Extraction of material parameters based on inverse modeling of three-dimensional interconnect fusing structures

Stefan Holzer\textsuperscript{a,}\textsuperscript{*}, Rainer Minixhofer\textsuperscript{b}, Clemens Heitzinger\textsuperscript{c}, Johannes Fellner\textsuperscript{b}, Tibor Grasser\textsuperscript{a}, Siegfried Selberherr\textsuperscript{c}

\textsuperscript{a}Christian Doppler Laboratory for TCAD in Microelectronics at the Institute for Microelectronics, Vienna, Austria
\textsuperscript{b}Austriamicrosystems AG, A-8141 Unterpremstätten, Vienna, Austria
\textsuperscript{c}Institute for Microelectronics, TU Vienna Gusshaussstrasse 27–29, A-1040 Vienna, Austria

Received 31 December 2003; received in revised form 6 April 2004; accepted 18 June 2004
Available online 24 July 2004

Abstract

An approach for determining higher order coefficients of the electrical and thermal conductivities for different materials is presented. The method is based on inverse modeling using three-dimensional transient electrothermal finite element simulations for electrothermal investigations of complex layered structures, for instance polycrystalline silicon (polysilicon) fuses or other multi-layered devices. The simulations are performed with a three-dimensional interconnect simulator, which is automatically configured and controlled by an optimization framework. Our method is intended to be applied to optimize devices with different material compositions and geometries as well as for achieving an optimum of speed and reliability.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Electro-thermal simulation; Parameter extraction; Polysilicon; Polycrystalline silicon; Polyfuse

1. Introduction

Semiconductor process technology nodes in the sub-micron regime often use polycrystalline silicon (polysilicon) fuses as one-time programmable devices which provide memories up to several kilobits and offer a cheap, efficient, and area-saving alternative to small non-volatile memories for System-on-a-Chip solutions. Approaches to increase the memory density by using 3-state fuses of layered materials have been reported [1]. Another important application is the use of simple field programmable gate arrays for trimming CMOS circuits to obtain a specific analog performance [2]. Furthermore, the fuses allow to provide variable elements such as trimable resistors or capacitor arrays [3]. Finally, the fuses may act classically as protective elements for improved protection and replacement of critical components before actual failures [4].

Programming is performed by sending a current pulse through the fuse, resulting in opening the polysilicon film after transition to a second-breakdown state. The transition occurs when parts of the polysilicon layer reach the silicon melting point, and the molten silicon is transported from the negative end through drift of ions in the applied field [5]. Fuses implemented in deep sub-micron technologies become more and more attractive in terms of power and area consumption, and hybrid approaches using other materials are losing importance [6]. Nevertheless, going to smaller ground rules below 350 nm implies decreasing the supply voltages to 1.5 V and below [7]. This constraint requires a careful optimization of the fuse layout, ensuring an efficient and reliable programming mechanism [8] and minimizing the necessary power consumption of the fusing process. As the fusing process takes place in a short time interval (between a couple of 10 ns for an applied voltage...
step up to the range of microseconds for a voltage ramp (c.f. Figs. 3 and 4), direct thermal measurements of this process are hard to obtain. Previously carried out work [9] already shed some light on the physics behind the fusing mechanism, but the optimization of the fuse structure for reliable and fast fusing was only possible via expensive experimental work by using test chips.

Our work focuses on gaining better insight into the material characteristics used in the structure, to enable a layout optimization through simulation. Since the electrical and thermal properties of polysilicon are complex functions of polysilicon film doping, grain size, and grain morphology [10], the average electrical and thermal properties as a function of temperature were obtained by measuring the transient resistivity response of the fuse through Joule self-heating and subsequent inverse modeling of the measured data to fit the observed behavior.

The electro-thermal self-heating simulations were carried out with the Smart-Analysis-Package (SAP) for three-dimensional interconnect simulation [11] in combination with Siesta, a TCAD optimization framework combining different simulators with gradient based and genetic optimizers [12].

This approach enables the optimization of the fuse layout significantly saving costs normally spent in design and production of layout test chips. Furthermore, a better insight into the transient electro-thermal effects occurring in the first couple of microseconds was gained.

