Thin Solid Films 613 (2016) 2–5

Contents lists available at ScienceDirect

Thin Solid Films

journal homepage: www.elsevier.com/locate/tsf

Random telegraph noise: The key to single defect studies in nano-devices

E. Simoen [a,b,*], W. Fang [a,c,d], M. Aoulaiche [a,1], J. Luo [d], C. Zhao [d], C. Claeys [a,d]

[a] Imec, Kapeldreef 75, B-3001 Leuven, Belgium
[b] Dept. of Solid-St. Physics, Ghent University, Gent, Belgium
[c] Key Laboratory of Microelectronic Devices & Integrated Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China
[d] E.E. Dept., KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 26 May 2015
Received in revised form 25 August 2015
Accepted 25 August 2015
Available online 29 August 2015

Keywords:
RTS
Generation-recombination centers
Low-frequency noise
Vertical polysilicon transistor

ABSTRACT

A review is given of the different methods to extract the main parameters from a Random Telegraph Signal (RTS) occurring in the channel current of small-area Metal-Oxide-Semiconductor Field-Effect Transistors, namely, its amplitude and its average up and down time constants. The advantages of using a so-called colored Time Lag Plot will be illustrated, enabling the detection of single defects in semiconductor materials and devices with high sensitivity. It will finally be shown that a detailed modeling of the RTS amplitude in vertical polycrystalline silicon transistors can yield the position of the trap in the channel with high accuracy.

1. Introduction

The charge transport in nano-scale devices is characterized by specific mechanisms, related to the quantum confinement in such structures. At the same time, the temporal fluctuations in the current through small devices, better known as noise, becomes dominated by only a few – sometimes one – dominant fluctuators. In the case that carrier trapping/detrapping is at the origin of the low-frequency noise, so-called Random Telegraph Signals (RTSs) can appear in the time domain. In the most simple form, the current switches between a high state (carrier capture) and a low state (carrier emission) with a certain amplitude [1]. This two-level switching with amplitude ΔI_D is illustrated in Fig. 1 for the drain current I_D through a Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET), whereby the time spent in the high (t_1) and in the low current states (t_0) is randomly distributed. Since the first report on RTS in small-area silicon MOSFETs [2], great progress has been made in the understanding of the underlying physics [1,3–5]. Not only the trap parameters (activation energy for emission and capture, capture cross section) can be derived from the study of the emission (τ_e) and capture (τ_c) time constants as a function of temperature but it is also possible to determine the location of the trap in the depletion region of the semiconductor by analyzing for example the current switching amplitude as a function of the operation bias [6].

As will be shown here, there exist different ways for the extraction of the RTS parameters [6], which are defined in Fig. 1. The most direct method is based on time domain measurements, representing the current amplitude at different measurement times in a histogram. Combined with the corresponding frequency-domain Lorentzian spectrum, ΔI_D, τ_c and τ_e can be determined as a function of the operation biases or temperature. However, in the case of complex, multilevel RTSs, a more refined analysis, based on so-called Time Lag Plots (TLPs), is required [7–10]. A TLP consists of representing $I_D(i+1)$ versus $I_D(i)$, with i the ith data point measured with a fixed time step Δt. A TLP is thus a graph showing the current measured at moment $(i+1)\Delta t$ versus the current at $i\Delta t$. Further improvements can be reached by using so-called colored or weighted TLPs [11]. Finally, it will be shown that analyzing the bias-dependence of the RTS parameters in vertical polycrystalline silicon (polysilicon) transistors for Non-Volatile Memory (NVM) applications enables estimating the trap location in the channel.

2. RTS parameter extraction

A simple way to extract the RTS amplitude is ranking the measured current values in a histogram [6]. In the case of a simple two-level signal, this results in two Gaussian shaped distributions, whereby the difference between the maxima yields ΔI_D, while the ratio of the areas underneath the peaks gives the average time constant ratio τ_1/τ_0. Time-domain analysis is well-established and has been reviewed in the literature [1]. The up and down time constants usually correspond with the capture (τ_c) and the emission time (τ_e) of a charge carrier. Studying these time constants at different temperatures allows the construction of an Arrhenius diagram, the slope of which gives the activation energy for capture and emission, respectively [1,12,13].

* Corresponding author at: Imec, Kapeldreef 75, B-3001 Leuven, Belgium.
1 Presently at Micron Technology Belgium, Imec Campus, Belgium.

http://dx.doi.org/10.1016/j.tsf.2015.08.037

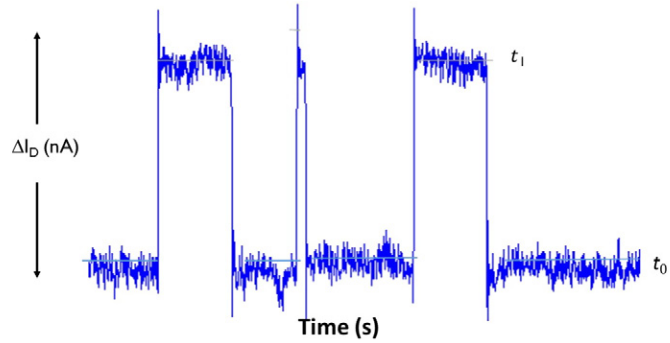

Fig. 1. Random telegraph signal in the drain current of a MOSFET and definition of the three relevant parameters: the amplitude ΔI_D; the high (up) t_1 and low (down) time t_0.

