AFM images).



a) b)

Figure 52. Typical outcome from noise experiments with single CNT FET. Data are from [274] for measurements performed in vacuum (18mTorr) of device fabricated and with parameters, as described in [171]. Details and comments are given in the text. a) Outcome from experiment with variable gate bias at low -V_D=0.1V. From top to bottom: ratio g_m/I_D; transfer I–V curve $I_D \leftrightarrow |V_G-V_T|$; spectral density of input referred (gate) noise $S_{V_G}^{0.5}$ at 1Hz in unit V/Hz$^{0.5}$; Hooge parameter α_H, as obtained from SPICE parameter $K_F=fS_{I_D}/I_D^2$ in bottom. b) Deviations observed in experiments with variable gate bias and with variable drain bias, all carried out in linear mode ($|V_D|<|V_G-V_T|$-0.5V). From top to bottom: non-linear transfer I–V curve $I_D \leftrightarrow |V_G-V_T|$, same data as in sub-figure a); non-linear output I–V curve $I_D \leftrightarrow V_D$ with step at low $|V_D|$; variation of SPICE parameter K_F with drain bias at low V_D and constant V_G; Different value for K_F when changing V_G at constant V_D, same data as in sub-figure a); Arrows denote the same voltage bias condition {–V_G=2V,–V_D=0.1V}, showing different currents and noise levels in different experimental trials. None of the deviations in sub-figure b) is supported by field-effect transistor theory and models.

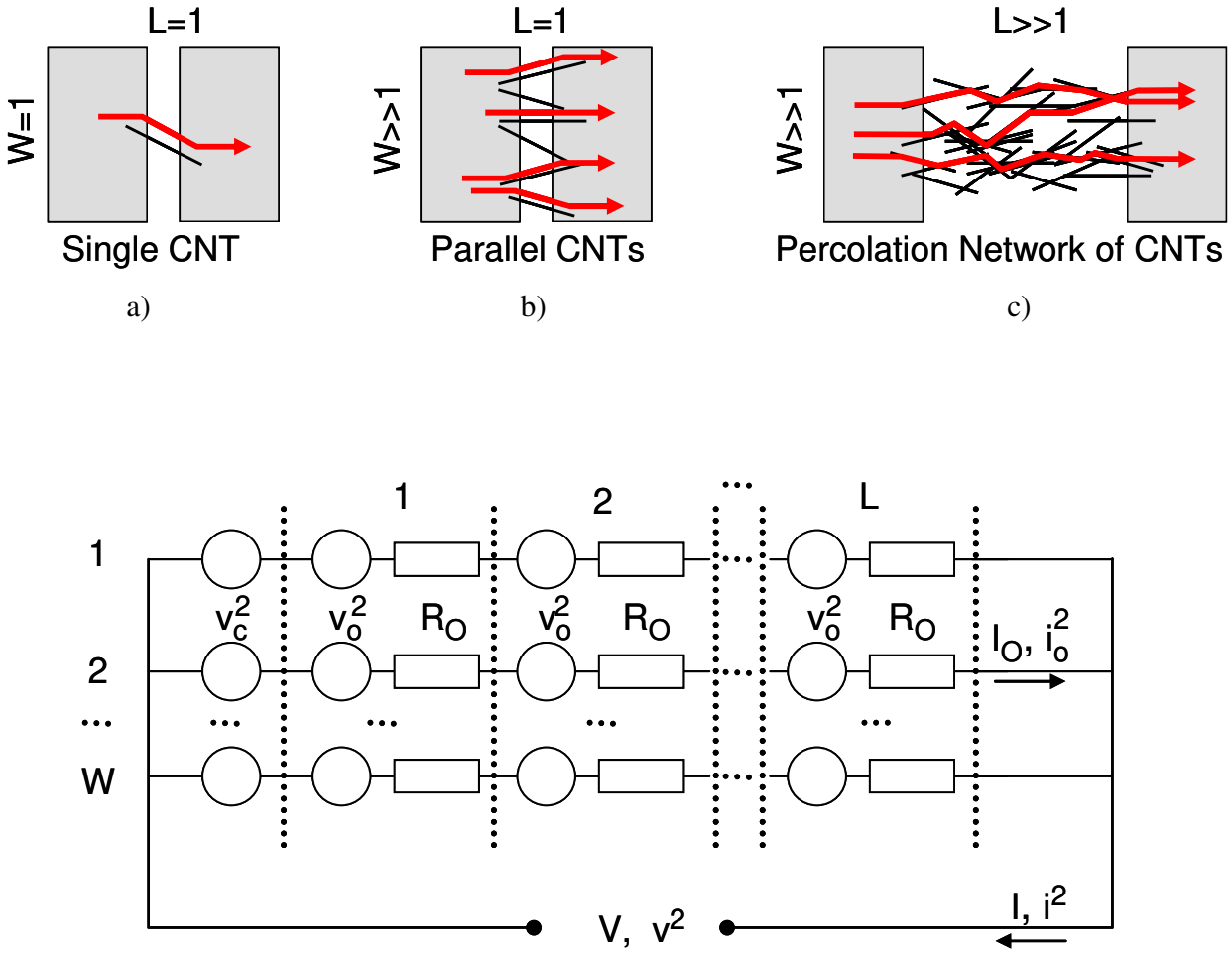


Figure 53. Typical arrangement of carbon nanotube devices and idealized model. a) Single CNT device. b) Short channel CNT device. c) Long channel CNT device with percolation network. d) Idealized resistor network, representing the percolation network in CNT devices. The arrows in a), b) and c) denote W-independent conduction branches in percolation network, which are shown with parallel branches in d). Each conduction branch in d) has L serially connected CNTs. Capital letters V, I in d) represent DC voltages and currents, and small letters v,i represent noise voltages and currents. The voltage noise sources v_o^2 are associated with every single CNT (with resistance R_O) in the network, while voltage noise sources v_c^2 are associated with contact between metal and CNT in a single conduction branch.

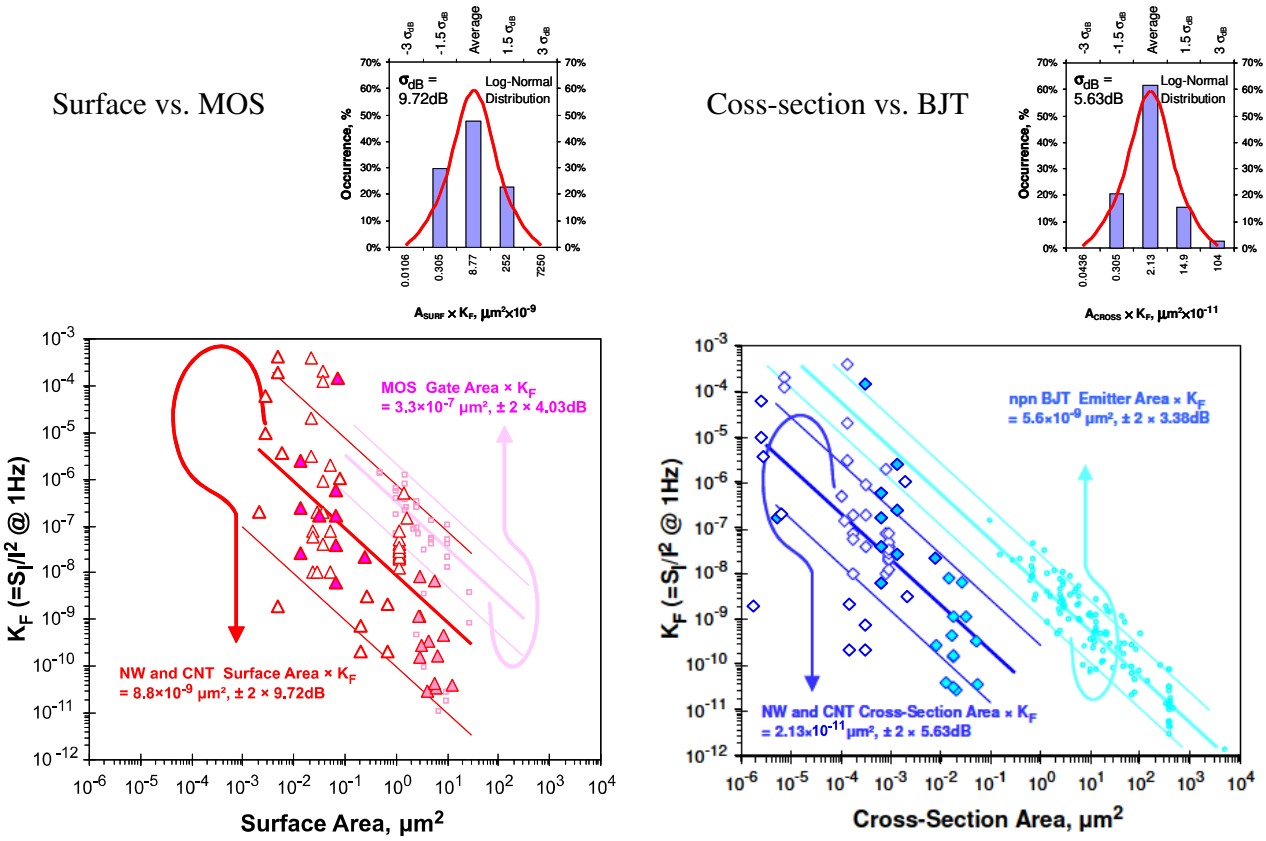
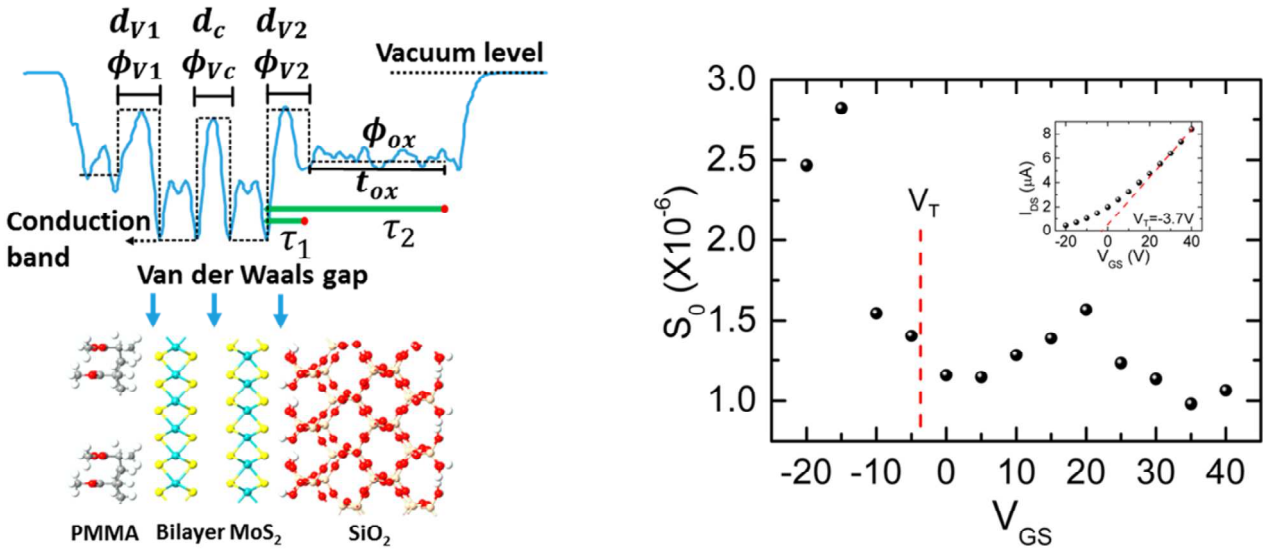