2. Polysilicon fuse measurements

An industry standard deep submicron polycide gate CMOS process is used for the fabrication of the investigated polysilicon fuse devices. On a specialized test chip, multiple different layout variations were placed to find the optimum layout for fast and reliable fusing. A more complicated example of a fuse structure is shown in Fig. 1. The first experiments were performed with rectangular pulses. Nevertheless, due to the steep slope of the fuse terminal voltage the initial fuse heating is not well resolved. Furthermore, the initial transient behavior of the measurement circuit yields high errors in the measured current. To overcome these problems a voltage ramp was applied and the resulting fusing resistance was calculated by assuming ohmic behavior. The polysilicon layer in the fuse is doped to solid solubility, and therefore its conductivity may be considered to be approximately ohmic. Since all materials in the fuse except the polysilicon layer are metallic, this assumption is expected to give a reasonable estimate for the fuse resistivity. The devices were stressed with different triangular voltage ramps for a few microseconds. A pulse generator was used to define the length of the pulse. As the generator has a typical output impedance of 50 Ω and the resistor of the polysilicon fuse is lower than that, the source has to be buffered by an operational amplifier with a high slew rate to get a stable voltage. To avoid an additional voltage drop on a serial resistor a current probe was used. In addition, the voltage on the fuse was monitored by an oscilloscope to calculate the correct resistance value. The measurement setup can be seen in Fig. 2.

The measurement data for three different source voltages and the corresponding currents through the fuse as functions of time are given in Figs. 3 and 4, respectively. The current shows a significant increase over 65 ms due to melting processes which speed up the heating of the fuse structure. The current peak in Fig. 4 indicates the presence of high conductive paths in the fuse structure, which reduces the resistivity dramatically before the fuse structure is completely molten.

The large value of the temperature coefficient of the electrical conductivity in all three curves is caused by the combined Joule self-heating of the polysilicon/polycide layer sandwich (see Fig. 1), which causes the saturation of the current, as shown in Fig. 4. The high noise in the data during the first 10 μs is due to the low voltage level in this time regime and the resulting low signal-to-noise ratio.
3. Simulation and inverse modeling

3.1. Mathematical models

For the numerical calculation of Joule self-heating effects two partial differential equations have to be solved. Euler’s equation

\[ \text{div}(\sigma(T) \text{grad } \phi) = 0 \]  

(1)

gives the electric potential \( \phi \) where \( \sigma(T) \) denotes the temperature dependent electrical conductivity. The power loss density \( p \) is obtained by computing

\[ p = \sigma(T)(\text{grad } \phi)^2. \]  

(2)

The heat conduction equation

\[ c_p \rho_m \frac{\partial T}{\partial t} - \text{div}(\lambda(T) \text{grad } T) = -p \]  

(3)

is solved to obtain the temperature distribution where \( \lambda(T) \) represents the temperature dependent thermal conductivity, \( c_p \) the specific heat, and \( \rho_m \) the mass density. The temperature dependence of the thermal and the electrical conductivities are modeled as

\[ \sigma(T) = \frac{\sigma_0}{1 + \alpha_1(T - T_0) + \beta_1(T - T_0)^2} \]  

(4)

and

\[ \lambda(T) = \frac{\lambda_0}{1 + \alpha_2(T - T_0) + \beta_2(T - T_0)^2}, \]  

(5)

where \( \sigma_0 \) and \( \lambda_0 \) are the electrical and thermal conductivities at \( T_0 = 300 \) K, and \( \alpha \) and \( \beta \) are the linear and quadratic temperature coefficients of the specified materials.

3.2. Simulation setup

The layout of the fuse was transformed into a three-dimensional representation of the device using a detailed process description of the interconnect deposition and etch steps. Using the well known electrical conductivity of the interconnect and barrier layers the structural setup was calibrated by calculating the overall resistance of the structure excluding self-heating effects. The polysilicon conductivity was determined from the observed overall resistance, and the resulting value was compared to independently measured sheet resistances of the polycrystalline layer in fabrication. As can be seen from Table 1, an excellent agreement between simulation and measurements has been achieved. The subsequent transient simulations were set up including the thermal coefficients of the electrical conductivity, the thermal conductivity, and the heat capacity of all layers in the structure. The initial values of these parameters were taken from literature data.

3.3. Inverse modeling

State-of-the-art simulation frameworks \cite{12–15} support a wide range of simulators and optimization strategies. In addition to commercially available software, for instance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Poly Si</th>
<th>Poly Si_{Ist}</th>
<th>WSi_{x}</th>
<th>WSi_{Ist}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_0 ) (1/\mu m)</td>
<td>0.12</td>
<td>–</td>
<td>1.25</td>
<td>0.1–18.8</td>
</tr>
<tr>
<td>( \alpha_1 ) (1/K)</td>
<td>( 9.1 \times 10^{-4} )</td>
<td>( 10^{-3} )</td>
<td>( 8.9 \times 10^{-4} )</td>
<td>( 5–10 \times 10^{-3} )</td>
</tr>
<tr>
<td>( \beta_1 ) (1/K²)</td>
<td>( 7.9 \times 10^{-7} )</td>
<td>–</td>
<td>( 8.1 \times 10^{-7} )</td>
<td>( 3.5 \times 10^{-7} )</td>
</tr>
<tr>
<td>( \lambda_0 ) (W/Km)</td>
<td>45.4</td>
<td>40</td>
<td>119.4</td>
<td>100–179</td>
</tr>
<tr>
<td>( \sigma_2 ) (1/K)</td>
<td>( 2 \times 10^{-2} )</td>
<td>( 10^{-2} )</td>
<td>( 2.98 \times 10^{-2} )</td>
<td>–</td>
</tr>
</tbody>
</table>
[16] and [17], the simulation framework Siesta provides numerous types of optimizers which can be chosen for a particular problem. Reference data for the optimization presented here are measurements of the resistance derived from Figs. 3 and 4.

