HAL
open science

Process Optimization and Downscaling of a Single Electron Single Dot Memory

Christophe Krzeminski, Xiaohui Tang, Nicolas Reckinger, Vincent Bayot, Emmanuel Dubois

To cite this version:

Christophe Krzeminski, Xiaohui Tang, Nicolas Reckinger, Vincent Bayot, Emmanuel Dubois. Process Optimization and Downscaling of a Single Electron Single Dot Memory. IEEE Transactions on Nanotechnology, 2009, 8 (9), pp.737-748. 10.1109/TNANO.2009.2021653 . hal-00601623

HAL Id: hal-00601623

https://hal.science/hal-00601623

Submitted on 19 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1

Process Optimization and Downscaling of a Single Electron Single Dot Memory

Christophe Krzeminski, Xiaohui Tang, Nicolas Reckinger, Vincent Bayot, Emmanuel Dubois

Abstract

This paper presents the process optimization of a single electron nano-flash electron memory. Self aligned single dot memory structures have been fabricated using a wet anisotropic oxidation of a silicon nano-wires. One of the main issue was to clarify the process conditions for the dot formation. Based on the process modeling, the influence of various parameters (oxidation temperature, nano-wire shape) have been investigated. The necessity of a sharp compromise between these different parameters to ensure the presence of the memory dot has been established. In order to propose an aggressive memory cell, the downscaling of the device has been carefully studied. Scaling rules show that the size of the original device could be reduced by a factor two. This point has been previously confirmed by the realization of single electron memory devices.

Index Terms

Process modeling, flash memories, non volatile memories, quantum dot, single-electron device, device scaling, scaling limits, silicon on insulator technology (SOI).

I. Introduction

Conventional flash memory devices are believed to be hardly scalable since the design space vanishes below the 40 nm technology node [1]–[3]. Since a few years, most of the area reduction has been achieved by the downscaling of isolation and interconnection regions [4], [5]. Alternative and innovative storage structures for nonvolatile memories are strongly needed as the number of stored electrons is declining to dangerous small numbers. Many emerging research devices are under investigation using a large diversity of materials and approaches: phase memory changes [6], polymer memory [7], nanocrystal memory [8]. Among all, Phase Change nanocrystal Memory (PCM) is now acknowledged to be the leading candidate [9]. However silicon nanocrystal remains an interesting approach due to a close compatibility with some standard industrial process and for 10 nm devices [10]. In this paper, the fabrication of single electron nanoflash memories with silicon nanocrystal is under investigation. Numerous studies have been published on multinanocrystal memories generated by low-pressure chemical vapor deposition [11], [12] by implantation and annealing step [13] or by aerosol deposition techniques [14]. One strong issue faced by the multi-dots approach is the impact of the dot distribution fluctuations on the electrical properties [15]. A relatively different approach is followed here since the idea is to improve and investigate the limits of a single dot memory. The fabrication of this alternative memory is based on an original combination of classical microelectronics process

steps. The main idea is to separate a silicon dot from the channel using anisotropic oxidation. It has already been shown that pattern dependent oxidation (PADOX) could be used to fabricate small silicon single-electron transistors (SETs) [16]. Self-limited oxidation effects and strain have been carefully engineered to generate silicon nanostructures embedded in silicon dioxide using pattern conversion [17]. A different approach has been undertaken in the present work. The anisotropic character is generated using the strong dependence of the oxidation kinetics on arsenic doping profile [18], [19]. One issue is to improve the understanding of the conditions for the dot formation. Based on that background, the standard process is first summarized in section II. Process simulation and models limitations are discussed in section III. Section IV reports on the conditions for the dot formation and on the influence of the oxidation temperature and of the nanowire width. Device downscaling is explored in the last section.

II. Device architecture and processing

This section briefly recalls the main process steps since a detailed overview can be found in reference [20]. The critical step corresponds to the implantation through a thin SiO_2 film of a high arsenic dose (1×10^{15}cm^2). The main objective is to create a sharply localized and a very doped area in the silicon active layer (200 nm thick) of a SOI wafer. A nitride layer is deposited on the substrate. E-beam lithography and reactive ion etching are used to pattern the nanowires. Fig. 1 presents the layout of the device. The central part of the device consists of a 130 nm by 130 nm square. Two 100 nm wide constrictions are created in order to realize the connection between the source and drain regions and the central part. The different dimensions are defined so that a dot can be created in the central part of the device while being separated from the source/drain regions. Finally, a wet oxidation step is performed in order to create the dot. Fig. 2 shows a SEM photography of the central part where a single dot memory is obtained. A nano-floating gate is clearly observed on top of the channel.

III. Process simulations models

Based on process simulation, the aim of this study is to clarify the conditions that lead to dot memory existence. Two steps have been identified to play a key role: the wet oxidation step and the lateral overetching effect during patterning. The first one is critical for the generation and the separation of the dot/channel structure. The second one helps indirectly to achieve the desired pattern conversion. The main objective of this section is to discuss the limitations of process simulation with these two process steps.

A. Modeling of the oxidation step

To date, the seminal model of Deal and Grove remains the most popular approach used in process simulators to model the oxide growth [21]. The kinetic of oxide growth can be simply described by the following equation :

$$X = \frac{A}{2} \cdot \left[\sqrt{1 + \frac{t}{A^2/4B}} - 1 \right]. \tag{1}$$

The oxide thickness X is described by a linear-parabolic relationship as a function of the oxidation duration t. The term $\left(\frac{B}{A}\right)$ characterizes the growth rate regime:

$$X \sim \frac{B}{A} t \quad \text{for} \quad t \ll \tfrac{A^2}{4B} \tag{2}$$

whereas B gives the parabolic rate constant that governs the diffusion limited regime in the limit of very long oxidation time :

$$X \sim \sqrt{Bt} \quad \text{for} \quad t \gg \tfrac{A^2}{4B}. \tag{3}$$

The diffusivity of the oxidant species and the reaction rate k_{Si} can directly be deduced from the linear parabolic constants:

$$\begin{cases} D = B \cdot \dfrac{N_1}{2C^*} \\ \\ k_{Si} = \dfrac{B}{A} \cdot \dfrac{N_1}{C^*} \end{cases} \tag{4}$$

where C^* is the oxidant solubility in the oxide and N_1 is the number of oxidant molecules incorporated into a unit volume of the oxide.

The well-known breakdown of the Deal and Grove model [22]–[24] in the ultra-thin regime is not a significant issue here since only very long wet oxidation are considered. However, three points have to be carefully addressed : the mesh refinement, the influence and the diffusion of arsenic on the oxidation reaction rate and the global influence of strain on oxidation.

1) Grid refinement: One critical issue is the mesh refinement since the dot radius is less than 20 nm. The refinement must be as dense as possible in the region of high doping and more relaxed in other parts of the system (nitride mask, buried oxide). The second aspect is that the grid refinement must be preserved during the simulation of the oxidation process. One solution to preserve the quality of the initial grid during oxidation is to perform simulation using very small time-steps (less than 1 ms) and regenerate and adapt the grid frequently.

2) Influence of arsenic: The influence of arsenic on the reaction rate is described by the theory of Ho and Plummer [18], [19]. The physical effect behind the oxidation rate enhancement is that the silicon Fermi level is shifted by doping. The change of the Fermi level enhances the concentration of vacancy that naturally provides more reaction sites for the oxidation mechanism. The linear oxidation rate in presence of dopants can be expressed by:

$$\left(\frac{B}{A}\right)^{doped} = \left(\frac{B}{A}\right)^{0} \cdot \Big[1 + \gamma \cdot (C_v - 1)\Big] \tag{5}$$

where $\left(\frac{B}{A}\right)^0$ is the intrinsic linear reaction rate for wet oxidation which depends on many other parameters (temperature, pressure), γ is an empirical temperature dependent parameter and C_v is the normalized total vacancy concentration. Since dopants are assumed to influence only the oxidation rate, it should be noticed that the anisotropy will be maximum in the reaction limited regime (Eq. 2) and negligible in the diffusion limited regime (Eq. 3). Fig.

3 shows the influence of the linear oxidation rate as a function of the arsenic concentration predicted by the Ho and Plummer model [18], [19]. The concentration should exceed 3×10^{19} at/cm^3 in order to observe an increase.

Another effect related to arsenic is the thermal diffusion during the oxidation process. A reduction of the initial arsenic gradient probably takes place during the oxidation process. In order to describe the diffusion of arsenic during this step, a standard five stream model has been used [25]. The main issue is to have a correct description of the amplitude and the shape of the arsenic peak. A correct description of the arsenic tail and TED (Transient Enhanced Diffusion) effects is not crucial here.