Figure 54. The geometric scaling rule for 1/f noise, the smaller is the area – the noisier the device is, applies to nanowire and nanotube devices too. a) Comparison of noise vs. surface area of nanowire (▲) and nanotube (△) devices to gate area of MOS transistors (□, data from [22, 48, 57, 79, 87, 89, 92, 208, 214]), inspecting surface number fluctuation. b) Comparison of noise vs. cross-section area of nanowire (◆) and nanotube (◇) devices to emitter area of bipolar transistors (○, data from Figure 2), inspecting injection noise due to weak contact. The data (▲,◆) for Si nanowire devices are from [271, 277, 278, 279, 280, 281]. The data for carbon nanotube devices (△,◇) are from [268, 272, 274, 275, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292], with details as discussed earlier in the text. The insets illustrate the distribution of the data around 1/area trend, in terms of log-normal distribution – see eqs. (83) and (84).



a) b)

Figure 55. Example from [293] for a 2D field-effect transistor with two atomic layers of MoS_2 semiconductor. (a) Energy diagram and spatial schematic diagram of the cross-section of the layers, showing energy barriers and spatial spacing due to van der Waals bonding. (b) Normalized noise referred to 1Hz, S_o=average(f×S_I(f)/I_{DS}^2), averaged over logarithmically spaced frequencies f in the range from 2Hz to 1000Hz, vs. gate bias voltage V_{GS}.

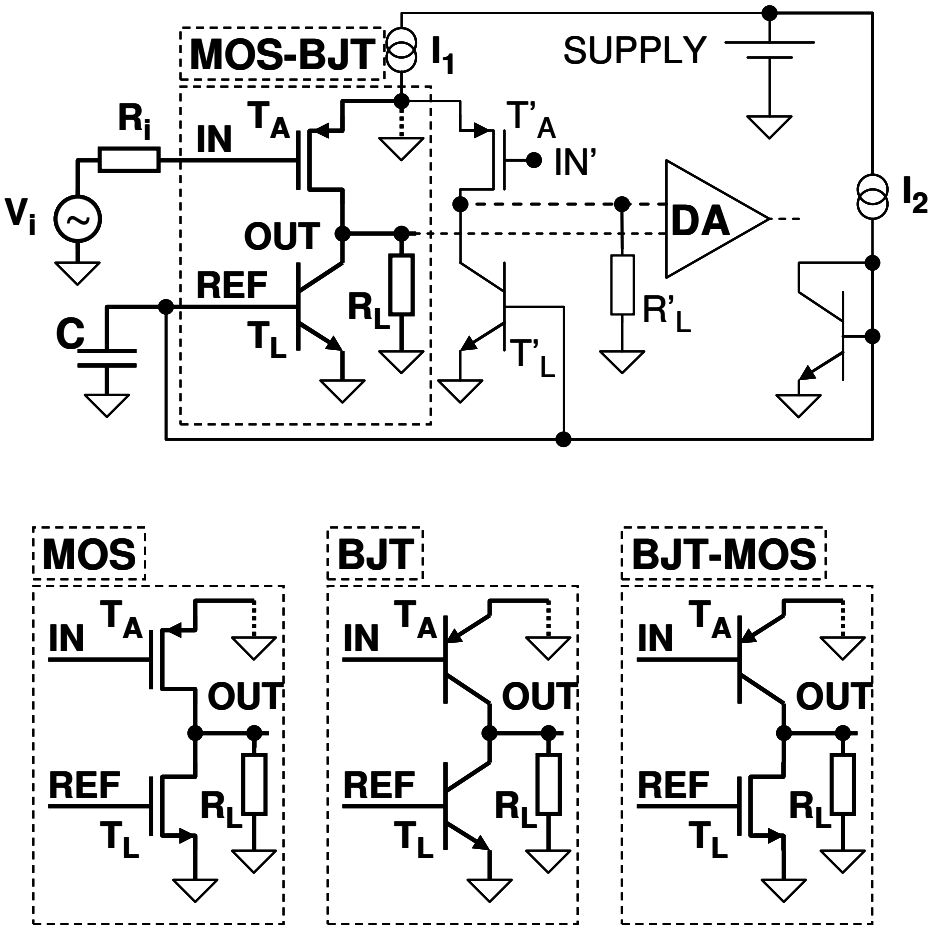
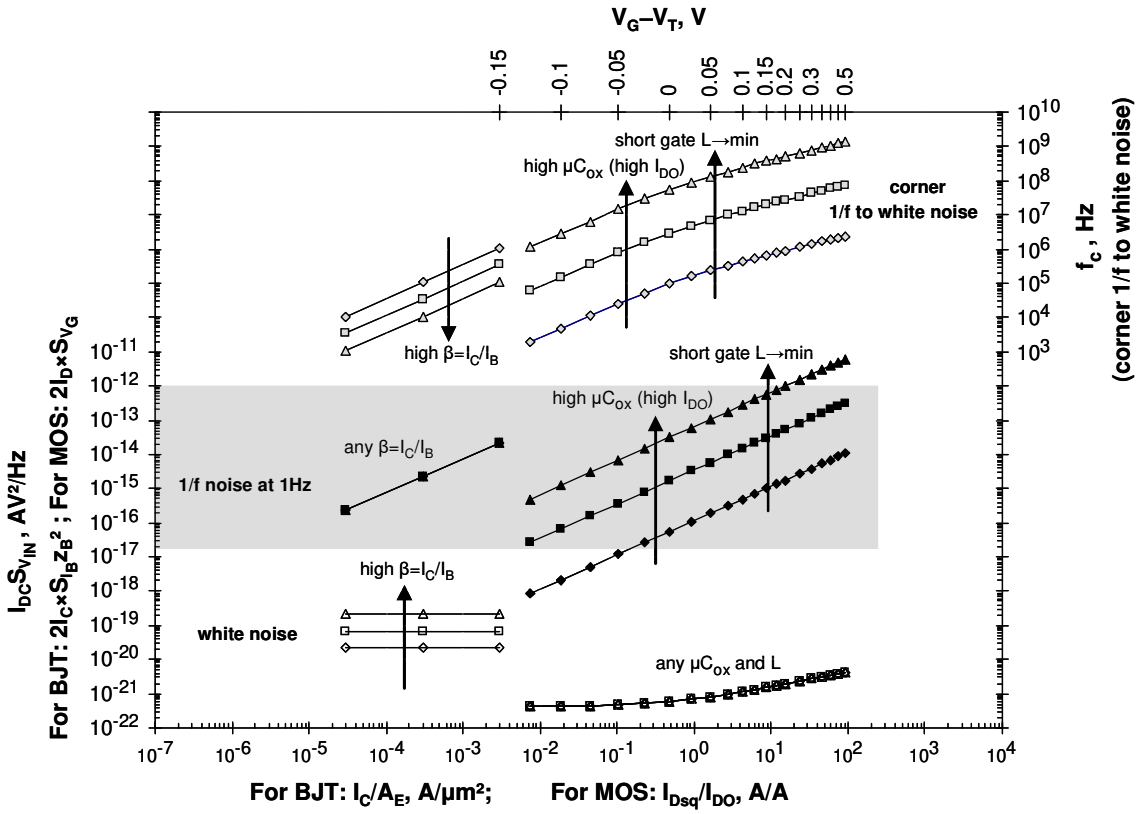


Figure 56. Typical configuration at the input of low-frequency amplifier. The amplification transistor T_A and the loading transistor T_L in the first differential stage are surrounded by a dashed line and can be MOS or BJT in BiCMOS technologies, as depicted in the three insets.



Legend		FOM for input 1/f voltage noise	Other parameters	corner f_c	1/f noise	white noise
BJT		FOM_{SVB}=3.8×10⁻¹² μm²V²/Hz at 1Hz	β=50	◇	◆	◇
			β=150	□	■	□
			β=500	△	▲	△
MOS		FOM_{SVG}=1.3×10⁻⁹ μm²V²/Hz at 1Hz	I_{DO}=0.1μA, L=1.5μm	◇	◆	◇
			I_{DO}=0.33μA, L=0.5μm	□	■	□
			I_{DO}=1μA, L=0.2μm	△	▲	△

Figure 57. Design window for noise in micropower low-frequency amplifiers, in which the product $I_{DC}S_{V_{IN}}$ of bias current I_{DC} and power spectrum density of noise $S_{V_{IN}}$ is desired to be low. The left-hand plots are for BJT amplifiers, and represent the input current noise by $S_{V_{IN}}=S_{I_{IN}}z_B^2$, where $z_B \sim \beta\varphi_t/I_C$ is the input resistance of the amplifier. The input voltage noise of the BJT amplifier is expected lower by a factor of $(r_b/z_B)^2$, where r_b is the contact resistance to the base of BJT. The right-hand plots are for MOS amplifiers, and represent the input referred noise voltage, because the input current noise is low at MOS gate, since $S_{I_{IN}}/S_{V_{IN}} \sim (2\pi f WLC_{ox})^2$ is low at low frequencies. The plots in the shaded area are for 1/f noise components. The plots for white noise components are below the shaded area. The plots above the shaded area are for the corner frequency f_c, at which the 1/f noise and the white noise have equal magnitudes. The symbols and the values of the parameters are explained in the legend. The ticks in the top axis are in steps of $50\text{mV} \sim 2\varphi_t$ for gate overdrive voltage (V_G-V_T), corresponding to the ratio I_{Dsq}/I_{DO} in the bottom horizontal axis for MOS transistors. The arrows illustrate the impact of several factors.