From the intercept, one derives the majority carrier capture cross section.

Alternatively, one can combine an amplitude histogram with a noise spectrum in the frequency domain to derive the RTS parameters. This is the Fourier transform of the power in the current fluctuations, yielding in the case of an RTS a Lorentzian spectrum, given by [1]:

$$S_I = \frac{A_0}{\left[1 + \left(\frac{f}{f_0}\right)^2\right]} \quad (1)$$

with S_I the current noise power spectral density (PSD), A_0 the plateau amplitude at low frequency f ($\ll f_0$) and f_0 the corner frequency of the generation-recombination (GR) level, corresponding with [1]:

$$2\pi f_0 = \frac{1}{\tau_c} + \frac{1}{\tau_e} \quad (2)$$

Combining Eq. (2) with the τ_c/τ_e time constant ratio derived from an amplitude histogram yields two equations with two unknowns so that both time constants can be calculated. The corner frequency is for example derived from the PSD spectrum by multiplying with f, yielding a peak versus frequency with maximum at f_0 [6,12,13].

Recently, the use of a TLP representation for the determination of the RTS parameters has been introduced [7]. It consists of plotting $I_D(i+1)$ against $I_D(i)$. An example is given in Fig. 2b, corresponding with the numerically generated time series of Fig. 2a. One can easily see that as long as the current stays on the high or on the low level, the TLP points will be on a diagonal through the origin, with two clouds of data points for the low and high current states. Only when a value is measured in the transition from high to low or vice versa, off-diagonal data points may be obtained [9]. The RTS amplitude is found from the distance between the center points of the two data clouds and the ratio of the high/low time constants follows from the ratio of the counts in both clouds. Combination with the PSD spectrum provides the necessary information to determine the capture and emission time constants.

The above methods work very well for a simple two-level RTS. However, as soon as two or more signals are present in the time domain, with different amplitudes and time constants the simple TLP method breaks down, as shown in Fig. 2b: only two data clouds can be discerned while the time series of Fig. 2a has been constructed from the Hidden Markov Model (HMM) in Matlab, containing a small- and a large-amplitude RTS. However, the use of a colored TLP [11], whereby a different weight is attached to the different data points in the clouds enables to unravel the four current levels, as shown in the histogram of Fig. 2c and the colored TLP of Fig. 2d. The method used here considers each column individually and normalizes all counts to 1 for each column. In a second step, the center of gravity of each cloud is determined by calculating the magnitude/degree of how many points are surrounding and how far are these points, taking account of all the grids in the data cloud. This procedure yields a weighted magnitude for one grid point. In this way, four different current levels can be distinguished which have to be assigned to two RTSs.

A final refinement of the RTS analysis in case of a small signal overwhelmed by random background (1/f) noise is the use of a time-series analysis based on the Hidden Markov Model [10]. In summary, there exists a whole arsenal from simple to more sophisticated techniques, enabling the accurate extraction of the trap parameters from a time series, containing simple or complex RTSs.

3. RTS in vertical polysilicon nMOSFETs

Currently, there is a strong interest in the development of poly-Si vertical transistors, schematically represented in the top of Fig. 3 for three-dimensional (3D) NVM development [14–16]. The channel consists of polycrystalline (poly-) silicon, with a grain size depending on the deposition parameters and containing grain boundaries (GBs).

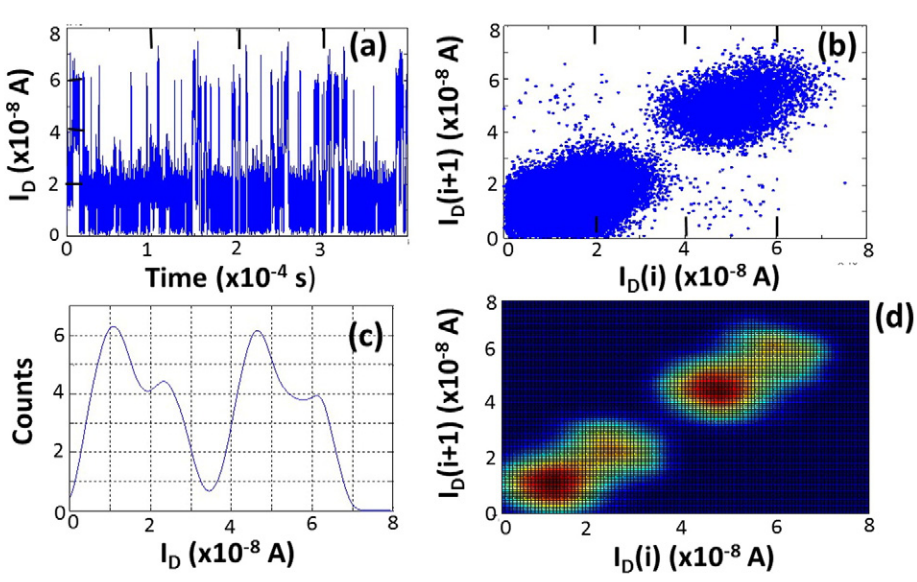

Fig. 2. (a) Time series generated by the hidden Markov routine in Matlab and containing a small and a large amplitude RTS; (b) corresponding simple TLP, showing only two clouds of data points; (c) amplitude histogram derived from the time domain signal and (d) colored TLP, showing the presence of the two RTSs.