3) Influence of strain: In our specific case, the presence of arsenic has a major impact on the reaction rate. On the other hand, strain are expected to reduce the oxidation rate. The presence of a nitride mask on top of the nanowire and the concave shape of the nanowire generates strain during the oxidation process. In most of process simulators, stress can be included in the basic Deal and Grove model by considering stress dependent parameters. As proposed by Kao et al. [26], [27] and Sutardja and Oldham [28], the hydrostatic pressure P is often assumed to influence the oxidation diffusivity by:

$$\begin{cases} D_0^\sigma = D_0 \cdot \exp\big[-\dfrac{PV_d}{k_B T}\big] & \text{for} \quad P > 0 \\ D_0^\sigma = D_0 & \text{for} \quad P \leq 0 \end{cases} \tag{6}$$

On the other hand, a compressive normal stress component at the Si/SiO$_2$ interface reduces the oxidation rate:

$$\begin{cases} \left(\dfrac{B}{A}\right)^\sigma = \left(\dfrac{B}{A}\right) \cdot \exp\big[-\dfrac{\sigma_{nn} V_k}{k_B T}\big] & for \quad \sigma_{nn} < 0 \\ \left(\dfrac{B}{A}\right)^\sigma = \left(\dfrac{B}{A}\right) \quad \text{for} \quad \sigma_{nn} \geq 0 \end{cases} \tag{7}$$

where k_B is the Boltzmann constant and T is the oxidation temperature in Kelvin, P is the hydrostatic pressure:

$$P = -\frac{1}{2}(\sigma_{xx} + \sigma_{yy}) \tag{8}$$

and σ_{nn} is the stress normal to the Si/SiO$_2$ interface, V$_d$ and V$_k$ are the corresponding activation volumes. A viscoelastic approach has been used in order to take into account strain relaxation in the oxide [29]

4) Model limitations: As previously described, dopant diffusion and oxidation are strongly related in our specific case. The use of a simulator where all these different models are implemented is required. Moreover, severe constraints are imposed on the grid refinement. Both DIOS [30] and TSUPREM4 [31] process simulators have been used. Relatively close results have been obtained as far as the kinetics of the dot formation is investigated. Thank to the DIOS features advanced mesh refinement it has been possible also to investigate the oxidation of the dot. In order to access the evolution of the various parameters of the memory cell, a monitoring tool has been developed [32]. The main limitation in our study is the description of the mechanical strain field in process simulators. Fig. 4 illustrates this point where the strain field in the oxide layer is reported for the simulation of the standard process described in the previous section. Two contours ($\pm$ 700 Mpa) of the hydrostatic pressure are reported based on

TSUPREM4 simulation. Very large compressive stress levels are observed in the vicinity of the dot whereas the nitride mask is subject to a high tensile stress. As far as we approach the outside silicon dioxide interface, stress decreases as the relaxation is much more important. Such a large level of stress at the nanometer scale indicates that we are too close to the limits for a finite elements simulations to expect a correct description of the mechanical stress around the dot. The use of coupled atomistic/continuous models could be a solution [33]. The second limitation is the use of a viscoelastic approach to describe the oxide deformation [29]. It has already been shown by Rafferty et al. [34] that linear viscous approach strongly overestimates the stress generated during cylinder oxidation at low oxidation temperature and large deformation [1]. In principle, an elastoplastic model would be necessary to properly estimate the stress field surrounding the silicon dot for low oxidation temperatures [35]. The third limitation is the modeling of the pile-up of arsenic at the Si/SiO_2 interface [36]. *Despite these fundamental modeling restrictions, it will be shown that interesting trends can be extracted based on the relatively good description of experimental oxidation kinetics in the presence of arsenic by the coupling of the Ho and Plummer model [18], [19] and the Deal and Grove models [21].*

B. Modeling of the overetching step

Due to the strong concentration of dopants, it has been observed experimentally that the top of the mesa structure is overetched during its creation. This overetching effect has been observed and explained by Winters et al. [37]. The concentration of negative etching ions is increased due to the electrostatic interaction between fluorine ions from the reactive plasma at the surface and the dopant charges in the silicon. The increase of the etching rate in the high doping region leads to a trapezoidal shape for the mesa-structure with a neck at the top concentration. This effect has a major impact experimentally on the final device structure and facilitates the creation of the dot. The SEM image (Fig. 5) clearly shows the presence of overetching in the presence of strong arsenic doping. This effect must be discussed and incorporated in our two dimensional simulations. An empirical approach has been adopted in order to match as finely as possible the experimental shape. Fig. 6 shows the shape of the mesa-structure matching the experimental configuration of Fig 5.b). The neck at the top is approximated by a simple trapezoidal shape. Three etching steps are used in the process simulator to generate the full structure. The nitride mask etching is simulated by a perfect anisotropic etching. Next 90 nm of the silicon layer are etched with an angle of 104° with respect to the horizontal direction to simulate the overetching and the presence of the neck. Finally, a classical anisotropic etch with an angle of 60° (from the horizontal direction) is performed to match the shape of the mesa-structure at the bottom. All of this process leads to an accurate shape of the mesa-structure matching the experimental one.

[1] A two dimensional implementation of an elastoplatic approach is very complex and subject to strong numerical issues. A rough estimation can be obtained by the application of the Rafferty model [34] to our system. Considering a silicon cylinder of 20 nm and a oxide cylinder of 50 nm gives a radial compressive strain component of 1.88 Gpa and an hydrostatic pressure of 454 Mpa at the Si/SiO_2 interface between the two cylinders

IV. Conditions of the dot formation

Experimentally, the creation of a silicon dot has always been observed for 80 minutes of wet oxidation at 800°C. Process variations show that the oxidation duration should be defined very finely in order to completely liberate the dot to avoid a complete consumption by the oxidation mechanism. Based on process simulation, the objective is therefore to investigate the best process conditions for dot formation. Two parameters have been identified for their influence on the dot formation: the oxidation temperature and the nanowire geometry.

A. Influence of the oxidation temperature

1) Lifetime of the silicon dot: The most important factor for the reliability of our device is the time window during which the dot is formed but not yet consumed. This point is governed by the difference between the oxidation time to create the dot and that to consume it. Fig. 7 reports the evolution of these parameters as a function of the oxidation temperature for various thermal budgets. First of all, an hyperbolic decrease ($\propto t^{-1}$) is observed for both the creation and consumption time as a function of oxidation temperature. For example, the creation time is 160 min. at 750°C and reduces to 24 min. at 900°C. This result tells us that the oxidation temperature must be higher than 750°C to keep a reasonable processing time. On the other hand, it has been observed that the difference between the creation time and the consumption time vanishes rapidly with increasing temperature. At 900°C, the dot is created and consumed at the same time, therefore the recommended wet oxidation temperature range is [775°C-850°C] in order to ensure the dot existence. The best compromise is around 800°C which ensures the presence of the dot in the central region and allows the full consumption of silicon in the constriction regions.

The oxidation temperature has also clear impact on the size of the dot. The evolution of the dot shape as a function of the oxidation temperature has been simulated. Fig. 8 reports both the channel and dot cross sections. It can be observed that the size of the dot is strongly reduced by increasing the temperature. Fig. 8 shows that the dot maximal half width is about 25 nm at 800°C and reduces to less than 5 nm for 850°C. For a higher temperature than 900°C, no dot is created.

Obviously, the oxidation temperature has a clear impact on the process reliability. It directly influences the oxidation parameters such as the oxidation rate or the strain relaxation mechanism. However, the main striking effect is the reduction of the arsenic gradient responsible for the dot creation, which in turn affects the dot formation. The evolution of the arsenic profile upon annealing in an inert ambience is reported. in Fig. 9 for fixed temperatures and durations corresponding to the onset of the dot creation, as obtained from Fig. 7. For temperature below 850°C, the arsenic profile remains essentially unaffected by thermal diffusion. Dopant diffusion becomes significant at 900°C with a strong broadening of the arsenic profile[2][3]. It is clear that the broadening and the flattening of the arsenic

[2]A relevant comparison can be performed between our simulation and the experiments on arsenic TED by Solmi et al. [38]. It has been observed that effectively, the maximum does not diffuse at 750°C and 800°C. When the temperature reaches 900°C, a clear reduction of the maximum level of doping is observed for a 30 min. annealing.

[3]It can also be noticed that the Arsenic maximum (2×10^{20}at/cm^3) is just below the concentration where the diffusivity dramatically increases with doping concentration [39].

profile reduce the anisotropic character of the oxidation step and explains the difficulty to observe the dot creation at temperatures above 850°C.