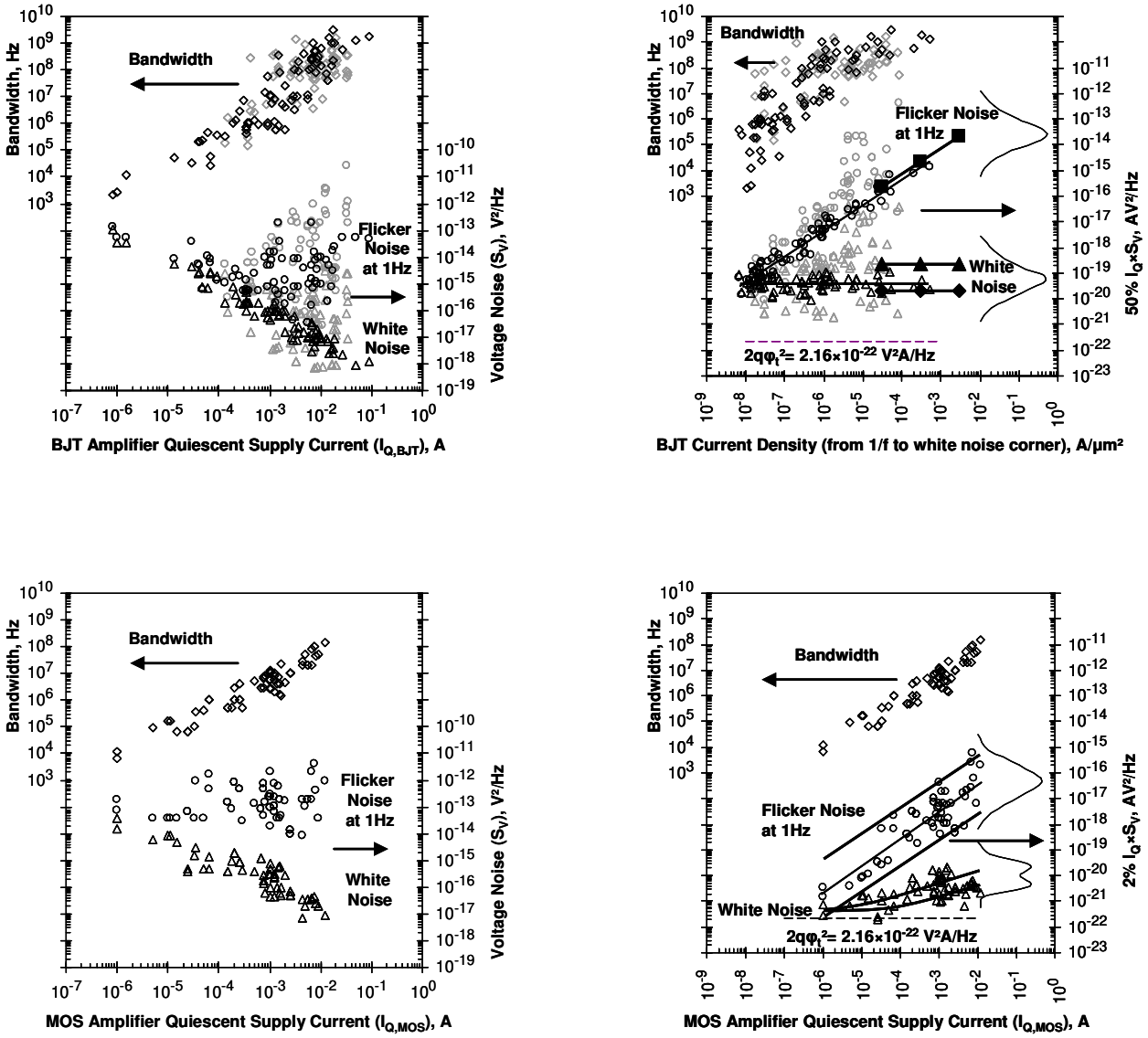
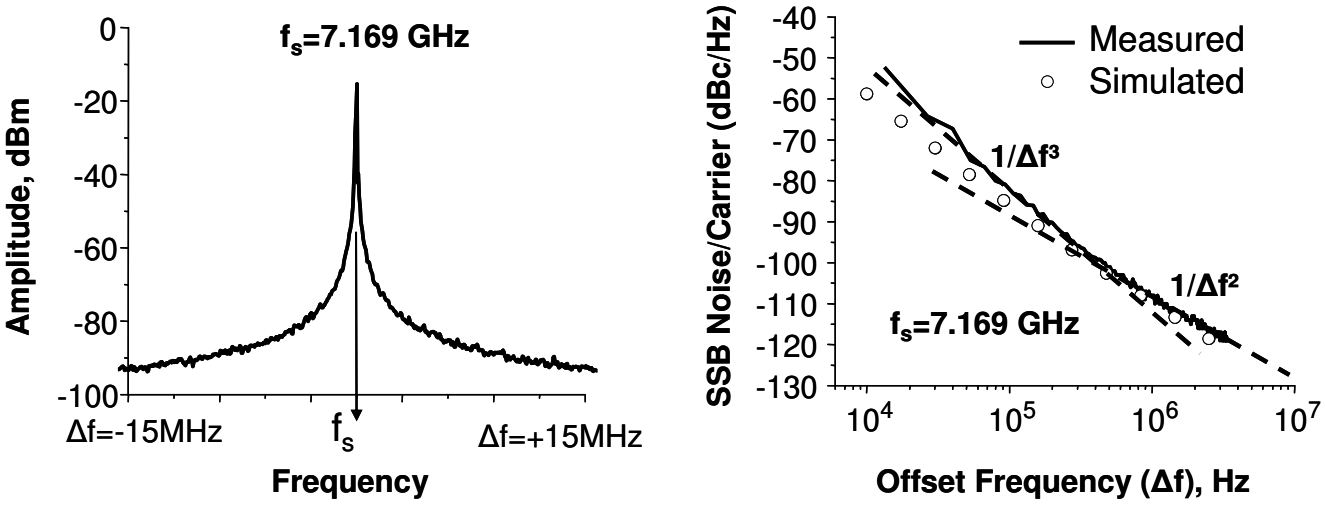


Figure 58. Bandwidth (◇), flicker noise at 1Hz (O) and white noise (△) in commercial operational amplifiers from several manufacturers (Linear Technology, Analog Devices, and Texas Instruments and Burr-Brown). Left-hand figures are data as given in datasheets. Upper figures are for 131 BJT amplifiers, in which the black symbols are for 81 amplifiers in regions, where the density of the data is higher. Bottom figures are for 67 MOS amplifiers. In right-hand figures, the noise level is multiplied by portion of amplifier quiescent current I_Q, 50%I_Q for BJT and 2%I_Q for MOS. The current density in BJT amplifiers is deduced from the corner frequency between 1/f and white noise, according to eq. (327), using typical values for β=150 and $FOM_{S_{VB}}$=3.8×10^{-12} µm²V²/Hz. The curves on right illustrate the distribution of the data around the trends. The trends are explained in the text and are in agreement with eq. (326) for BJT and eq. (336) for MOS amplifiers. The distribution for white noise in MOS amplifiers is bimodal, with I_{DO}(W/L)=2µA and 150nA.



(a) (b)

Figure 59. Typical spectrum at output a) and normalized single-side band noise b) of a CMOS oscillator. The dash-lines in b) denote the components in SSB noise with slopes Δf^3 and Δf^2. Data are for 7.2GHz CMOS voltage controlled oscillator [297].

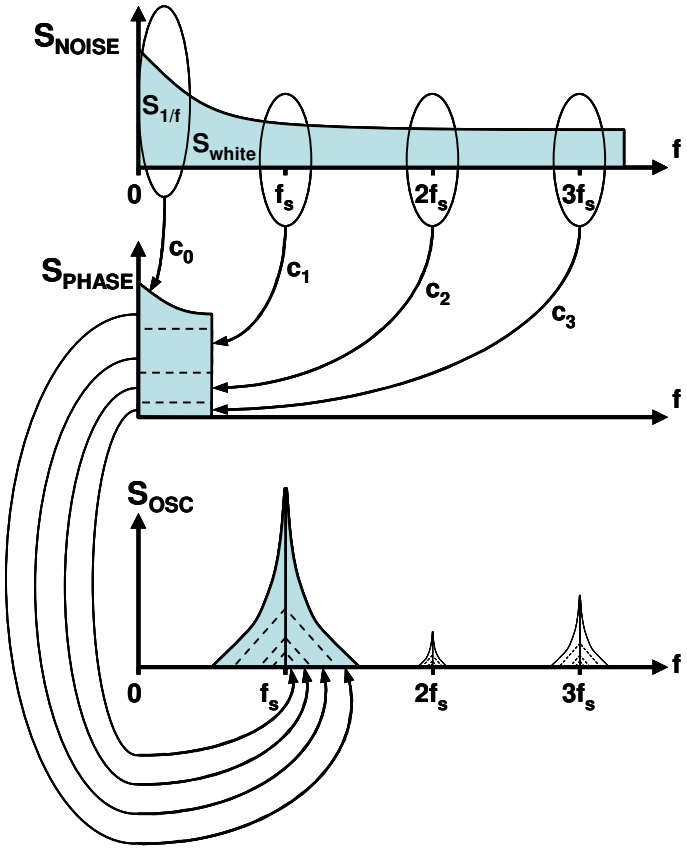


Figure 60. Conversion of circuit (nodal) noise (S_{NOISE}) to phase fluctuations (S_{PHASE}) and phase-noise sidebands (S_{OSC}) around the fundamental harmonic (f_s; shaded area) and higher order harmonics ($2f_s$, $3f_s$,…, kf_s, blank areas) of the oscillator. The conversion is conceptually explained according to [315], but the conversion process is valid for other phase noise theories, that use harmonic balance and perturbation. Note that the low-frequency noise contributes only a portion to the phase noise.

