A new approach for studying experimental results of a-SiC:H thin-films alloys based on statistical physics

D. TSIOTAS[a*], L. MAGAFAS[b]
[a]*Department of Regional and Economic Development, Agricultural University of Athens, Greece, Nea Poli, Amfissa, 33100, Greece*
[b]*Laboratory of Complex Systems, Department of Physics, International Hellenic University, Kavala Campus, St. Loukas, 65404, Greece*

This paper applies a multilevel analysis based on statistical physics to detect structures of semiconductor a-SiC:H thin-film alloys with the best possible electrical performance expressed as a function of the temperature and hydrogen flow. Toward promoting the multidisciplinary demand in statistical physics, this paper broadens the conceptualization of the natural visibility graph algorithm by applying it to conductivity activation energy instead of time series; to study datasets described by insufficient information as a complex network. The statistical analysis shows that the conductivity activation energy is statistically indifferent to the temperature without the supply of hydrogen flows. However, the variation captured amongst conductivity activation energy levels supplied by hydrogen flows does not appear statistically significant. Hydrogen flows to the semiconductor body leads to better semiconductor structures at lower temperatures, where a zone of 17-20sccm has better conductivity activation energy levels. Finally, the network analysis reveals a rich-club configuration of temperatures and leads to three distinct conductivity activation energy states, corresponding to different structural (semiconductor) behaviors of the a-SiC:H thin-film alloys. The overall analysis proposes a framework of dealing with complexity under insufficient information, where traditional methods of statistical physics are of marginal functionality.

(Received February 4, 2021; accepted November 24, 2021)

Keywords: Structural optimization, Complex network analysis of time series, Community detection

1. Introduction

The high cost of energy that emerged in the post-industrialization era has led humanity to renewable energy sources and to the exploitation of solar energy, where crystal semiconductor photovoltaic panels suggest a popular solar energy device [1,2]. The amorphous semiconductor materials have caused a particular research interest because they have more advantages than mono-crystal semiconductor materials [2]. Such benefits concern their ability to apply in a thin-film format to large surfaces [3], their less expensive fabrication due to their high absorption factor [4], and their compatibility with other semiconductor elements (Ge, Si, C) for composing alloys of desiring properties [5]. Especially the amorphous silicon (a-SiC) thin-film alloys have attracted much interest due to their successful use in photovoltaic panel applications [5,6]. The prime techniques of constructing a-SiC thin-film alloys are either RF sputtering natural process for the non-hydrated (a-SiC) alloys or the chemical process for the (a-SiC:H) hydrated alloys [7], the Glow Discharge plasma process [8], and the chemical vapor deposition (CVD) process [9]. Amongst these three techniques, RF sputtering is the most commonly used [5,7] due to its lower application cost, its free-of-toxicity applicability, easy-to-use in industrial applications, and repeatability in providing standardized outcomes [5,10].

The use of a-SiC:H thin-film alloys in p-i-n (a-SiC:H(p)/a-Si:H(i)/a-Si:H(n)) photovoltaic components, which are of high photovoltaic performance, has widely attracted the research interest, especially towards the direction of optimizing the optoelectronic attributes of a-SiC:H that aim to the decrease of the energy gap and thus to the exploitation of a broader range of the solar spectrum [5,10]. A prime effort towards this direction concerns increasing carbon concentration in the a-SiC:H thin-film alloys, which causes a consequent increase of their optical energy gap up to a critical ceiling value, while afterward, the energy gap decreases [11,12]. However, carbon concentration increase induces the alloys' structural disordering that is related to a subsequent downgrade of their optoelectronic attributes [12]. Therefore, another direction for the photovoltaic performance optimization regards the control of the hydrogen concentration to the a-SiC:H thin-film alloys composition [11,12], which results in better structures compared to the carbon control techniques.

Within this context, based on the RF-sputtering development of a-SiC:H thin-film alloys, this paper examines their structural and electric properties as a function of the temperature and hydrogen flow. The overall approach employs techniques of statistical physics that can be applicable in cases of insufficient information, which are mainly related to laboratory or experimental restrictions due to high application costs. In particular, this paper uses experimental data of the RF-sputtering development of a-SiC:H thin-film alloys, in which the temperature ranges between 30 and 330^{o}C and the

hydrogen flow belongs to the set {0,9, 14, 20sccm}. The methodological approach builds on statistical physics due to its ability to deal with non-linear patterns resulting from the complex structure and topology related to the amorphous crystal configuration [10]. In general, the non-linear conceptualization in the research of amorphous semiconductors has been evident in the literature for almost three decades [13-15], in terms of structure [16], dynamics [17,18], optical behavior [17,20], energy applications, and physical properties [16]. Indicative approaches in the research of amorphous semiconductors employ non-linear stochastic processes [17], quantum-physics modeling of conductivity [15,19], non-linear and chaotic time series analysis [19-22], and structural modeling of super-lattice topologies [18].