B. Influence of the nanowire width

1) Critical linewidth (experiment) : The main objective is to estimate the influence of the nanowire initial width (at the top of the mesa-structure) on the dot existence. All the other process parameters related to the implantation or the oxidation step are those reported in section II. In the present experimental study, the top linewidth of the mesa-structure defined by the lithography step varies from 200 nm down to 50 nm. Fig. 10 presents some cross sections after the oxidation step for various linewidths (80, 120 and 200 nm). Precise data about the dot size could not be estimated from SEM analysis but the nanofloating gate diameter is about 20 nm with a significant error bar [40]. The critical experimental linewidth that ensures the dot existence for this set of experiments is observed for 120 nm $\pm$ 10 nm. The dot is not separated from the channel for a linewidth larger than 130 nm which is therefore favorable to generate the source/drain regions. On the other hand, for a narrower linewidth than 100 nm, no dot is observed from SEM cross section analysis and the configuration seems suitable for the constriction regions. However, it must be kept in mind that the experimental technique used to observe the dot could weaken the structure since some of the SiO_2 cover is removed and that a very small dot could be present but not observed.

2) Critical linewidth (modeling aspects): A theoretical study has been undertaken in order to quantify the influence of this parameter. The top linewidth of the mesa-structure used in simulation varies from 200 nm down to 50 nm. The overetching effect is taken into account as described in section III.B.

Fig. 11 presents the simulation of the oxidation step for various linewidths (80, 120 and 200 nm) corresponding to the constriction, central part and source/drain regions respectively. The direct comparison between the simulated configuration (Fig. 11) and their experimental counterpart (Fig. 10) is of particular interest. It can be noticed that there is a global agreement. Moreover, the triangular shape of the channel as well as the dot geometry (shape, position) is well described. The dot-channel distance is estimated above 45 nm and can be compare to the experimental value of $\sim$ 40nm in Fig. 11. The main discrepancy is observed for the bottom channel since clearly less oxidation is predicted [4]. Fig. 12 presents also a nice results on the dot creation line as simulation predicts almost exactly the experimental linewidth where a dot is first observed. From the simulation point of view, the dot is just separated from the channel for a 125 nm nanowire linewidth. On the other hand, the theoretical critical range for the dot existence is clearly much larger 40 nm compared to the experimental one. Two possible explanations can be proposed about this difference. The first one is related to the limits of the experimental technique used to observe the single dot as discussed just before-mentioned. The other one is related to the models limitations discussed in section III.A. Stress levels with decreasing silicon dot size are probably overestimated which could lead to the underestimation of wet oxidation kinetics and a larger critical range for the dot existence. On the other hand with respect to the

[4]Since the evolution of all the other geometrical parameters are reasonably described, such a large difference could be reasonably explained reasonably by the presence experimentally of a thinner silicon after etching compare to the configuration of Fig. 5. This layer plays a major role as it tends to delay the oxidation of the bottom part of the channel

experimental limits, simulation confirms that the silicon dot is completely consumed for a 80 nm wide nanowire configuration and that self-limited oxidation effects [41] are not sufficient to stop the reaction in our present case.

3) Simulation of the dot characteristics: Process simulation can help to have an estimate of the variation of the dot (shape, size, position)and the parameter range that allow for dot formation. The dot generated by our process has a triangular shape with a dot height often more important than the dot width. In order to simplify the discussion, the use of a mean radius which matches as closely as possible the dot geometry is proposed [5]. The variation of the dot properties as a function of the linewidth are reported Fig. 12. The dot radius r_{dot} decreases linearly with the initial linewidth. It can be described by the following relationship with a small set of parameters:

$$r_{dot} = 0.33(d - d_{max}) + r_{max} \quad (nm) \tag{9}$$

where d is the linewidth, d_{max} is the critical line for the dot creation (125 nm) and r_{max} is the associated maximum radius (20 nm).

Another interesting parameter is the variation of the dot position with the nanowire initial width. In general, the dot center is located 30 nm below the top silicon surface close to the midpoint between the top surface and the arsenic profile maximum. It can be observed that the thickness of the oxide separates the dot from the channel increases as the initial linewidth decreases (Fig. 12). As clearly shown in Fig. 14, the dot position approaches 15 nm for a linewidth around 90 nm. This trend can be simply described by the following linear empirical law as a function of the difference between the linewidth and d_{max} = 125 nm :

$$r_{center} = 0.484(d - d_{max}) \quad (nm) \tag{10}$$

This effect can be explained by the fact that, for thinner structures, the oxidation is more important in the region of strong arsenic gradient since there is a smaller amount of silicon to oxidize for the same thermal budget.

Finally, the evolution of the tunnel oxide thickness t_{ox} (the dot-channel distance) is also reported. The tunnel oxide increases almost linearly with downscaling as the thermal budget is kept constant and not adjusted for the critical linewidth study.

$$t_{ox} = 26.59 + 2.616(d - d_{max}) \quad (nm) \tag{11}$$

[5]The mean radius is extracted from the calculation of the dot surface defined by the integration of the various finite elements.

C. Conclusion

The existence of two critical linewidth (125 nm and 80 nm) for the dot existence has been clearly established by the process simulation in agreement with the experiments. The different structures are correctly described by process simulation despite the limitations in section III. The difference (15 %) observed on the critical line for the dot consumption could be explained by the limits of the experimental technique and also by the limitations related to strain effects at the nanometer scale, two dimensional arsenic dopant diffusion and segregation. Process simulation also provides interesting trends for dot geometry characteristics. The mean radius of the nanofloating gate scales down almost linearly with the linewidth. On the other hand, the dot-channel distance and the dot center position increases due to a strong oxidation rate at the bottom of the nanofloating gate. These variations have been quantified and empirical variation laws extracted.

V. Downscaling

One challenge faced by alternative nanometric devices is their scalability. In addition, aggressive size reduction would open the perspective to operate close to single electron mode of operation. The main objective of this section is to evaluate the possibility of downscaling this alternative device using the same simulation strategy that was above detailed.

A. Scaling issues

Downscaling of the standard device by a factor 2 is our target. The first issue is to adapt our process for the implementation on a thinner silicon film of 100 nm. The film thickness reduction mainly impacts the implantation conditions. The objective is to create a sharp arsenic gradient in a more localized area. First simulations show that the arsenic peak should not be too close to the silicon surface in order to create the dot since oxidation takes also place under the nitride mask. Therefore, the arsenic maximum has to be located into the top third of the silicon film as in the standard configuration. A reduction of the implantation energy around 60 keV is therefore recommended. The scaling of the dose must be investigated but other implantation conditions (tilt angle) could remain unchanged. The second issue is that in order to be able to form the constrictions, the region where the channel will be formed should be significantly thicker that where the dot will.

B. Linewidth variation

In the downscaling study, the initial configuration shown in Fig. 13 has been used. Arsenic implantation has been performed with the parameters chosen in the previous section (Energy 60 keV, Arsenic dose 1e15 at/cm^2). A nitride film has been deposited on top of the oxide layer. No overetching is considered since the main objective is to identify the trends. Moreover it will complexity the discussion of the results. A standard sidewall has been chosen to define the shape at the bottom of the mesa-structure.

An extensive set of simulations has been performed as a function of the initial linewidth. A low temperature wet oxidation performed at (800°C), the optimal temperature. The final resulting configurations (Fig. 15) are reported,

close to the dot creation. First, large configurations (80 nm to 60 nm) have been simulated. From the simulation standpoint, it clearly appears that downscaling is feasible. The final structure is well organized since the dot is in the top half of the film and the constriction region in the bottom half of the device. Thinner structures have also been investigated (from a 50 nm configuration down to 20 nm). Very surprisingly, the structure dot-channel has been observed to be scalable down to 20 nm with a dot around 5 nm as shown in Fig. 16.

C. Evolution of dimensional properties of the dot

Top Fig. 16 summarizes the evolution of two parameters (the dot mean radius and the thermal budget of the wet oxidation step)as a function of the nanowire linewidth. For a large configuration of 80 nm, the mean radius approaches the 10 nm target. As expected, the objective to reduce the size of the dot by a factor two has been achieved. The dot radius reduces for narrower configurations down to 5 nm for a 20 nm configuration. This figure emphasizes the interest of this approach to generate a large spectrum of dot using single parameter variation. The bottom graph in Fig. 16 shows the evolution of the oxidation time at the onset of the dot creation. The various final structures presented just before obtained by setting the oxidation duration close to the creation time for each linewidth. As shown in Fig. 16, the thermal budget downscales almost linearly with decreasing linewidth. For a large configuration, the oxidation time is about 80 min. and is less than 20 min. for the narrowest device. The impact of dose reduction by a factor 2 has also been evaluated. The linewidth range for the dot existence is clearly much sharper indicating that the doping gradient should not be further reduced.