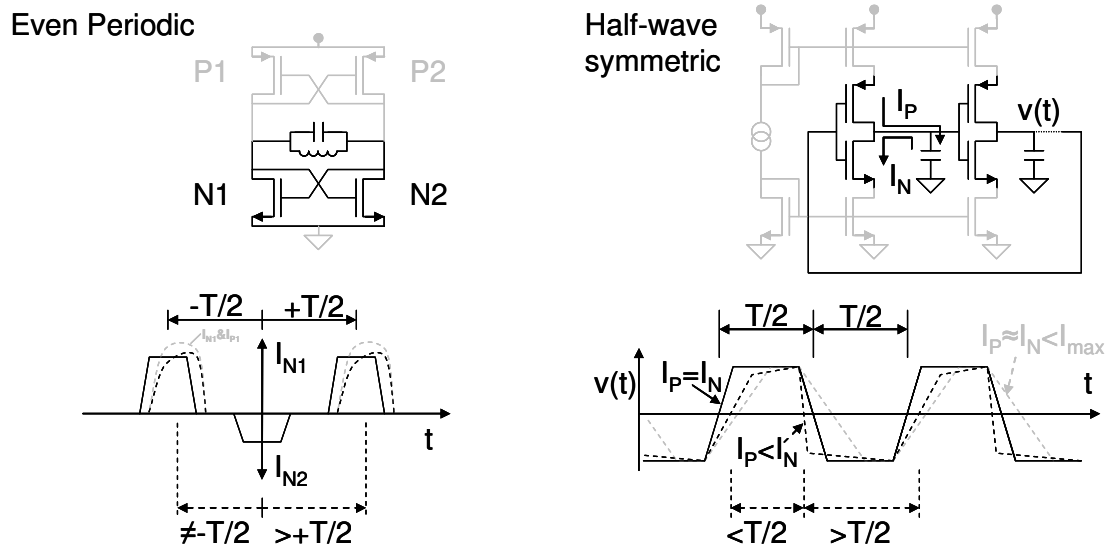


Figure 61. Examples for difficulties of obtaining waveforms with zero value for the DC component ISF_{dc} of impulse sensitivity function (ISF). Left-hand figures are for LC oscillator, in which the current from transistors is "tilted", and this does not allow to obtain perfect even periodic symmetry $v(t)=v(-t\pm nT)$ in LC oscillator signal. Right-hand figure is for ring oscillator, in which the currents from pMOS and nMOS transistors are not equal in magnitude, resulting in different rise and fall times in the waveform, which does not allow to obtain perfect half-wave symmetry $v(t)=-v(t+T/2\pm nT)$ in ring oscillator signal. The ideal signals are shown with solid lines, the distorted waveforms are shown with black dash-lines, and gray lines illustrate improvements that can be achieved by using the additional circuitry, the later also given by gray color in upper figures.

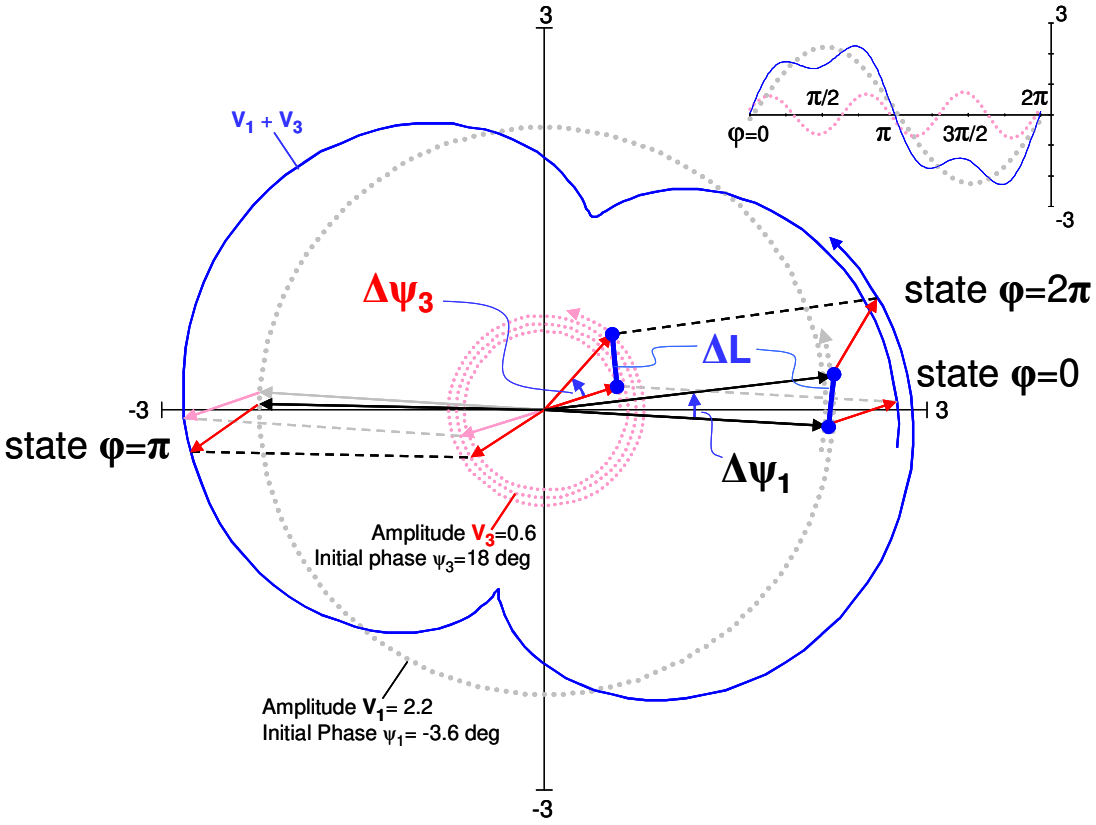
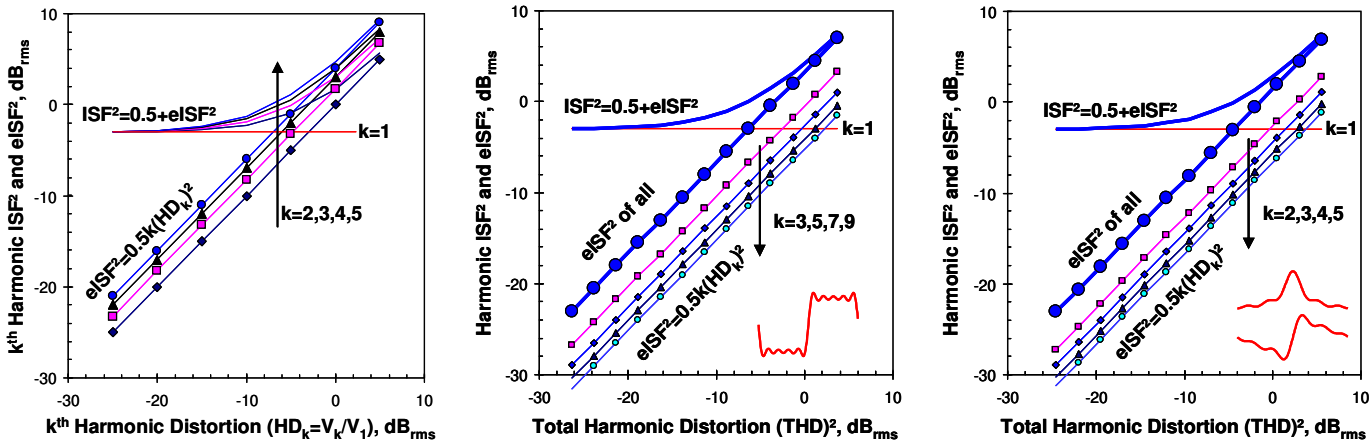
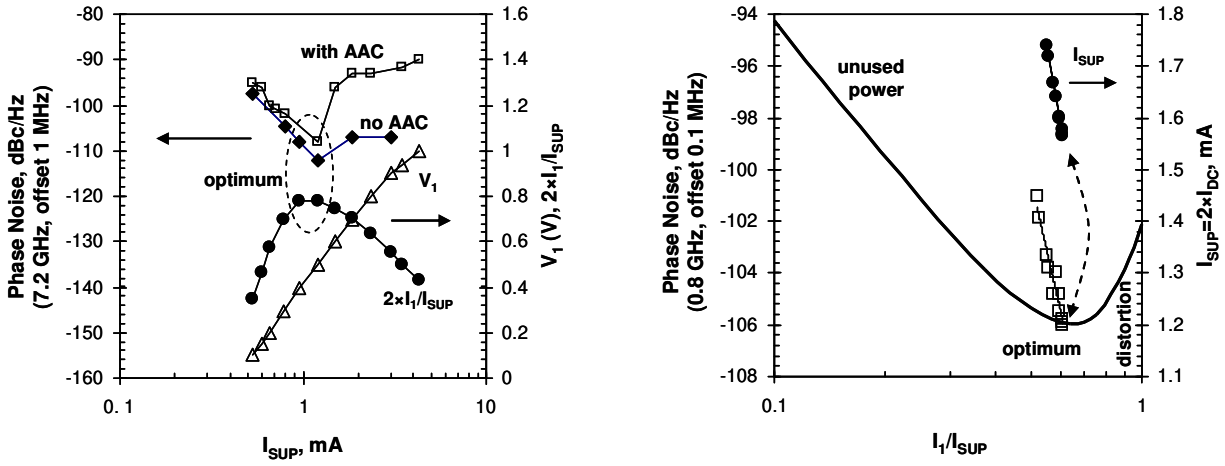


Figure 62. Contours of third harmonic (k=3, inner spiral, V_3, with 3 turns) and fundamental harmonic (p=1, outer spiral, V_1, with 1 turn) of an oscillating signal with limit cycle (cycloid, V_1+V_3) and distorted waveform (inset). The contours are drawn open to illustrate that for one period of oscillation, V_1 makes p=1 turn and V_3 makes k=3 turns. If no phase noise is present, then the vectors at state $\varphi=0$ and $\varphi=2\pi$ would overlap. The RF noise, uncorrelated at harmonic frequencies f_s and $3f_s$, diverges independently the harmonics from their ideal phase, as shown for states $\varphi=\pi$ and $\varphi=2\pi$. The accumulated phase deviation $\Delta\psi_3$ in k=3rd harmonic changes the state $\varphi=2\pi$ of the oscillation contour (V_1+V_3), "pulling" V_1 to that state, and inducing a phase deviation $\Delta\psi_1$ in the fundamental p=1 harmonic. Vice versa, the accumulated phase deviation in V_1 will "push" the phase of V_3. In the limit of small deviations, both V_3 and V_1 are displaced at the same distance ΔL along their contours, thus, $V_1\sin(\Delta\psi_1)\approx\Delta L\approx V_3\sin(\Delta\psi_3)$.