Among such approaches, some works build on an aspect of statistical physics related to complex network analysis and network science. For instance, the authors of [23] developed queuing network models to design and analyze semiconductor wafer-fabs. The authors of [24] used complex network analysis to study cluster synchronization of mutually-coupled semiconductor lasers (SLs) networks with complex topology.

Network science is a modern discipline using the network paradigm to model communication systems into pair-sets of nodes and links [25,26] and enjoys multidisciplinary research that has been proven fruitful in providing insights about the structure and functionality of complex systems consisting of interconnected entities. An indicative example of the composite nature of network science is the modern research field of *complex network analysis of time series*. This approach builds on a combined statistics and physics conceptualization and enjoys applications in various disciplines, such as econometrics, finance, medicine, engineering, and more [27,28]. This field employs methods transforming a time series into complex networks that allow studying a time series as a complex network within a higher order of complexity context [29-33]. Toward promoting this multidisciplinary demand, this paper broadens the conceptualization of complex network analysis of time series by applying a fundamental technique to a conductivity activation energy instead of time series data; to study the performance of a-SiC:H thin-film alloys as a function of temperature and hydrogen flow. The overall approach proposes a framework of dealing with complexity under insufficient information, where traditional methods of statistical physics are of marginal functionality. The results of applying the proposed methodological framework to the case of a-SiC:H thin-film alloys are compared with available theoretical and experimental findings to evaluate the effectiveness of the proposed framework.

The remainder of this paper is structured as follows: Section 2 presents the methodology and data of the study; describing in detail the laboratory and the proposed methodological framework based on statistical physics and complex network analysis of time series. Section 3 shows the analysis results and discusses them under available experimental findings on a-SiC:H thin-film alloys research. Finally, Section 4 provides the conclusions.

2. Methodology and data

The methodological framework of the study employs techniques of statistical physics to study the electrical properties (expressed by the amount of conductivity activation energy E_a) of semiconductor a-SiC:H thin-film alloys as a function of the temperature (ranging between 30 and 300ºC) and hydrogen flow (ranging between, 0, 9, 14, and 20sccm). The methodological framework consists of a multilevel analysis including three steps; the first builds on statistical inference analysis, the second on pattern recognition, and the third on network analysis. Each step is described in brief in the following sections.

2.1. Laboratory background and data

Laboratory experiments measured the dependence of structural, optical, and electrical properties of a-SiC:H thin films [34,35], deposited by using the RF-sputtering process on the substrate temperature. This procedure took place for different hydrogen flow rates so that to gain the optimum material quality. The characterization of a-SiC:H used measurements of transmission electron microscopy [36], scanning electron microscopy [37], electron microanalysis [37], electron Auger spectroscopy [38], optical transmission in the range of the visible and infrared spectrum; of dark electrical conductivity in different temperatures; and electrical conductivity under certain conditions of illumination [34]. Also (based on a-SiC:H, characterization of Schottky diodes and "heterojunctions") was conducted by using measurements of current-bias voltage (*I-V*) [39], in different temperatures, differential capacitance-bias voltage (*C-V*), and spectral response (n-λ) measurements [34].

Within this context, the available data of the study regard measurements of the conductivity activation energy (E_a) of a-SiC:H thin-film alloys [34,35], which are expressed as a function of temperature (Ts) and hydrogen flow (H_f), as shown in Fig.1.

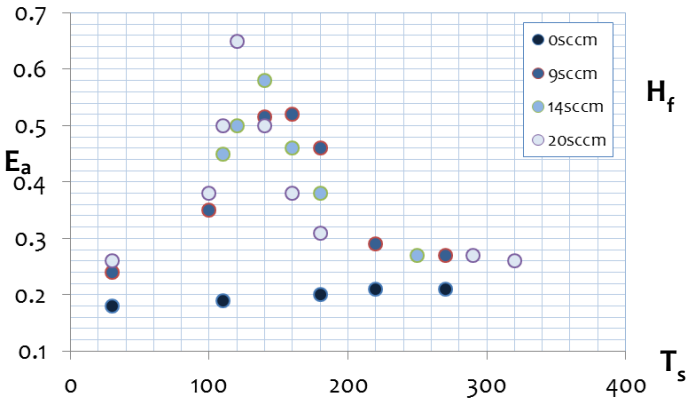

Fig.1. The available a-SiC:H conductivity activation energy (E_a) dataset, expressed as a function of temperature (Ts) and hydrogen flow (H_f). Data source: [34] and [35] (color online)