D. Conclusion

Based on the downscaling study, the fabrication of a single electron memory cell has been achieved with a reduction by a factor 2 with respect to the original design [40]being kept constant. The silicon film thickness has been reduced to 100 nm and the implantation energy reduced to 55 keV (close to the best theoretical value of 60 keV). The formed dot memory has been observed through the gate and has a diameter of about 5-10 nm. The downscale device proposed by simulation is the 60 nm configuration of Fig. 15 with a dot size of 10 nm, a gate oxide of about 20 nm and a tunnel oxide of 15 nm. A preliminary device simulation study has shown that the device can operated for relatively thick oxides (15 nm for the gate oxide and 35 nm the for tunnel oxide) [42]. As already reported in a previous paper, single hole effects have been observed by several electrical characterizations at room temperature in the implementation of the downscale device [40]. As the process space allows some process modifications, more work should be done in order to optimise the device space for this device.

VI. Conclusion

Important conclusions can be drawn on the two major topics. First, the concept of single silicon dot generated by oxidation is of interest. It has been shown that a large range of single dots can be generated from 20 nm nanometer down to 5 nm. This method is very complementary to other techniques for which smaller dot are often obtained. Single dots generated with this technique hold the noteworthy property of tight dot size control. The proposed

flash memory design and the corresponding process open new perspectives for the integration of this concept in self-assembled nanowires to generate mass flash memory or the design of complex device with several single dots in the central part to fabricate single electron pumps [43].

Secondly, on the process modeling point of view, this study is one example of the interest of process modeling to speed-up the development of nanodevices. The impact of various parameters and configurations has been investigated. Major trends and sharp compromises can be determined. This work also illustrates the challenges faced by process simulation. The development of new devices needs the simulation of various coupled phenomenas (diffusion, oxidation, segregation, strain relaxation). Atomistic simulations are often proposed as an interesting alternative to simulate the process fabrication of such devices. But as clearly shown by this study, the main issue is to tackle the simulation of process fabrication and the interaction of relatively large silicon pattern with nanometric objects. Improving the limitations of continuous models and developing multi-scale approaches remain mandatory.

ACKNOWLEDGMENT

This work was supported by the European Union under Contract IST-2001-32674 SASEM Project (Self-Aligned Single Electron Memory and Circuits). One of the authors (C. K.) would like to thanks the reviewers of the SASEM project, Professor Asen Asenov (University of Glasgow) and Professor Niemeyer (PTB) for giving very fruitful directions during the project. The support of Patrick Van Hove from the European Commission (Future and Emerging Technologies) was also appreciated [44]. This work benefits also from fruitful discussions with E. Lampin (IEMN) on implantation and diffusion simulations, with V. Senez (IEMN) on oxidation and mechanical strain and with C. Bonafos (CEMES-Toulouse) on silicon nanocrystals oxidation kinetics. The regular support of Jean-Michel Droulez on workstations, software, licence, script for this project is also acknowledged.

REFERENCES

[1] *The International Technology Roadmap for Semiconductors*, 2005.

[2] K. Kim and J. Choi, "Technology for sub-50nm dram and nand flash manufacturing," *IEDM Tech. Digest.*, vol. 144, p. 323, 2005.

[3] C.-Y. Lu, K.-Y. Hsieh, and R. Liu, "Future challenges of flash memory technologies," *Microelectronic Engineering*, vol. doi:10.1016/j.physletb.2003.10.071, 2008.

[4] J. D. Blauwe, "Nanocrystal nonvolatile memory devices," *IEEE Nanotechnology*, vol. 1, p. 72, 2002.

[5] A. Fazio, "Future directions of non-volative memory technologies," *Mat. Res. Symp. Proc.*, vol. 830, p. D1.1, 2005.

[6] M. Lankhorst, B. Ketelaars, and R. A. Wolters, "Low-cost and nanoscale non-volatile memory concept for future silicon chips," *Nature Materials*, vol. 4, p. 347, 2005.

[7] J. Ouyang, C.-W. Chu, C. R. Szmanda, L. Ma, and Y. Yang, "Programmable polymer thin film and non-volatile memory device," *Nature Materials*, vol. 3, p. 918, 2004.

[8] S. Tiwari, F. Rana, H. Hanafi, A. Hartsein, E. F. Crabbé, and K. Chan, "A silicon nanocrystals based memory," *Appl. Phys. Lett.*, vol. 68, p. 1377, 1996.

[9] S. J. Hudgens, "The future of phase change semiconductor memory devices," *J. Non-Cryst. Solids*, vol. doi:10.1016/j.jnoncrysol.2007.09.111, 2008.

[10] D. Lammers, "Flash vendors facing scaling challenges," *Semiconductor International*, 2008. [Online]. Available: http://www.semiconductor.net/article/CA6624774.html

[11] T. Baron, A. Fernandes, J. F. Damlencourt, B. D. Salvo, F. Martin, F. Mazen, and S. Haukka, "Growth of si nano-crystals on alumina and integration in memory devices," *Appl. Phys. Lett.*, vol. 82, p. 4151, 2003.

[12] S. Lombardo, B. D. Salvo, C. Gerardi, and T. Baron, "Silicon nanocrystal memories," *Microelectronic Engineering*, vol. 72, p. 388, 2004.

[13] C. Bonafos, H. Coffin, S. Schamm, N. Cherkashin, G. B. Assayag, P. Dimitrakis, P. Normand, Carrada, V. Paillard, and A. Claverie, "Si nanocrystals by ultra-low-energy ion beam-synthesis for nonvolatile memory applications," *Solid State Electronics*, vol. 49, p. 1734, 2005.

[14] J. D. Blauwe, M. Ostraat, M. L. Green, G. Weber, T. Sorsch, A. Kerber, F. Klemens, R. Cirelli, E. Ferry, J. L. Grazul, F. Baumann, Y. Kim, W. M. J. Bude, J. Lee, S. J. Hillenius, R. C. Flagan, and H. A. Atwater, "A novel, aerosol- nanocrystal floating-gate device for non volatile memory applications," *IEDM Tech. Digest.*, vol. 139, p. 683, 2000.

[15] L. Pernolia, B. D. Salvo, A. F. P. G. Ghibaudo, G. Pananakakis, V. Vidal, T. Baron, and S. A. Lombardo, "Modeling of the programming window distribution in multinanocrystals memories," *IEEE Trans. on Nanotechnnology*, vol. 2, p. 277, 2003.

[16] Y. Ono, K. Yamazaki, M. Nagase, S. Horiguchi, K. Shiraishi, and Y. Takahashi, "Fabrication of single electron transistors and circuits using sois," *Solid State Electronics*, vol. 46, p. 1723, 2002.

[17] M. Uematsu, H. Kageshima, K. Shiraishi, M. Nagase, S. Horiguchi, and Y. Takahashi, "Two-dimensional simulation of pattern-dependent oxidation of silicon nanostructures on silicon-on-insulator substrates," *Solid State Electronics*, vol. 48, p. 1073, 2004.

[18] C. P. Ho and J. D. Plummer, "Si sio_2 interface oxidation kinetics: A physical model for the influence of high substrate doping levels," *J. Electrochem. Soc.:solid state science and technology*, vol. 126, p. 1516, 1979.

[19] ——, "Si sio_2 interface oxidation kinetics: A physical model for the influence of high substrate doping levels ii," *J. Electrochem. Soc.: solid state science and technology*, vol. 126, p. 1523, 1979.

[20] X. Tang, X. Baie, J.-P. Colinge, A. Crahay, B. Katschmarskyj, V. Scheuren, D. Spôte, N. Reckinger, F. V. de Wiele, and V. Bayot, "Self-aligned silicon-on-insulator nano flash memory device," *Solid State Electron.*, vol. 44, p. 2259, 2000.

[21] B. E. Deal and A. S. Grove, "General relationship for the thermal oxidation of silicon," *J. Appl. Phys.*, vol. 36, p. 3770, 1965.

[22] H. Z. Massoud, J. D. Plummer, and E. A. Irene, "Thermal oxidation in dry oxygen growth-rate enhancement in the thin regime," *J. Electrochem. Soc.*, vol. 132, p. 2685, 1985.

[23] ——, "Thermal oxidation in dry oxygen growth-rate enhancement in the thin regime," *J. Electrochem. Soc.*, vol. 132, p. 2693, 1985.

[24] C. Krzeminski, G. Larrieu, J. Penaud, E. Lampin, and E. Dubois, "Silicon dry oxidation kinetics at low temperature in the nanometric range: Modeling and experiment," *J. App. Phys.*, vol. 101, p. 064908, 2007.