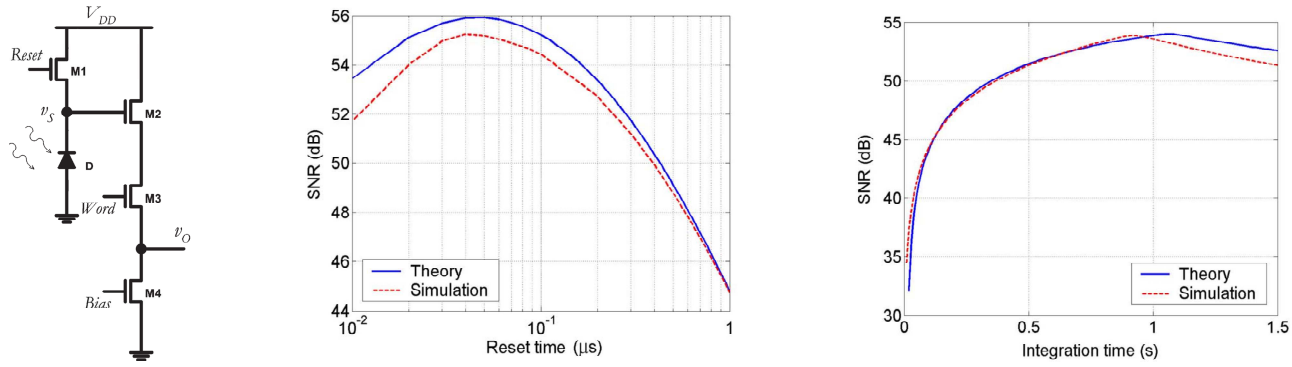
a) b) c)

Figure 63. Relative increase of phase noise in presence of harmonics. a) Higher order of harmonics cause higher increase in eISF at same harmonic distortion. b) Case of "rectangular" waveform distortion with odd harmonics only, amplitudes $V_{2k+1} \propto 1/(2k+1)$. c) Case of "peaking" or "sawtooth" waveform distortions with harmonic amplitudes $V_k \propto 1/k$. The waveform of the distortions are shown in the insets.



a) b)

Figure 64. Minimum phase noise at maximum ratio I_1/I_{SUP} of fundamental current harmonic I_1 to supply current ($I_{SUP}=2I_{DC}$ for oscillators with cross-coupled transistor pairs). a) CMOS oscillator from [297]. b) BJT oscillator from [323]. Note in both cases the phase noise is at minimum when the ratio I_1/I_{SUP} is at maximum, but the oscillation voltage (triangles in a) for V_1) is not necessarily at maximum. The solid curve in b) is when assuming "rectangular" approximation for the distortion in transistor current.



a) b) c)

Figure 65. Typical circuit of an APS (a), and variations of the SNR with reset time (b) and integration time (c), as reported in [349]

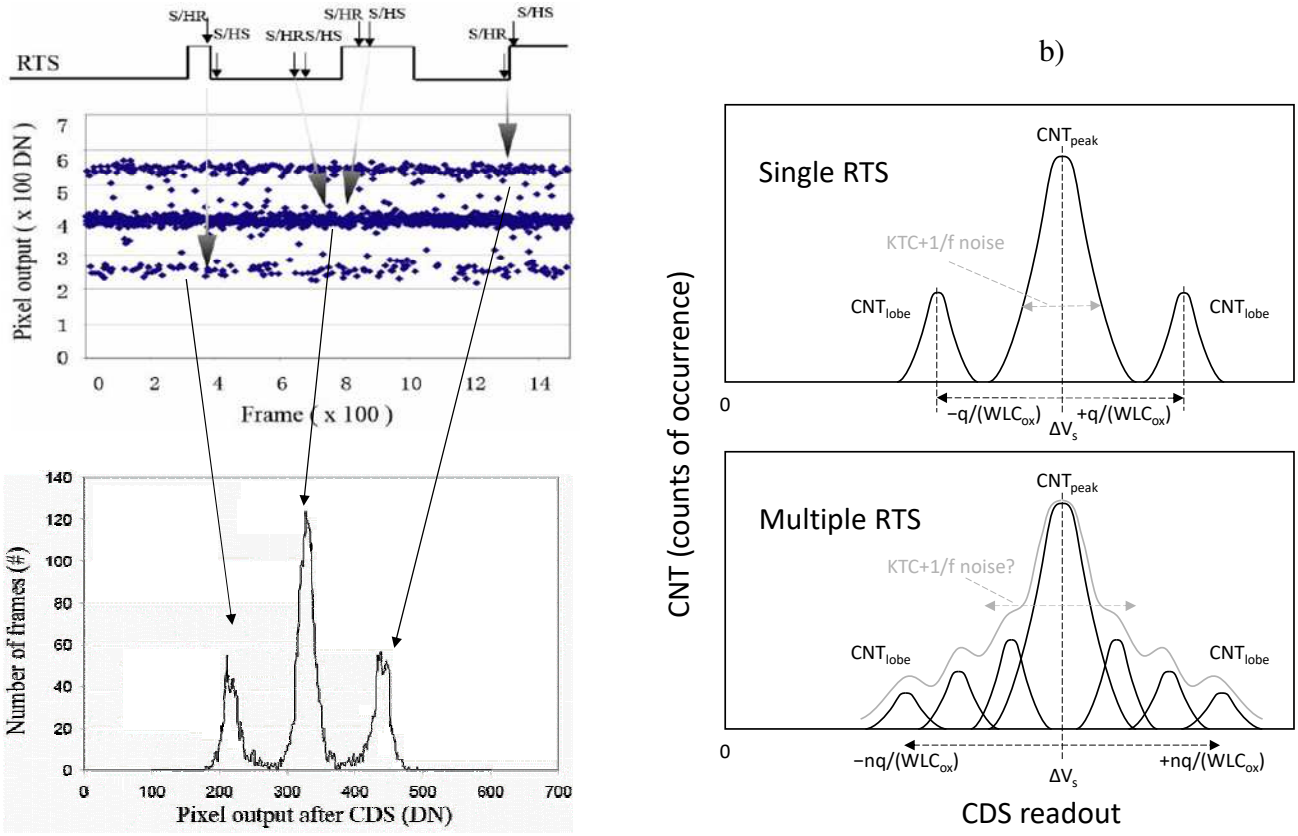


Figure 66. Histograms of the RTS inn APS. (a) experimental results for an APS with single RTS [351]. "Digital Number (DN)" denotes ΔV_s after ADC and "Frame" denotes the number of CDS acquisitions (CNT). (b) heuristic explanation of the histogram of single RTS noise in APS, showing the $q/(WLC_{ox})$ spacing between the peaks in the histogram and indicating the peak widening due to KTC and 1/f noise. (c) same as (b) for multiple RTS, indicating the reduction of the $q/(WLC_{ox})$ spacing in APS with larger-area buffering transistor (M2 in Figure 65 (a)), which causes overlap between the peaks in the histogram that eventually leads to inability to distinguish the peaks in the histogram envelope (gray curve), and superimpose and attribute the individual RTS noise into 1/f noise.

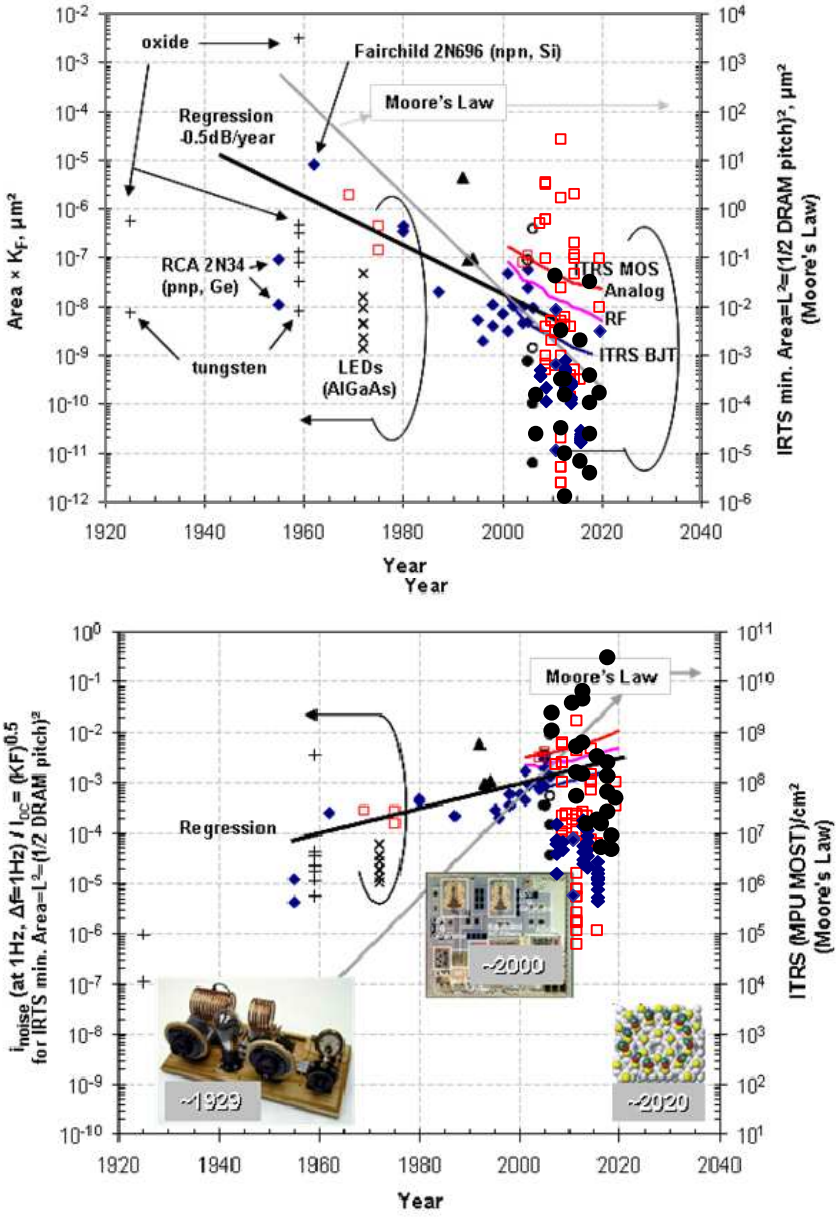


Figure 67. Evolution of a) the factor (Area×K_F), see eq. (2), and b) relative RMS noise i_{noise}/I_{DC} (at 1Hz for bandwidth of 1Hz) over almost of a century of research on 1/f noise. Data are: for vacuum tubes (+) from [359, 364]; for Si, Ge and SiGe bipolar transistors (◆) data are aggregated from Figure 2, Figure 14 and Figure 40, and from [365, 366, 367] for the period before year 1980; by using surface area of nanowire and CNT devices (○), and for MOS transistors (□), the data are aggregated from Figure 54a, and from [360, 362] for the period before year 1980 for MOS transistors; by using cross-section area of nanowire and CNT devices (●), the data are aggregated from Figure 54b; data are from [368, 369] for III-V semiconductor HBTs (▲); data are from [370] for optical noise of light-emitting diodes (×). See the text for the regression lines. The prediction lines (top to bottom) for MOS-Analog, MOS-RF and BJT (left-hand axes) and the Moore's law (right-hand axes) are from ITRS for the period 2006-2020 [3].