In particular, the E_a measurements extracted from four temperature flows referring to 0, 9, 14, and 20sccm

(standard cubic centimeters per minute), each configuring a temperature curve ranging between 30 and 300ºC. More specifically, the temperature range is calibrated at 30, 100, 120, 140, 160, 180, 220, 250, 270, 290, 300, and 320ºC. Each a-SiC:H conductivity activation energy curve $E_a=f(T_s, H_f)$ referring to a certain level of hydrogen flow $E_a(H_f=i$, where i=0,9,14,20sccm) is expressed as a function of temperature $E_a=f(Ts \mid H_f=i$, where i=0,9,14,20sccm) and therefore four (4) distinct conductivity activation energy curves $E_a(1)=f(H_f$:0sccm), $E_a(2)=f(H_f$:9sccm), $E_a(3)=f(H_f$:14sccm), and $E_a(4)=f(H_f$:20sccm) are available. In terms of data analysis [40,41], the available experimental data are insufficient in applying high-reliability statistical analysis. Within this context of insufficient information, the methodological framework builds on a multilevel consideration consisting of different methodological aspects (statistical inference analysis, pattern recognition based on parametric fitting, and network analysis); to compare the separate results of each analysis and configure conclusions based on the detected commonalities.

2.2. Statistical inference analysis

At the first step of the analysis, a statistical inference analysis applies to the available levels of a-SiC:H conductivity activation energy (E_a) to detect whether they are statistically different. The inference analysis builds on computing the confidence intervals (CIs) for the parameter θ (theta), according to the relation [40,41]:

$$\theta_{a,\{l,u\}} = \theta \pm t_{1-a/2,n-2} \cdot s_{\theta} \tag{1}$$

where: $t_{1-a/2,n-2}$ is the value of the Student's t-distribution (used instead of normal due to insufficient data) computed for a% confidence coefficient (1–a% confidence level) and n–2 degrees of freedom; s_θ is the standard deviation of the available dataset X on which we estimate the theta parameter; and $\{l,u\}$ are indicators referring to lower and upper bound of the CI respectively. In this step, the theta (θ) parameters that are estimated are the *average* ($\theta=\mu\{X\}$) and the *maximum* (θ=max$\{X\}$) value, respectively. Overlaid CIs ($\theta_{a,\{l,u\},1} \bigcap \theta_{a,\{l,u\},2} \neq \varnothing$) imply that the two estimated parameters θ_1 and θ_2 are likely to be equal (with up to 1–a% likelihood), whereas those not overlaying are 1–a% likely to differ [41]. On the one hand, the average-values testing applies amongst variables of different hydrogen flows (between flow levels), expected to provide insights into the emergence of structural differences between these flow levels in the a-SiC:H thin-film alloys conductivity activation energy (E_a) values. On the other hand, the maximum-values testing is implemented both amongst variables of different hydrogen flows (between flow levels) and amongst values of the same hydrogen flows (within flow levels). The between-flow consideration aims to support and complement the insights into the average-value testing and tests whether the a-SiC:H thin-film alloys structure is affected by changes of hydrogen flows. On the other hand, the within-flow consideration is expected to test whether the a-SiC:H thin-film alloys structure is affected by the temperature changes when hydrogen flows is constant.

2.3. Pattern recognition

We apply pattern recognition of the a-SiC:H thin-film alloys structure by using parametric fitting [40] to the available conductivity activation energy data $E_a=f(T_s, H_f)$ expressed as a function of temperature (T_s) and hydrogen flows (H_f). The types of the available fitting curves examined in this part of the analysis are: *linear* (1st-order polynomial), *quadratic* (2nd-order polynomial), *cubic* (3rd-order polynomial), *power*, *Gaussian*, *exponential*, and *logarithmic* fittings.

Generally, all available fitting-curve types can have the following mathematical expression [40,41]:

$$\hat{y} = f(x \mid b_1, b_2, ..., b_n) \tag{2}$$

where n is the number of the desired fitting parameters and the $f(x)$ function can be either logarithmic $f(x)=(\log(x))$, or polynomial $f(x)=x^m$, or Gaussian $f(x)=\exp\{(x-\mu)/\sigma$, or exponential $f(x)=(\exp\{x\})$, or any other. Within this context, the fitting analysis aims to estimate the b_i parameters that best fit the observed data y, so that to minimize the square differences $y_i - \hat{y}_i$ [41], according to the relation:

$$\min\left\{e = \sum_{i=1}^{n}\left[y_i - \hat{y}_i\right]^2\right\} = \min\left\{\sum_{i=1}^{n}\left[y_i - \left(\sum b_i f_i(x) + c\right)\right]^2\right\} \tag{3}$$

where $f(x_i)=y_i$. The beta coefficients (b_i) are estimated using the Ordinary Least-Squares (OLR) method [41], based on the assumption that the differences e in relation (3) follow the normal distribution $N(0,\sigma_e^2)$ [40,41]. Due to the insufficient information of the available dataset, the results of parametric fitting are evaluated at the macroscopic level, namely according to the type (typology) of the best fitting instead of their specialized (exact) mathematical formula.