[25] S. Dunham, "A quantitative model for the coupled diffusion of phosphorus and point defects in silicon," *J. Electrochem. Soc.*, vol. 139, p. 2628, 1992.

[26] D.-B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-dimensional thermal oxidation of silicon: Modeling stress effects in wet oxides," *IEEE Trans. Electron Devices*, vol. 35, p. 25, 1988.

[27] ——, "Two-dimensional thermal oxidation of silicon: Experiments," *IEEE Trans. Electron Devices*, vol. 34, p. 1008, 1987.

[28] P. Sutardja and W. G. Oldham, "Modeling of stress effects in silicon oxidation," *IEEE Trans. Electron Devices*, vol. 36, p. 2415, 1989.

[29] V. Senez, D. Collard, B. Baccus, and J. Lebailly, "Analysis and application of a viscoelastic model for silicon oxidation," *J. Appl. Phys.*, vol. 43, p. 720, 1994.

[30] *DIOS, SYNOPSYS TCAD Manuals, Release 6*.

[31] *TSUPREM-4, SYNOPSYS TCAD*.

[32] C. Krzeminski, E. Dubois, X. Tang, N. Reckinger, A. Crahay, and V. Bayot, "Optimisation and simulation of a alternative nano-flash memory: the sasem device," *Mater. Res. Soc. Symp.*, vol. 830, p. D1.6, 2005.

[33] W. A. Curtin and R. E. Miller, "Atomistic/continuum coupling in computational materials science," *Modelling Simul. Mater. Sci. Eng.*, vol. 11, p. R33, 2003.

[34] C. S. Rafferty and R. W. Dutton, "Plastic analysis of cylinder oxidation," *App. Phys. Lett.*, vol. 54, p. 1815, 1989.

[35] C. S. Rafferty, L. Borucki, and R. W. Dutton, "Plastic flow during thermal oxidation of silicon," *App. Phys. Lett.*, vol. 54, p. 1516, 1989.

[36] C. Steen, A. Martinez-Limia, P. Pichler, H. Ryssel, L. Pei, G. Duscher, and W. Windl, "Characterization of the pile-up of as at the sio2/si interface," *37th European Solid State Device Research Conference, ESSDERC*, 2007.

[37] H. F. Winters and D. Haarer, "Influence of doping on the etching of si(111)." *Physical Review B*, vol. 36, p. 6613, 1987.

[38] S. Solmi, M. Ferri, M. Bersani, D. Guibertoni, and V. Soncini, "Transient enhanced diffusion of arsenic in silicon," *J. Appl. Phys.*, vol. 94, p. 4950, 2003.

[39] A. N. Larsen, K. K. Larsen, P. E. Andersen, and B. G. Svensson, "Heavily doping effects in the diffusion of group iv and v impurities in silicon," *J. App. Phys.*, vol. 73, p. 691, 1993.

[40] X. Tang, N. Reckinger, V. Bayot, C. Krzeminski, E. Dubois, A. Villaret, and D. Bensahel, "Fabrication and room temperature single charging behavior of self-aligned single dot memory devices," *IEEE Trans. on Nanotechnology*, vol. 5, p. 649, 2006.

[41] H. Coffin, C. Bonafos, S. Schamm, N. Cherkashin, G. B. Assayag, A. Claverie, M. Respaud, P. Dimitrakis, and P. Normand, "Oxidation of si nanocrystals fabricated by ultralow energy ion implantation in thin sio_2 layers," *J. App. Phys.*, vol. 99, p. 044302, 2006.

[42] X. Tang, X. Baie, J.-P. Colinge, C. Gustin, and V. Bayot, "Two-dimensional self-consistent simulation of a triangular p-channel soi nano-flash memory device," *IEEE Trans. On Electron Devices*, vol. 49, p. 1420, 2002.

[43] A. B. Zorin, S. V. Lotkhiv, H. Zangerie, and J. Niemeyer, "Coulomb blockade and cotunneling in single electron circuits with on-chip resistors: Towars the implementation of the r pump," *Journal of applied Physics*, vol. 88, p. 2665, 2000.

[44] P. V. Hove, "European research on emerging nanoelectronic information processing components," *physica status solidi (a)*, vol. 203, p. 1051, 2006.

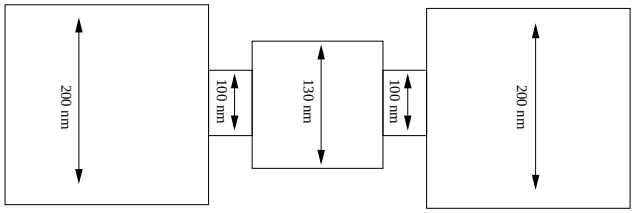

Fig. 1. Layout of the nanodevice. In order to generate the structure made of a single silicon dot in a channel separated by oxidation from the source/drain contact, different regions have to defined by a lithography step. The central part of the nanodevice where a single memory silicon dot separated from the channel is implemented is defined by a square area of 130 nm^2. This central part is connected to thin constriction regions (with less volume to oxidize) promoting the isolation of the dot from the source/drain regions. Finally, larger source/drain contact regions are fabricated in order to control the transport in the channel as a classical FET.

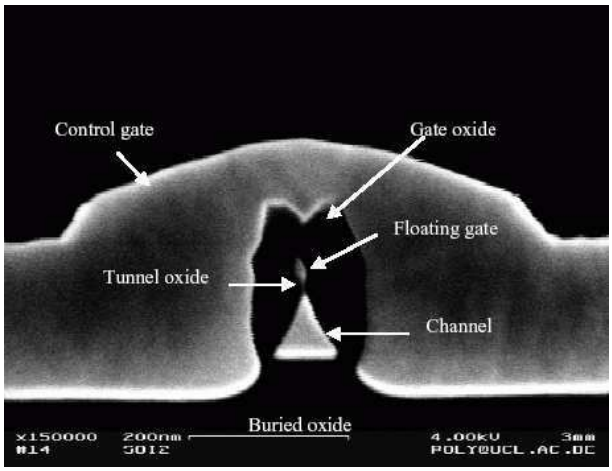

Fig. 2. SEM photography of the central part of the nanoflash memory.

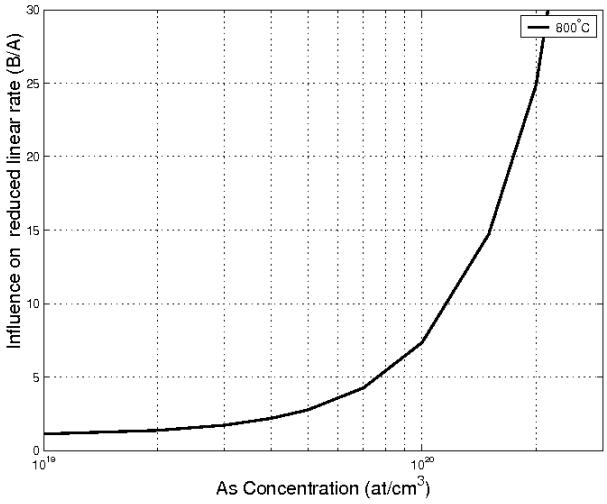

Fig. 3. Direct influence of the Arsenic concentration on the reduced oxidation linear rate predicted by the Ho and Plummer theory [18], [19]. Above an arsenic concentration of 10^{19} at/cm^3, $(B/A)^{doped} = (B/A)^0$ whereas an increase by a factor of 7 is observed at 10^{20} at/cm^3.

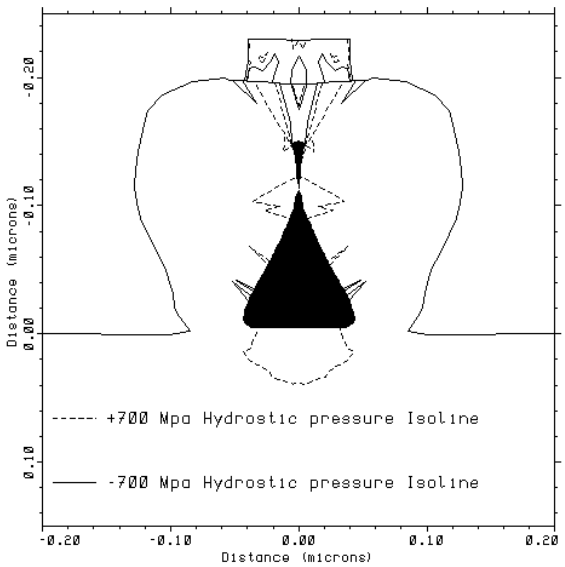

Fig. 4. The hydrostatic pressure characteristic of the strain field in the oxide generated by the dot formation after the wet oxidation step is reported. Black regions correspond to the two silicon parts (channel and dot). Both the compressive and the tensile isoline for an hydrostatic pressure of 700 Mpa have been calculated.