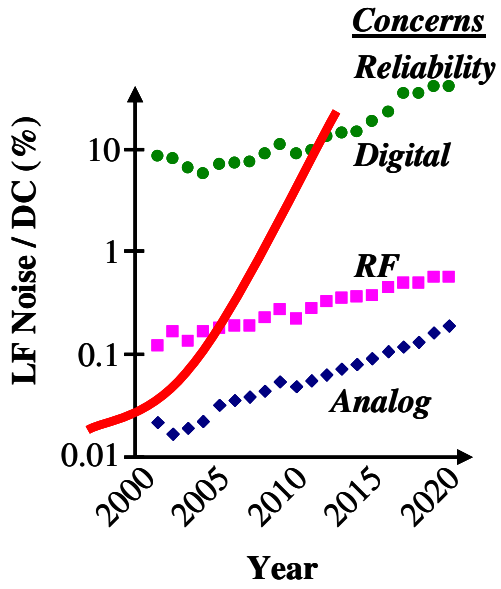


Table 4. Typical assumptions for 1/f noise in Analog, RF and Digital applications of MOS transistors

Application	Size W×L	Bandwidth f_{max}/f_{min}	Confidence Max to RMS ratio
Digital ●	$3\times1\ L_{min}^2$	10^{12} =1/error rate	4 (4σ~99.994%)
RF ■	$1000\times1\ L_{min}^2$	1MHz/1Hz	3 (3σ~99.7%)
Analog ◆	$500\times20\ L_{min}^2$	100kHz/1Hz	2 (2σ~95%)

Figure 68. Concerns for the application of MOS transistors related to the increase of 1/f noise by the device down-scaling [29]. When the thick line crosses and is above the data, the concern with 1/f noise becomes important for the particular application. In the vertical axis, (LF Noise) denotes peak-to-peak values that would be present at the conditions listed in Table 4. The data are based on predictions in ITRS [3] for the figure of merit (Area×K_F) – see the text for further details.

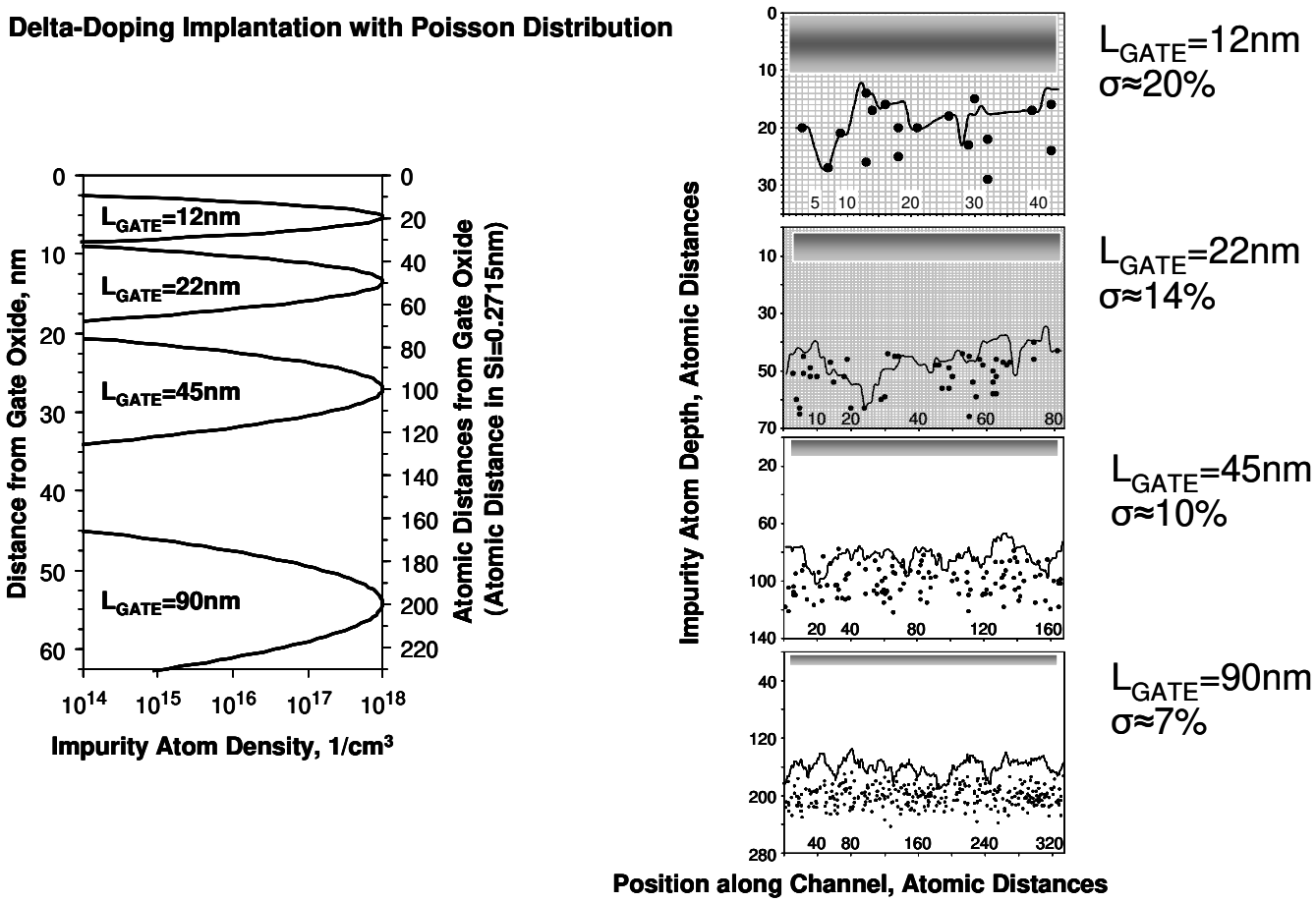
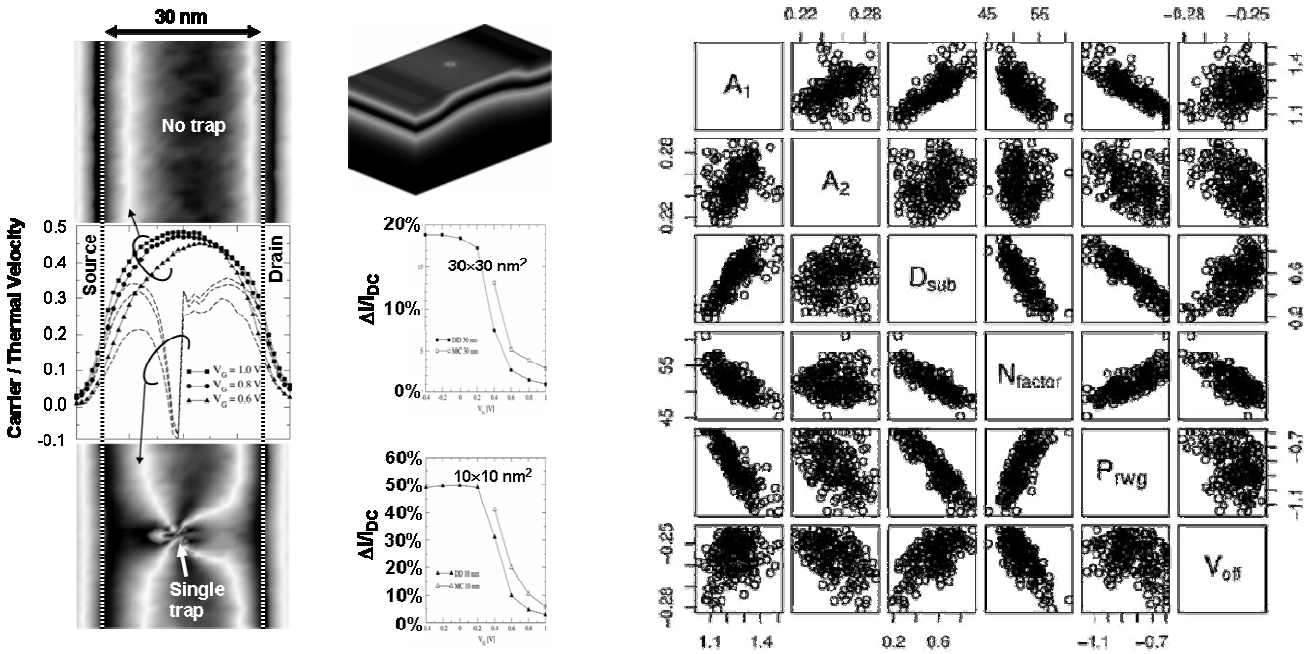


Figure 69. Delta doping as the best choice to reduce DIBL with minimum threshold voltage shift in MOS transistors. The profiles on left are calculated assuming Poisson distribution during ion implantation. On right, the corresponding location of impurity atoms are given, decreasing the gate length from 90nm to 12nm (from bottom to top), and showing with lines the effective depth of the depletion capacitance, and with gradually-shaded areas the region, where the current transport in the MOS transistors occurs. The grid represents the atomic distance in silicon, and it is visible in the smallest device (L_{GATE}=12nm).



a) b) c)

Figure 70. Typical outcome from "atomistic" and Monte-Carlo simulations of MOS transistors. a) Carrier velocity in 30nm×30nm bulk MOS transistor without and with one trap in the channel [388, 389]. b) Structure simulated in a), and increase of the relative difference in the drain current for the two cases (with and without trap), when decreasing the transistor sizes from 30nm×30nm to 10nm×10nm. c) Scatter plots for several BSIMSOI parameters for 200 devices of Ultra-Thin-Body SOI nMOS transistors with 10 nm channel length [372]. The BSIMSOI parameters shown, are: A_1 and A_2 are non-saturation factors for saturation regime; D_{sub} is the channel length factor for drain-induced barrier-lowering (DIBL) effects; N_{factor} is an ideality factor, used to improve BSIMSOI in sub-threshold regime; P_{rwg} characterizes the gate-bias effect on access resistance, and reflects access resistance variation caused by the body thickness; V_{off} is the offset in sub-threshold regime from the threshold voltage in strong inversion.