A brief overview of the structure and the data flow of Siesta is given in Fig. 5. At startup Siesta guesses the initial values of the free parameters within the user defined constraints for the three-dimensional interconnect simulator STAP of the SAP package, as introduced in [11]. The output of the simulation is parsed by Siesta in order to compare it with the reference data. A score value is computed which indicates how good these two data sets match. This value is submitted to the optimizer which generates the next $n$-tuple of free parameters corresponding to the score value. This improves the next score value which will be evaluated after the next simulation run with the currently produced values.

The optimizer mainly used in Siesta is a genetic optimizer, which relies on the theory of evolutionary computation and genetic algorithms described in [12–15]. The population of the $n$-tuples of free parameters are chosen randomly with respect to a Gaussian normal distribution. Several distribution and generation parameters can be configured and tuned to fit special needs. Furthermore, the simulation of the population can be distributed on a computer cluster to significantly decrease the optimization time.

Large intervals of free parameters can result in convergence problems because of non-physical parameter values which would result in negative resistance or negative doping. To avoid these problems, the simulation framework Siesta provides a divergence detection algorithm where it is signaled when the simulator has problems to converge.

4. Results and discussion

With the simulation framework SIESTA the thermal coefficients of the conductivities have been computed in order to minimize the difference between the reference data and the simulation results. To check the consistency of the setup, all thermal and electrical parameters were used for the automated simulation run, resulting in a total of 10 parameters. The resulting best fit to the measured reference data is given in Table 1. The electrical and thermal...
conductivities $\sigma_0$ and $\lambda_0$ as well as the linear temperature coefficient of the thermal conductivity $\alpha_2$ for polysilicon are in excellent agreement compared to data reported in Ref. [10]. The electrical conductivity of the polysilicon/tungsten silicide sandwich as a function of temperature is comparable to data measured electrically by external heating of the layers.

The optimized parameter set results in resistance characteristics as shown in Fig. 6 where an excellent match with the measurements is obtained. The strong increase of the resistance, as a function of time due to self-heating is evident. After a certain critical temperature is reached the resistance drops dramatically and the ohmic approximation loses its validity. To generalize this result for other fuse geometries this critical temperature has to be extracted. The critical temperatures of all the three samples are at about 1150 K (cf. Fig. 7). This value is much smaller than the silicon melting point of 1414 °C and the tungsten silicide (WSi$_x$ phase) melting point of 2015 °C [18]. The maximum temperature of the polysilicon fuse is observed in the center of the tungsten silicide layer as shown in Fig. 8. Several mechanisms for this low critical temperature are possible. First, the disordered region between the tungsten silicide and the silicon may have a stoichiometry closer to the eutectic point of the tungsten silicide system and therefore a lower melting point. But since the lowest eutectic temperature of the W–Si system is 1389 °C [18], this is not likely for pure alloys. Second, the high doping concentration of the polysilicon layer reduces the melting temperature as reported for silicon glasses with high boron and phosphorus contents. And finally, the assumption that all materials show ohmic behavior over the full temperature range between 300 and 1200 K does not hold for higher temperatures.

The intended target for getting the possibility to optimize fuse layouts for better performance is not affected since it is obvious from Fig. 7 that the melting begins always at approximately the same temperature. Therefore, the method should be applicable to other geometries as well. The agreement between experiment and simulation is excellent and provides a reliable base for carrying out predictive simulations of the transient temperature distribution during the initial heating phase of the fusing process.

5. Conclusion

We have presented a method to obtain important material parameters by inverse modeling using transient finite element simulations of complex interconnect structures. This method is capable of describing the electrical behavior of interconnect materials over a significant temperature range. Furthermore, it uses the transient thermal self-heating effect to separate different materials and their electrical and thermal properties. Nevertheless, the exact conduction mechanism inside the polysilicon layer is still not well reflected in this analysis. The impact of the grain boundary barriers and their behavior at high temperature could be addressed by implementing a more accurate model like the model of Mandurah [19]. However, we have demonstrated that our method is consistent and gives an excellent match to experimental results. With the extracted critical temperature where the material loses its ohmic properties, the geometry can be optimized in terms of reliability and speed.

References