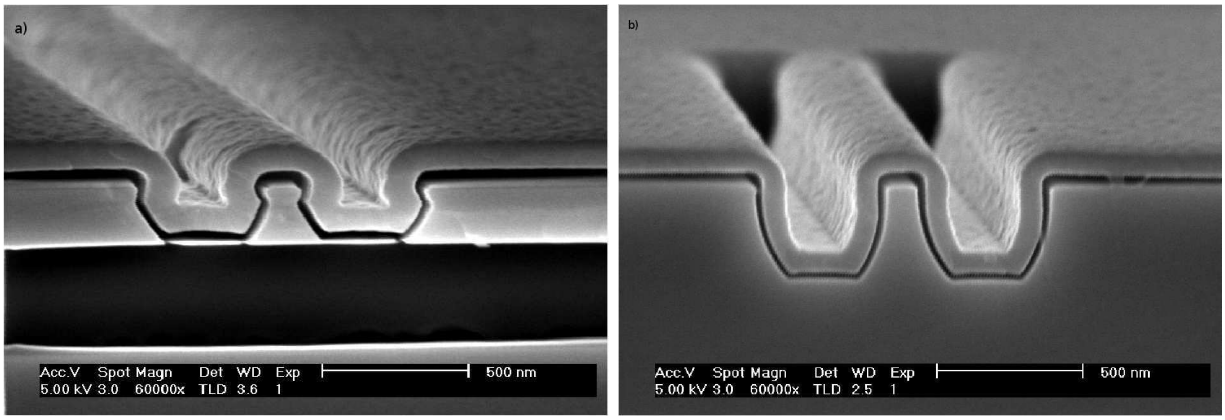

Fig. 5. Influence of the doping concentration on the shape of the mesa structure. a): The arsenic (1×10^{15}cm^2, 70 keV) profile gives overetching at the top of the structure. This leads to a trapezoidal shape with a neck. b): The classical trapezoidal shape for undoped silicon.

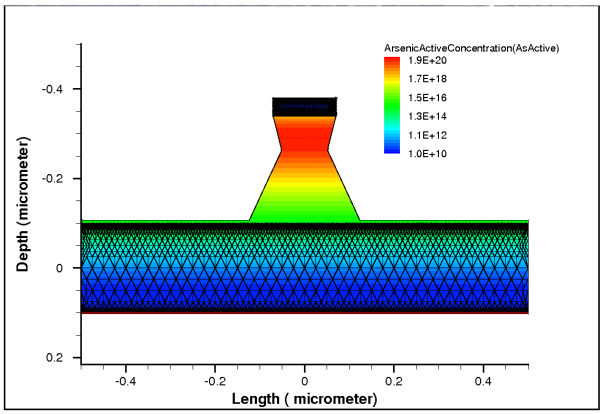

Fig. 6. The simulated counterpart for a 130 nm configuration. The arsenic gradient in the silicon layer is shown.

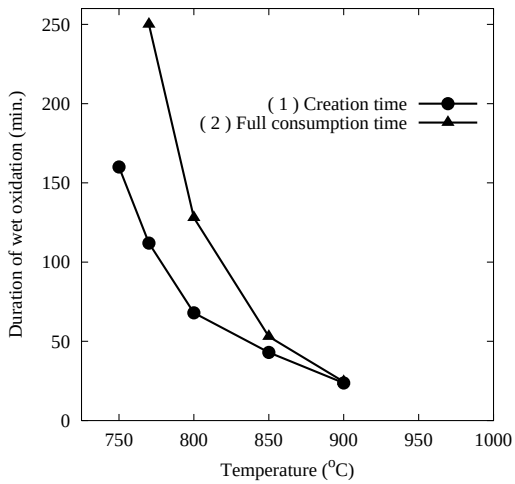

Fig. 7. Evolution of the times of oxidation to (1) create and totally (2) consume the silicon dot, for the two configurations, as a function of the oxidation temperature.

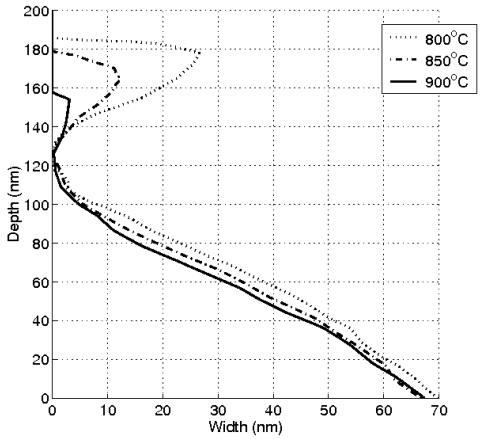

Fig. 8. Evolution of the mesa structure for various oxidation temperatures (800°, 850°, 900°). For each temperature, the structure has been reported just at the oxidation time previously determinated in Fig. 7

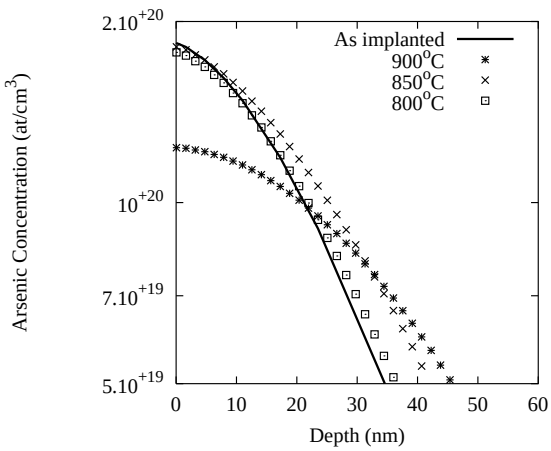

Fig. 9. Evolution of the arsenic profile with different thermal budgets. For each temperature, the oxidation duration is set to correspond exactly to dot creation (as reported in figure 7).

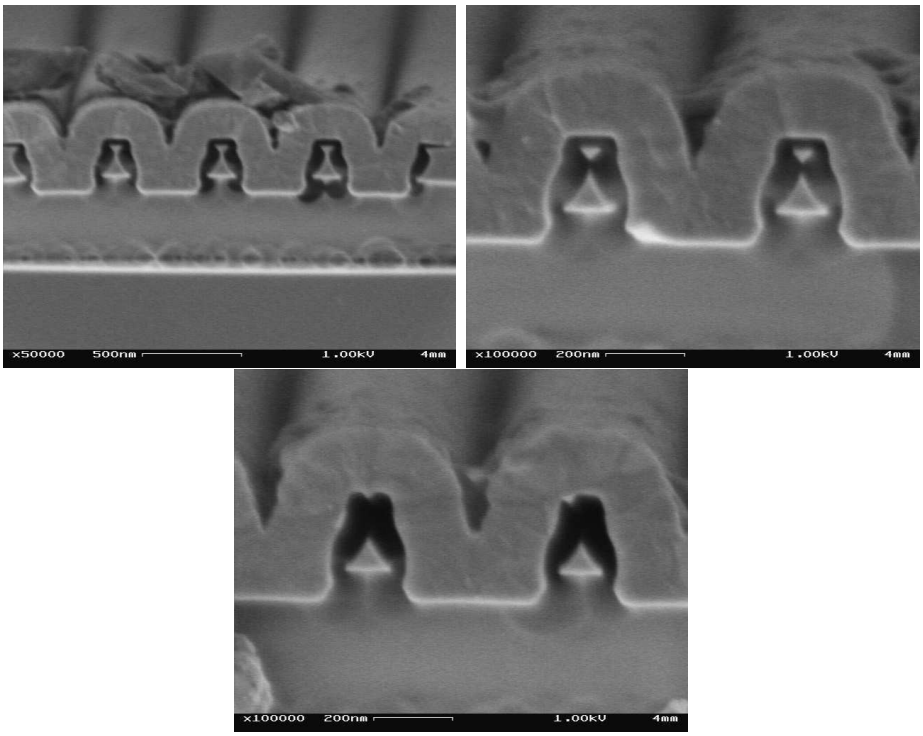

Fig. 10. Cross sections after the oxidation step for (200 nm, 120 nm and 80 nm) line widths corresponding to the different part of the device.