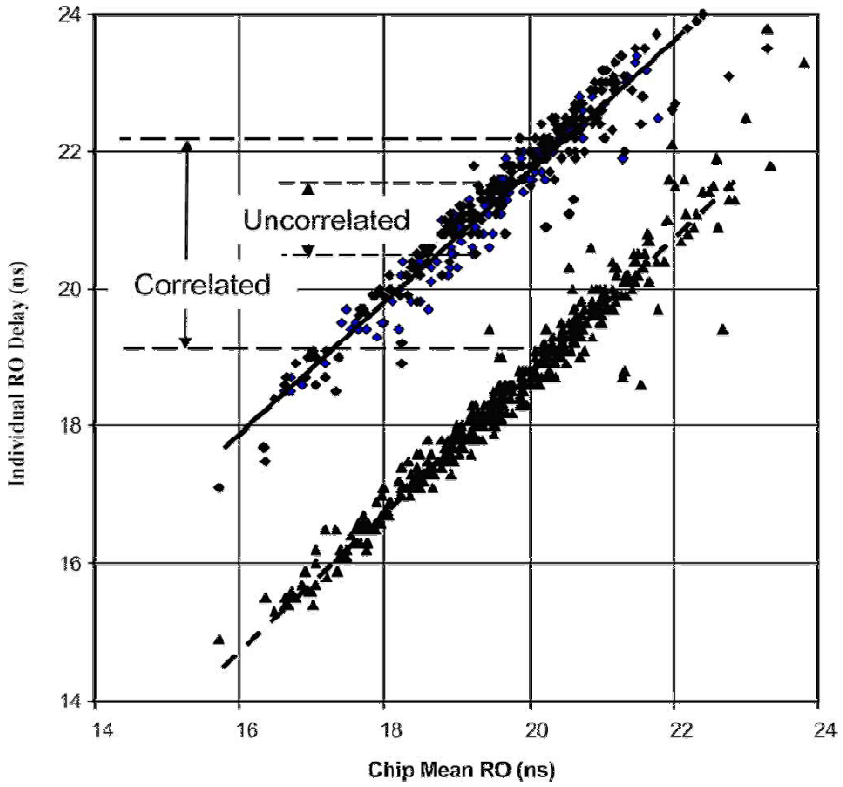


Figure 71. Variations of maximum and minimum ring delays of individual ring oscillators found in groups of ring oscillators vs. the average ring delay in each group [387]. The correlation is between individual and average – see the text.

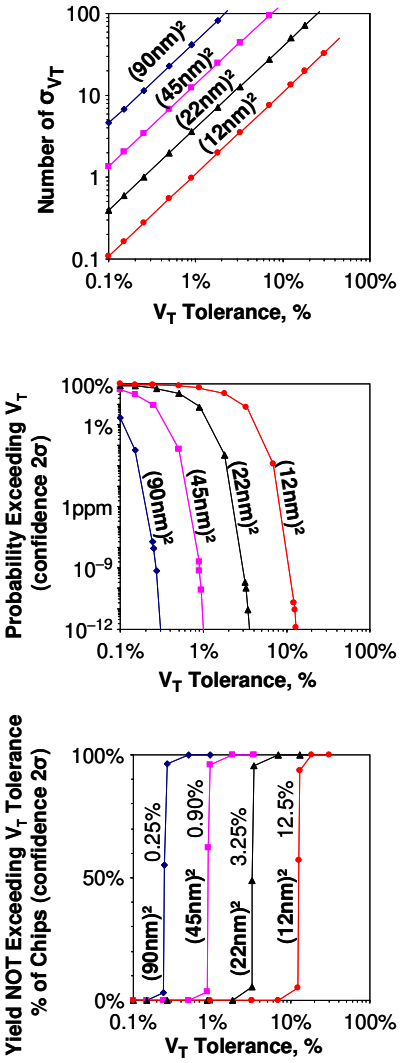


Figure 72. Impact of Poisson distribution in ion implantation for δ-doping on variability of the threshold voltage of MOS transistors. a) ratio of standard deviation σ_{V_T} in V_T to given tolerances of V_T (x–axes). b) Probability exceeding the tolerance for V_T. c) Yield of 6–T SRAM chips at given limit for V_T tolerance. The assumed sizes of SRAM are: 1MByte for cells made with transistors WL=90×90nm²; 10MByte for WL=45×45nm²; 100MByte for WL=22×22nm²; and 1GByte for WL=12×12nm². The percentages in the last figure are the minimum tolerances for V_T, which should be allowed in the SRAM design in order to achieve chip yield higher than 50% by Poisson distribution in ion implantation for δ–doping, and 2σ standard deviations are considered in the calculations for b) and c) to achieve confidence 95% in the calculation.

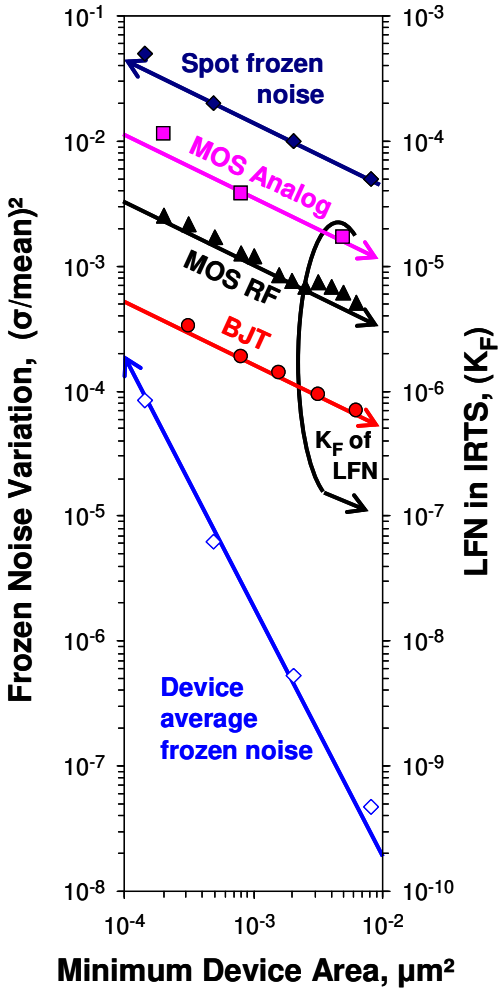


Figure 73. Comparison between "frozen noise", eqs. (443) and (444), and low-frequency noise from ITRS [3]. The "Spot frozen noise" is as obtained from Poisson distribution for the depth of ion implanted δ-doping – see eq. (443). The "Device average frozen noise" is as obtained for variation of V_T, and its variance is the variance of the (Spot frozen noise)² divided on (Number of impurity atoms), $[(\sigma_{dep}/t_{dep})_{avg}]^2=(\sigma_{dep}/t_{dep})^2/N_{imp}$, – see after eq. (444). The low-frequency noise is as projected by ITRS, and it corresponds to Figure 67. Evidently, the low-frequency noise and the "Spot frozen noise" have the same dependence on device area, namely, $(\sigma/\text{mean})^2 \propto 1/(\text{Area})$.

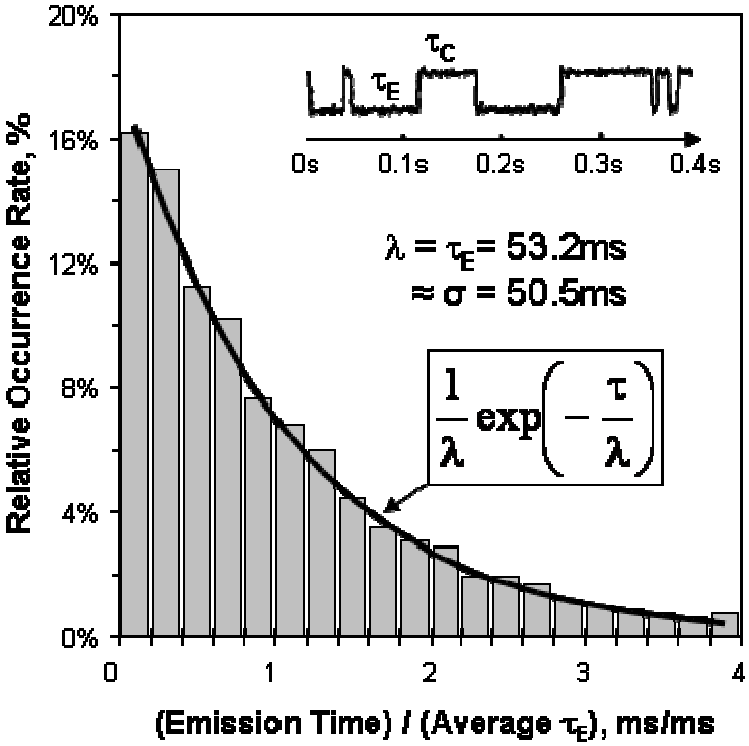


Figure 74. Exponential distribution of emission time, τ_E, experimentally observed for RTS noise in MOS transistor [68].

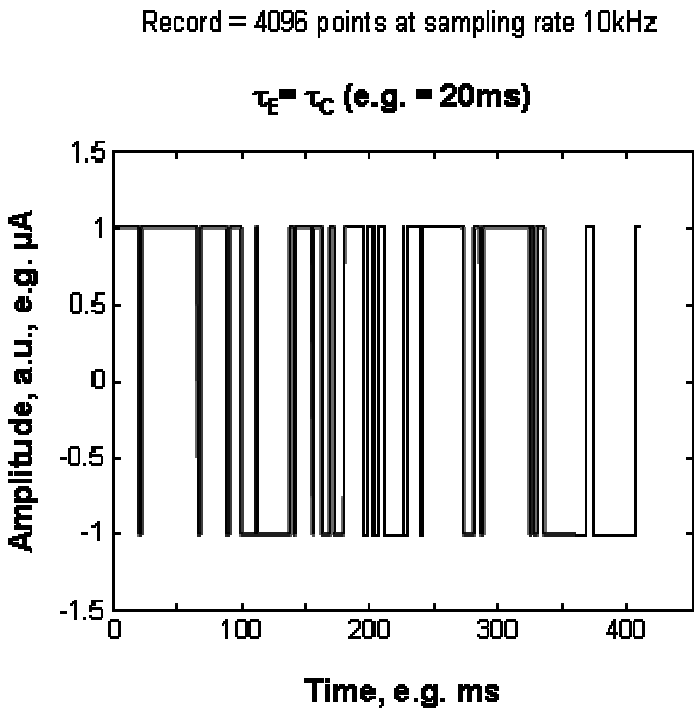


Figure 75. Sample of RTS with time constants $\tau_E=\tau_C$=20ms. The record length is 4096 points at sampling frequency 10kHz, and 2048 records were generated in order to evaluate averaging techniques for noise spectra – see also the next two figures.