2.4. Network analysis

Network analysis is applied to graph models, which are generated by a transformation based on the (natural) visibility graph (NVG) algorithm that is proposed by the

authors of [32] to transform a time series into a complex network. However, this study broadens the conceptualization of the NVG by applying it to discrete a-SiC:H conductivity activation energy datasets of the form $E_a=f(T_s, H_f)$ instead of to time series. This allows generating visibility graphs (associated with countable/discrete datasets) beyond the temporal conceptualization of time series analysis [28,30,31,42], introducing avenues of further research. After the NVG transformation, the topological properties of the associated visibility graphs are studied by computing network measures capturing different aspects of network topology and by detecting structural properties based on the modularity optimization approach. The network analysis is complementary to the previous statistical-based methods to the extent that it studies a graph structure (associated with a discrete dataset) instead of its source dataset and therefore it broadens the (source) data-space that is described by insufficient information. The parts composing the network analysis are described in brief in the following paragraphs.

■ Graph transformation: the natural visibility graph algorithm

The natural visibility (NVG) algorithm was proposed by the authors of [32]. The algorithm conceptualizes a time series as a landscape [28,42] and particularly as a sequence of successive mountains of differential height. Within this context, an observer standing on each mountain (expressing a time series node) can see, in both directions, as far as no other node obstructs its visibility, as shown in Fig. 2.

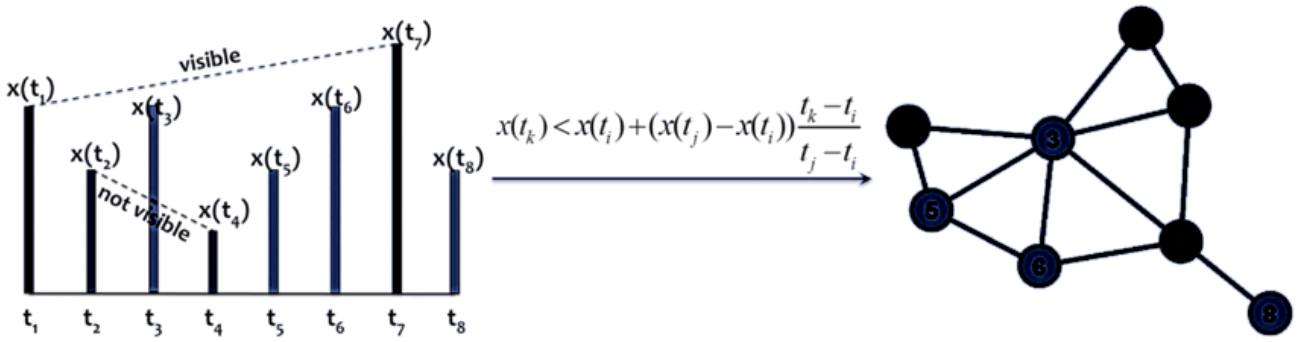

Fig. 2. (left) Example of a pair of visible (x_1,x_2) and non-visible (x_2,x_4) time series according to the natural visibility algorithm (NVG), (right) the visibility graph associated with the time-series shown at the left side (color online)

For a sequence $(t_k, x(t_k))$, k=1,2,…, where $x(t_k)$ are the numerical value of the time series' nodes at time t_k, we can correspond a time series node $(t_k, x(t_k))$ to a graph node n_k, namely $n_k \equiv (t_k, x(t_k)) \in V$. In the visibility graph $G(V,E)$ associated with the time series, where V expresses the node-set and E the edge-set, we can define that a pair of nodes $n_i,n_j \in V$ are connected $(n_i,n_j) \in E$ whether the NVG connectivity criterion (inequality) is satisfied, as described by the relation:

$$x(t_k) < x(t_i) + (x(t_j) - x(t_i))\frac{t_k - t_i}{t_j - t_i} \quad (4)$$

where $n_i \equiv (t_i,x(t_i))$ are the graph nodes and $(t_j, x(t_j))$ their corresponding time series nodes. In geometric terms, a visibility line is drawn between two nodes $(t_i,x(t_i))$, $(t_j,x(t_j))$ of the time series whether no other node $(t_k,x(t_k))$ that intermediates them obstructs their visibility. This implies that no other node intermediating the pair $(t_i,x(t_i))$, $(t_j,x(t_j))$ is higher so that to intersect their visibility line, as it is shown in Fig.2. Therefore, two nodes $n_i \equiv (t_i,x(t_i))$, $n_j \equiv (t_j,x(t_j))$ in the time series can enjoy a connection $(n_i,n_j) \in E$ in the associated visibility graph $G(V,E)$ when they are visible through a visibility line. The NVG interprets the time series as a landscape and generates a visibility graph corresponding to this landscape so that complex network analysis to be further applied [28,32,42].

■ Network measures

The visibility graph is an undirected and unweighted graph model [28,42], where complex network analysis applies to examine its topological and structural features. The network measures that are used for this part of the analysis are shown in Table 1, and they were extracted from the relevant literature [43-45].

Table 1. Network measures(*) used in the analysis

Measure	Symbol	Description	Math Formula
Graph density	ρ	The fraction of the existing connections of the graph to the number of possible connections. It expresses the probability to meet in the GMN a connected pair of nodes.	$\rho = m \Big/ \binom{n}{2} = \frac{2m}{n \cdot (n-1)}$
Node Degree	k	The number of edges that are adjacent to a given node, expressing the node's communication potential.	$k_i = k(i) = \sum_{j \in V(G)} \delta_{ij}$, where $\delta_{ij} = \begin{cases} 1, \text{ if } e_{ij} \in E(G) \\ 0, \text{ otherwise} \end{cases}$
Node strength	s	The sum of edge weights that are adjacent to a given node.	$s_i = s(i) = \sum_{j \in V(G)} \delta_{ij} \cdot w_{ij}$, where $d_{ij} = w(e_{ij})$ in km
Average path length	$\langle l \rangle$	Average length $d(i,j)$ of the total of network shortest paths.	$\langle l \rangle = \frac{\sum_{v \in V} d(v_i, v_j)}{n \cdot (n-1)}$
Clustering coefficient (*local*)	$C(i)$	Probability of meeting linked neighbors around a node, which is equivalent to the number of the node's connected neighbors $E(i)$ (i.e. the number of triangles), divided by the number of the total triplets shaped by this node, which equals to $k_i(k_i-1)$.	$C(i) = \frac{E(i)}{k_i \cdot (k_i - 1)}$
Modularity	Q	Objective function expressing the potential of a network to be subdivided into communities. In its mathematical formula, g_i is the community of node $i \in V(G)$, $[A_{ij} - P_{ij}]$ is the difference of the actual minus the expected number of edges falling between a particular pair of vertices $i,j \in V(G)$, and $\delta(g_i,g_j)$ is an indicator function returning 1 when $g_i=g_j$.	$Q = \frac{\sum_{i,j} [A_{ij} - P_{ij}] \cdot \delta(g_i, g_j)}{2m}$
Closeness centrality	CC	The inverse of total binary distance $d(i,j)$, computed on the shortest paths originating from a given node $i \in V(G)$ with destinations all the other nodes $j \in V(G)$ in the network. This measure expresses the node's reachability in terms of steps of separation.	$CC(i) = \left(\frac{1}{n-1} \cdot \sum_{j=1, i \neq j}^{n} d_{ij} \right)^{-1} = \left(\bar{d}_i \right)^{-1}$
Betweenness centrality	CB	The proportion of the (σ) shortest paths in the network that pass through a given node i.	$CB(i) = \sigma(i)/\sigma$

*. Sources: [43-45]

■ Community detection based on modularity optimization

At the third step of network analysis, the associated visibility graph is divided into connected communities by using the *modularity optimization algorithm* proposed by [46]. Modularity is an objective function quantifying a network's potential to be subdivided into communities [47]. A modularity optimization algorithm is a greedy approach that divides a graph into communities. This is possible under the criterion of maximizing connectivity within the communities (the intra-community connectivity), which results in the minimization of the connectivity between the communities (the inter-community connectivity), as expressed in the relation [47]:

$$\text{maximize } [Q \propto (m_{within\ communities} - m_{between\ communities})] \quad (5)$$

where Q is the modularity function and m is the number of network links.

The modularity optimization algorithm is applied in two steps [46,47]. At the first, each graph node is registered into a separate community. Next, network nodes are step-wisely being swept and placed to collective communities, whether the gain in the weighted modularity function ($Q=Q(w_{ij})$) of the initial graph is increased when a node is assigned into a collective community. At the second step, the collective communities are replaced by super-nodes and the process is repeated until the modularity function cannot increase any more [47,48]. For a thorough review, see the review article of [47].