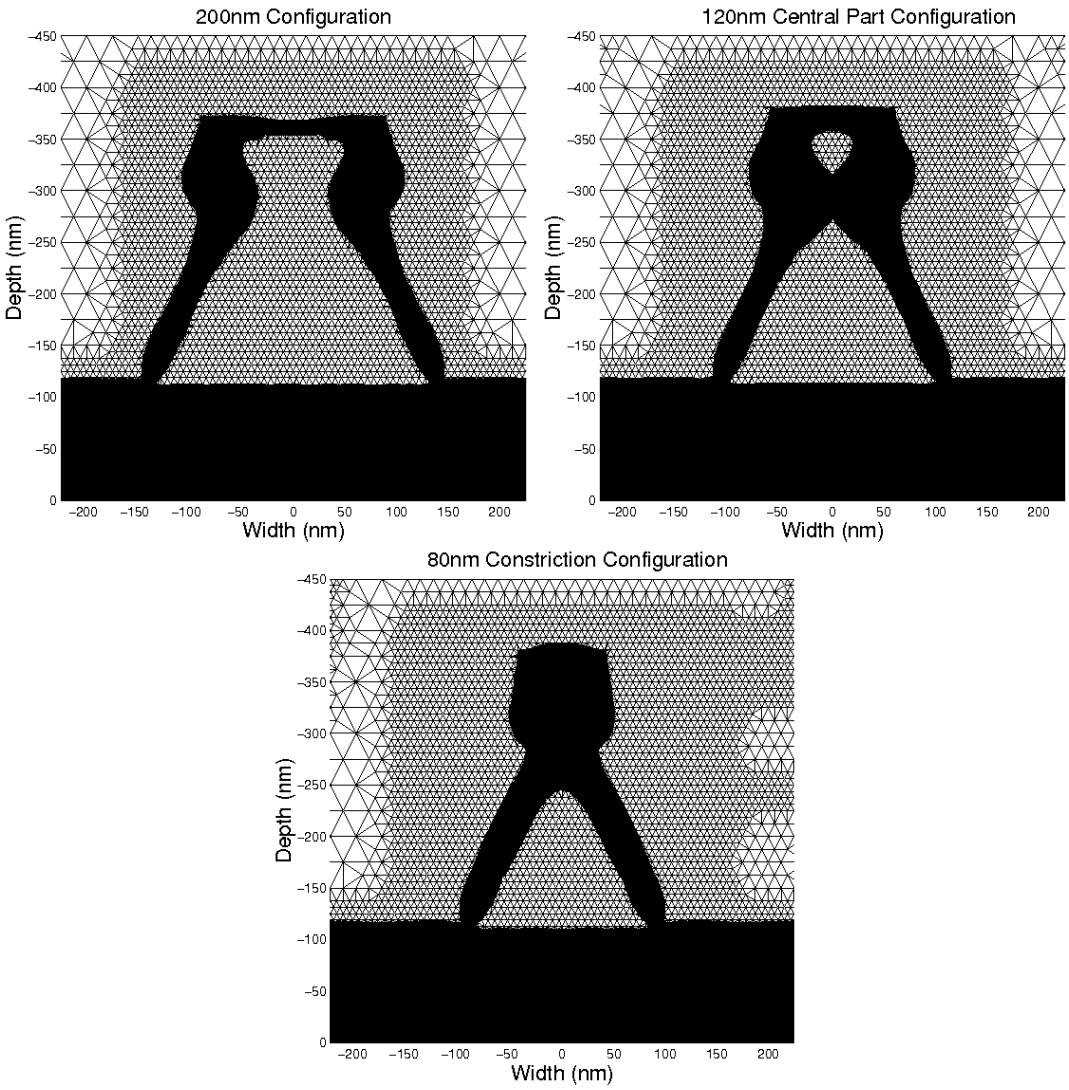

Fig. 11. Some configurations illustrating the critical linewidth experiments (respectively 180 nm, 120 nm, and 100 nm). All the process parameters have been kept constant (oxidation temperature 800°C, duration 88 min.) and the linewidth varies from 200 nm down to 50 nm. For a linewidth larger than 130 nm, the dot is not liberated from the channel. A dot is observed for 120 nm $\pm$ 10 nm.

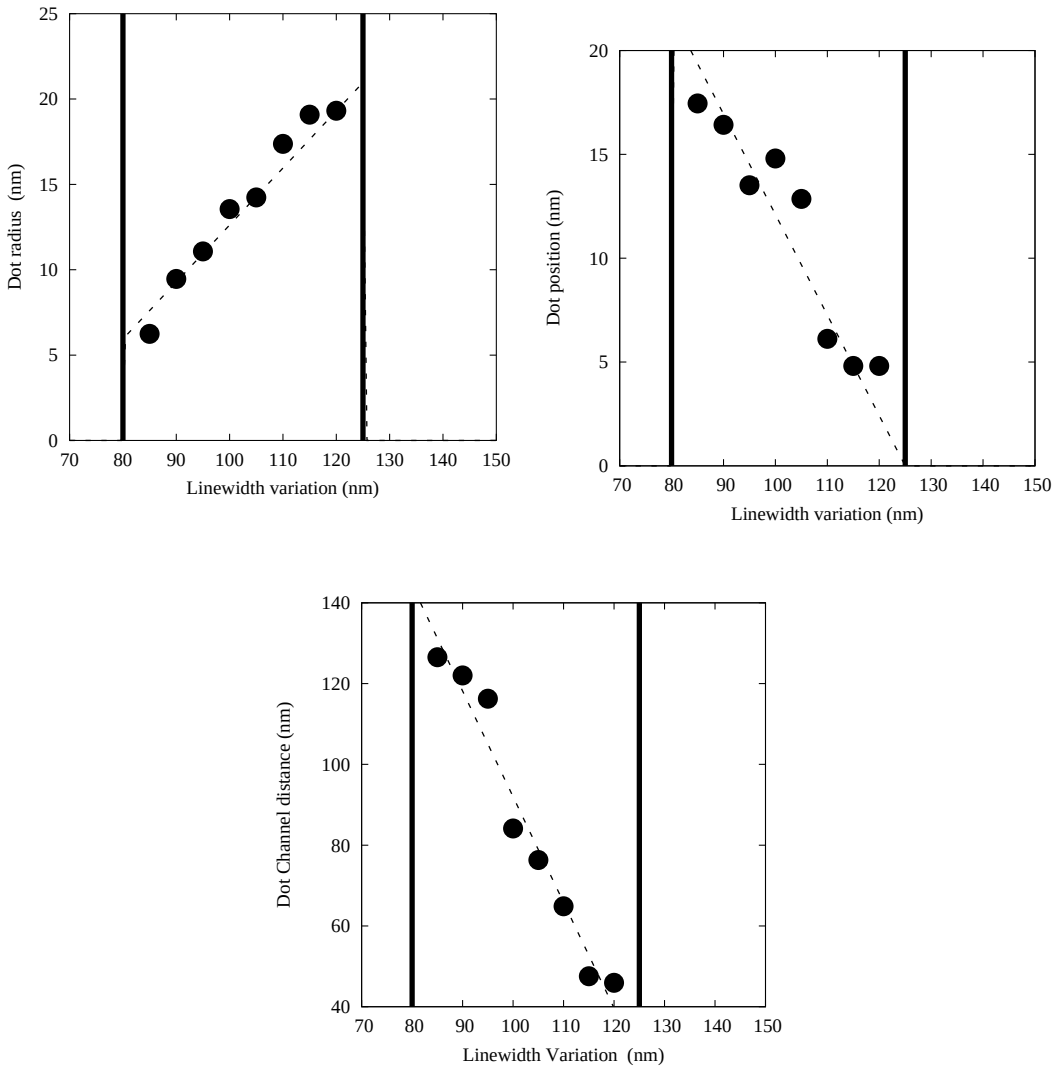

Fig. 12. Evolution of the main parameters (dot size, dot position, tunnel oxide thickness) between the two critical linewidths.