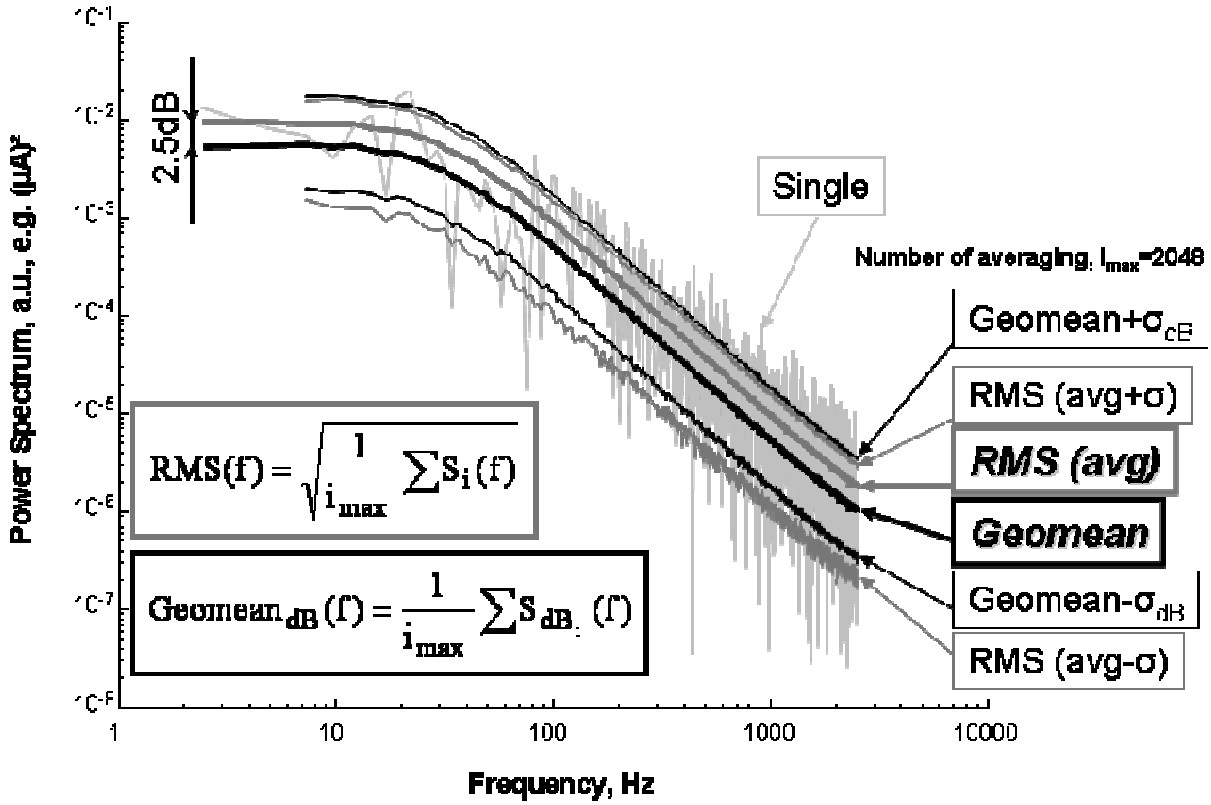
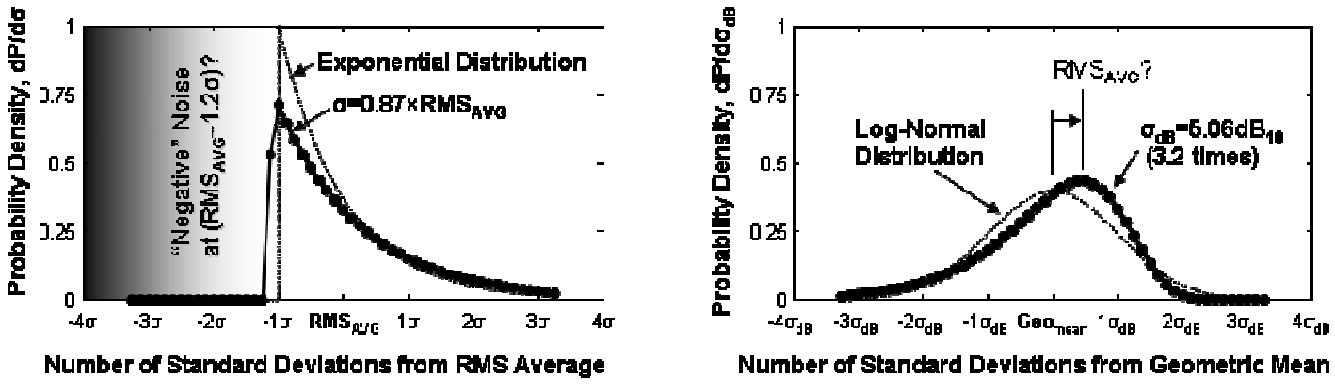
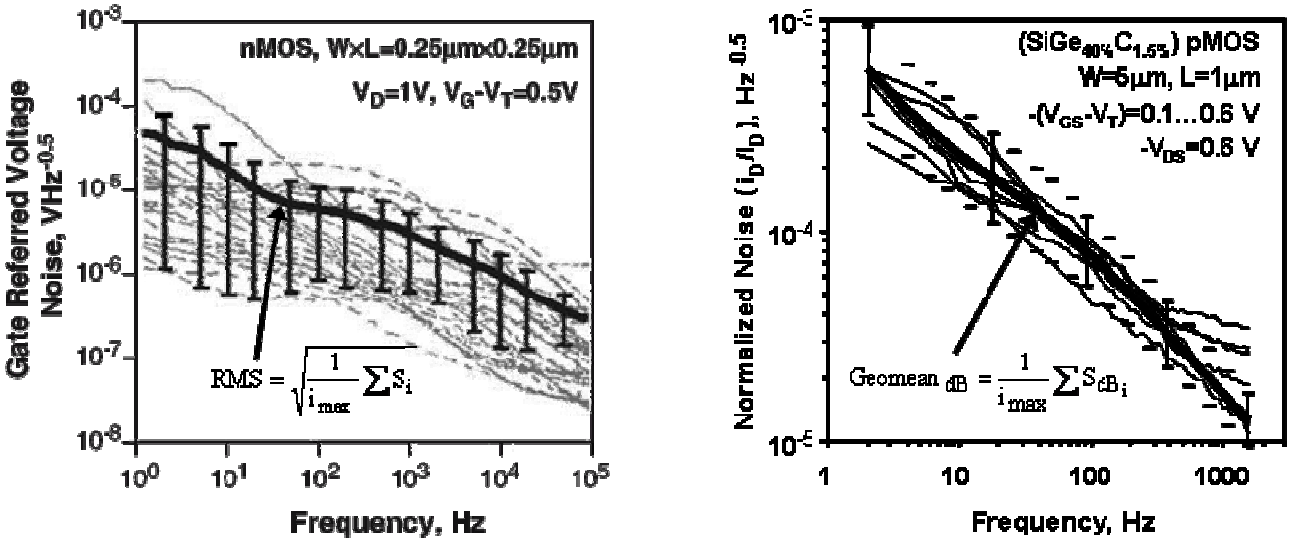


Figure 76. Spectra of RTS noise. Geometric averaging (Geomean) underestimates the root-mean-square average (RMS) with 2.5dB. Note that the upper limits RMS(avg+σ) and (Geomean+σdB) are similar, Geomean is well centered between its limits, whereas the lower limit RMS(avg−σ) is unrealistically low.



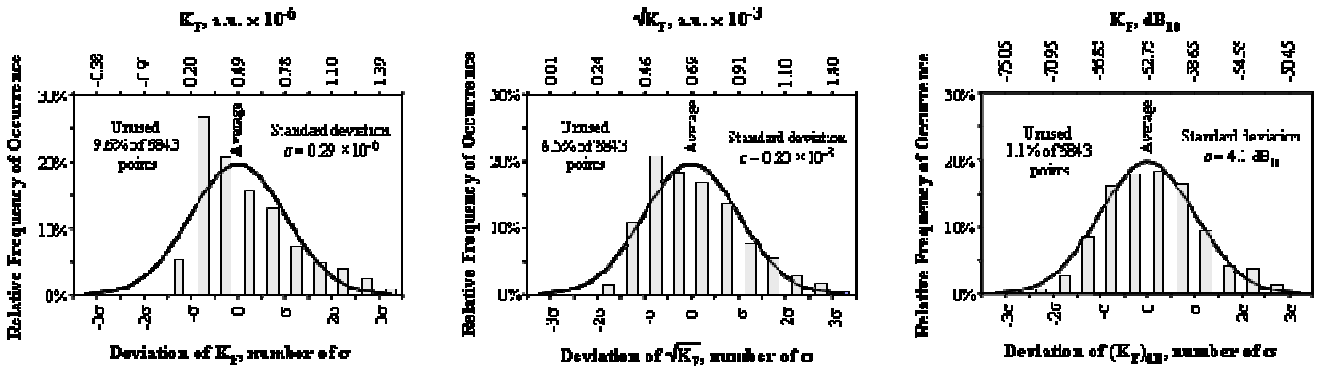
a) b)

Figure 77. Distributions of power spectrum amplitudes, normalized to standard deviations of the distributions. a) After root-mean-square (RMS) averaging (x-axis is linear), compared to exponential distribution. b) After geometric averaging (x-axis is logarithmic), compared to log-normal distribution.



a) b)

Figure 78. Behaviors of RMS averaging a) and geometric averaging b), when the scattering in the noise spectra is large [28]. RMS average is pulled up above where data usually are. The geometric mean is well in the middle where the majority of the data points are. The error bars correspond to one standard deviation.



Noise Power | Noise Amplitude | Logarithm of Magnitude

a) b) c)

Figure 79. Evaluation of the distribution of noise magnitudes in terms of SPICE parameter K_F=S(f)×f, by means of a) Noise Power ∝ K_F, b) Noise Amplitude ∝ $(K_F)^{0.5}$, and c) as logarithm of K_F. The distribution of Noise Power K_F tends to exponential distribution, log(K_F) tends to normal distribution, and the distribution of the Noise Amplitude $(K_F)^{0.5}$ is between these cases. The histograms are from [79] and are based on the same 6843 data points from a measurement of $SiGe_{40\%}C_{1.5}$% pMOS transistor (W=5µm, L=1µm) at different biases. "Unused points" are those data points, which deviate more than ±3σ from particular average, and were removed from histogram calculation. The solid lines represent normal distributions.