3. Results and discussion

■ Statistical inference analysis

At the first part of the analysis, a statistical inference analysis is applied to the available levels of a-SiC:H conductivity activation energy (E_a) to detect whether they

are statistically different. Fig.3 shows the results of the confidence interval (CI) analysis that examines whether the maximum values of the conductivity activation energy are statistically different; for each hydrogen flow $E_a=f(Ts \mid H_f=i$, where i=0,9,14,20sccm).

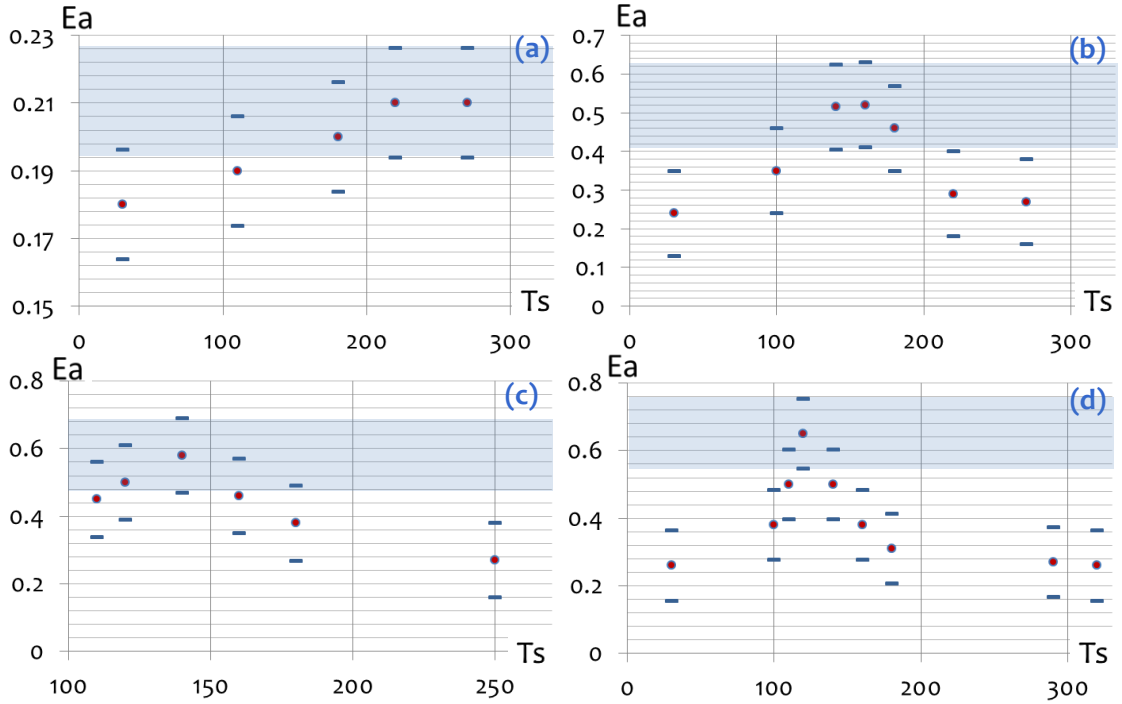

Fig. 3. Error bars of 95% CIs showing how the a-SiC:H conductivity activation energy (E_a) ranges as a function of temperature (T_s), for hydrogen flows (H_f) of (a) 0sccm, (b) 9sccm, (c) 14sccm, and (d) 20sccm. Coloured zones imply the range of the maximum values per hydrogen flow level (color online)

As it can be observed in Fig.3, for all cases except for H_f=0 the hypothesis is retained, indicating that the values of the a-SiC:H maximum conductivity activation energy $\{E_{a,9,max}, E_{a,14,max}$, and $E_{a,20,max}\}$ are statistically different than the other values for a certain hydrogen flow. This implies that, for zero hydrogen flow (H_f=0), the conductivity activation energy can be considered statistically indifferent to the temperature, which verifies the experimental hypothesis of [34] and [35], whereas, for non-zero hydrogen flows (H_f=9,14, and 20), it cannot (i.e. the conductivity activation energy can be considered statistically dependent to the temperature).

Further, Fig.4 shows the results of the 95% CIs computed on the average (Fig.4a) and maximum (Fig.4b) values of a-SiC:H conductivity activation energy. This analysis aims to examine whether these conductivity activation energy values are statistically different amongst the levels of hydrogen flows.