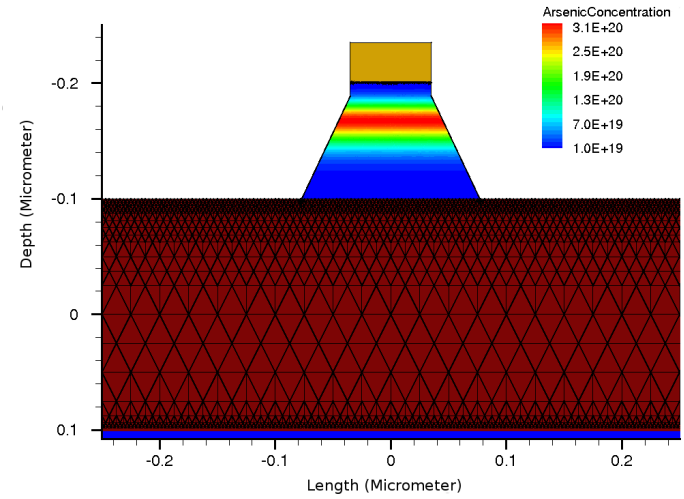

Fig. 13. Initial configuration (70 nm baseline) used in the downscaling study. Overetching is not considered here and a standard anisotropic etching has been perform to match the nanowire shape

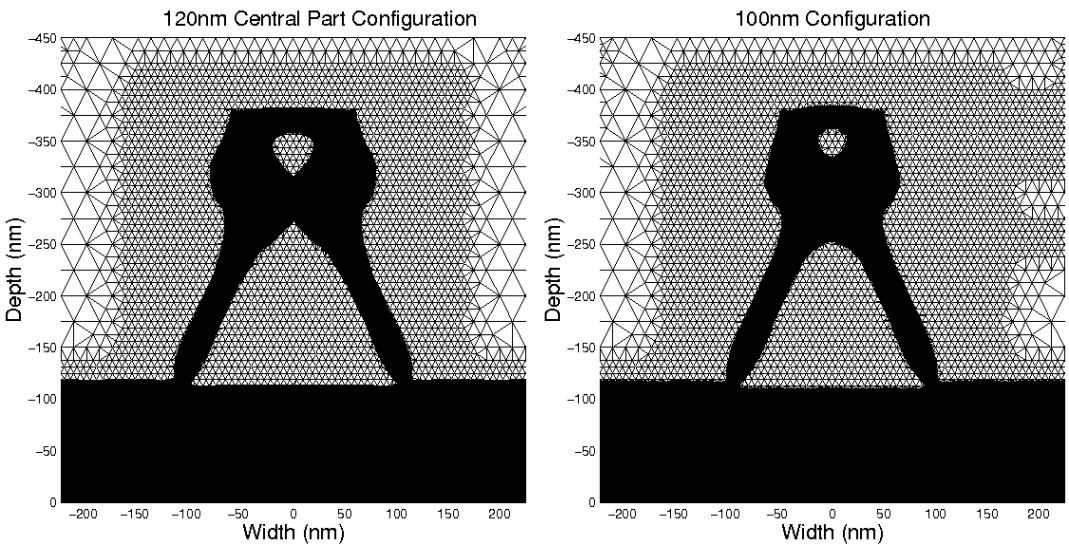

Fig. 14. Evolution of the central configuration for 120 nm and 100 nm linewidth.

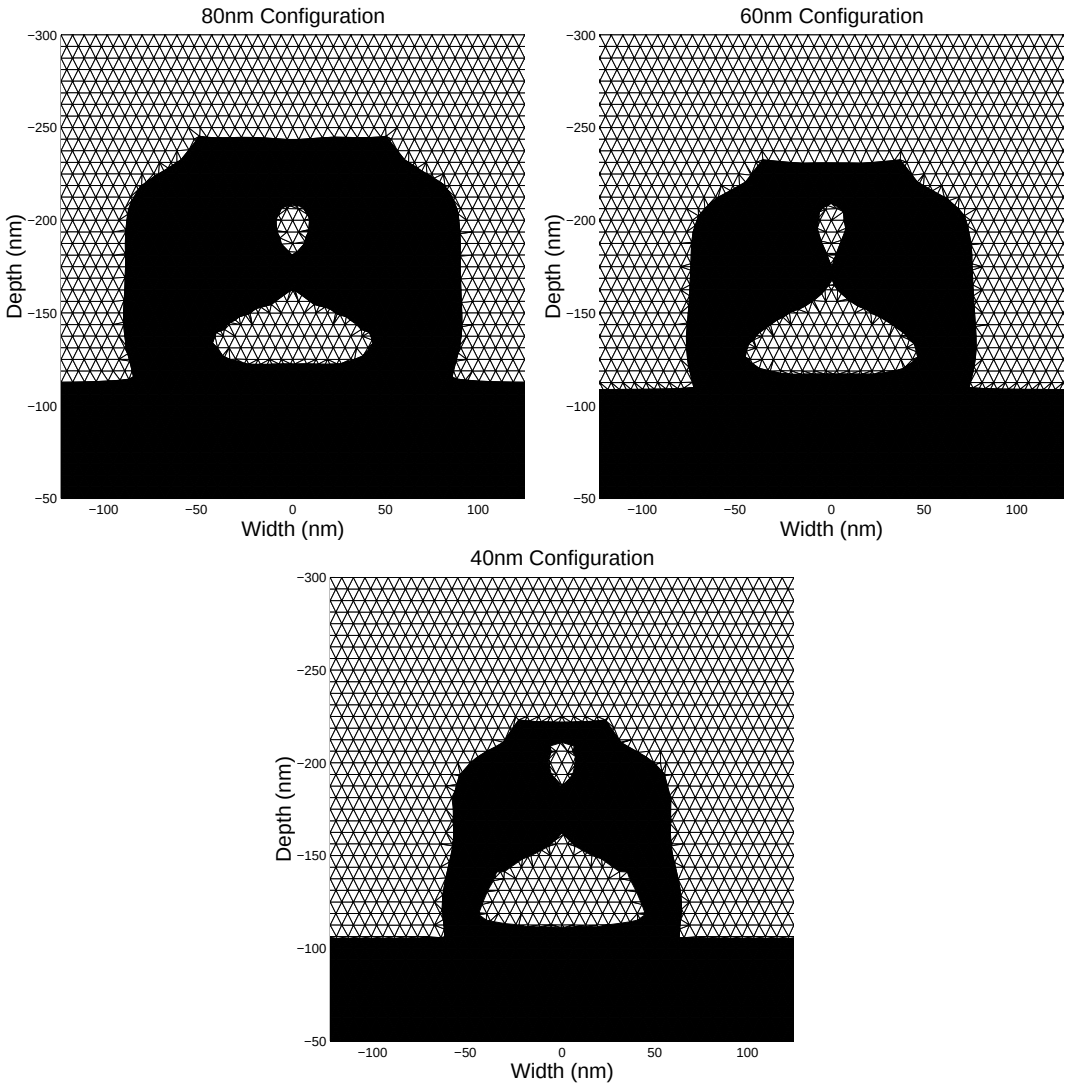

Fig. 15. Simulated 80 nm, 60 nm and 40 nm linewidth downscaled configurations. The oxidation time is set according to the estimation reported in Fig. 16.

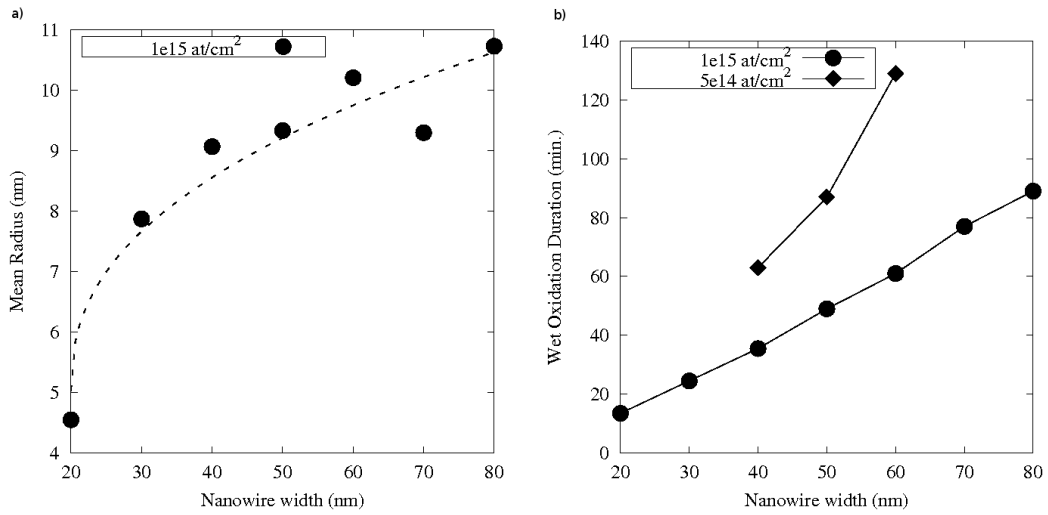

Fig. 16. Summary of the parameters evolution in the downscaling study. a) The dot radius evolution as a function of the initial nanowire linewidth is presented. b) The oxidation time necessary to create the silicon dot with decreasing linewidth for the standard dose (1×10^{15} at/cm^2) and with a lower dose (5×10^{14} at/cm^2) is reported. It can be observed that the thermal budget downscales linearly with decreasing linewidth and that reducing the dose by a factor two is not suitable because the range for the dot existence is clearly sharper [60-40] nm only.