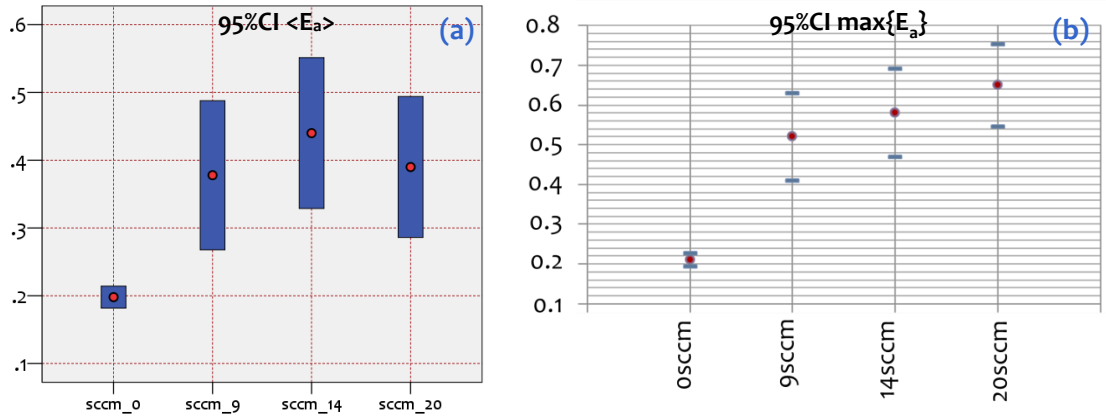

Fig. 4. Error bars of 95% CIs showing how the average a-SiC:H activation-energy (E_a) ranges as a function of temperature (T_s), for hydrogen flows (H_f) of (a) 0sccm, (b) 9sccm, (c) 14sccm, and (d) 20sccm. Coloured zones imply the range of the maximum values per hydrogen-flow level (color online)

As it can be observed, for both the average-value (Fig.4a) and maximum-value (Fig.4b) cases the zero-flow level (H_f=0) is statistically different than the others, but no such observation can be made for the other hydrogen flow levels because their CIs overlay. These results imply that the application of non-zero hydrogen flows can statistically change the levels of a-SiC:H conductivity activation energy (and thus the semiconductor's structure), comparatively to the hydrogen-free state, but the conductivity activation energy levels amongst non-zero hydrogen flows cannot be considered as statistically different.

■ Pattern recognition

At the second part of the analysis, pattern recognition based on parametric (curve) fitting is applied to the available levels of a-SiC:H activation energy (E_a). The results of this analysis are shown in Fig.5, where the best possible fittings that can be applied to the available dataset are the *linear* (Poly1) for zero hydrogen flow (H_f=0) level, the *two-term Gaussian* (Gauss2) for the 9sccm hydrogen flow (H_f=9) level, the *two-term exponential* (Exp2) for the 14sccm hydrogen flow (H_f=14) level, and the *two-term Gaussian* (Gauss2) for the 20sccm hydrogen flow (H_f=20) level. All best-fittings associated with non-zero hydrogen flow levels are convex-shaped patterns, which implies that a global maximum can be found in each case. This result complies with the experimental findings of [34] and [35], who observed the existence of a substrate temperature, which is a function of the hydrogen flow rate at which the optimum quality of a-SiC:H is achieved.

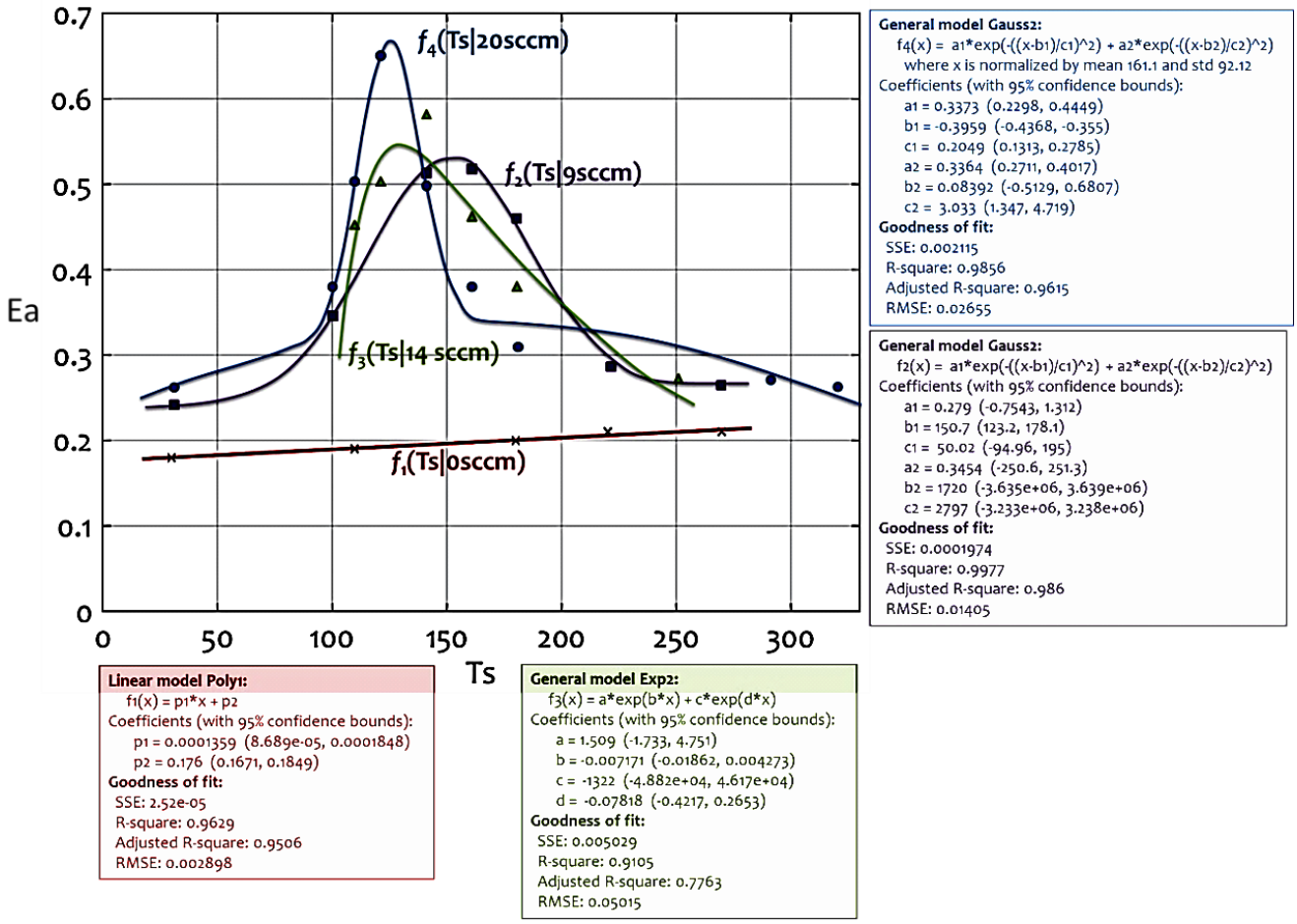

Fig. 5. Fitting curves applied to the a-SiC:H conductivity activation energy (E_a) dataset that is expressed as a function of temperature (T_s), for different levels of hydrogen flows (H_f) (color online)

At a further approach, pattern recognition based on parametric (curve) fitting is applied to the levels of a-SiC:H maximum conductivity activation energy ($E_{a,\max}$) that is expressed as a function of hydrogen flow $E_{a,\max}=f(H_f)$. The results of this analysis are shown in Fig.6, where the three best possible fittings that can be applied to the available dataset are the *linear* (Poly1), with an adjusted (adj.) coefficient of determination equal to $R^2_{Poly1,adj}$ = 0.883, the *Gaussian* (Gauss1), with $R^2_{Gauss1,adj}$ = 0.948, and the *quadratic* (Poly2), with $R^2_{Poly2,adj}$ = 0.984. As it can be observed, amongst these best fittings, the *Gaussian* (Gauss1) and *quadratic* (Poly2) have the highest coefficient of determination, which expresses that there is the most likely the conductivity activation energy function $E_{a,\max}=f(H_f)$ to have a convex shape. Within the context of insufficient information describing the available dataset,

this observation implies that the behavior of the a-SiC:H conductivity activation energy as a function of hydrogen flow is described by a global maximum point, which for both the Gauss1 and Poly2 fittings is defined for the 20sccm hydrogen flow. Although is not available from the experimental data, according to these fitting patterns, values of conductivity activation energy for hydrogen flows greater than 20sccm (H_f > 20sccm) are expected to correspond to lower levels of conductivity activation energy. This implies that higher than 20sccm hydrogen flow levels are the most likely to be described by lower values of conductivity activation energy than this corresponding to the hydrogen flow level of 20sccm and thus that is the most likely to be described by worse structures.

Finally, we apply curve fitting pattern recognition to the levels of the temperature of a-SiC:H maximum conductivity activation energy (T_{max}=T($E_{a,\max}$)), expressed as a function of hydrogen flow. The $T_{\max}$ is defined as the temperature where the maximum conductivity activation energy appears for the available levels of hydrogen flows {0, 9, 14, and 20 sccm}.

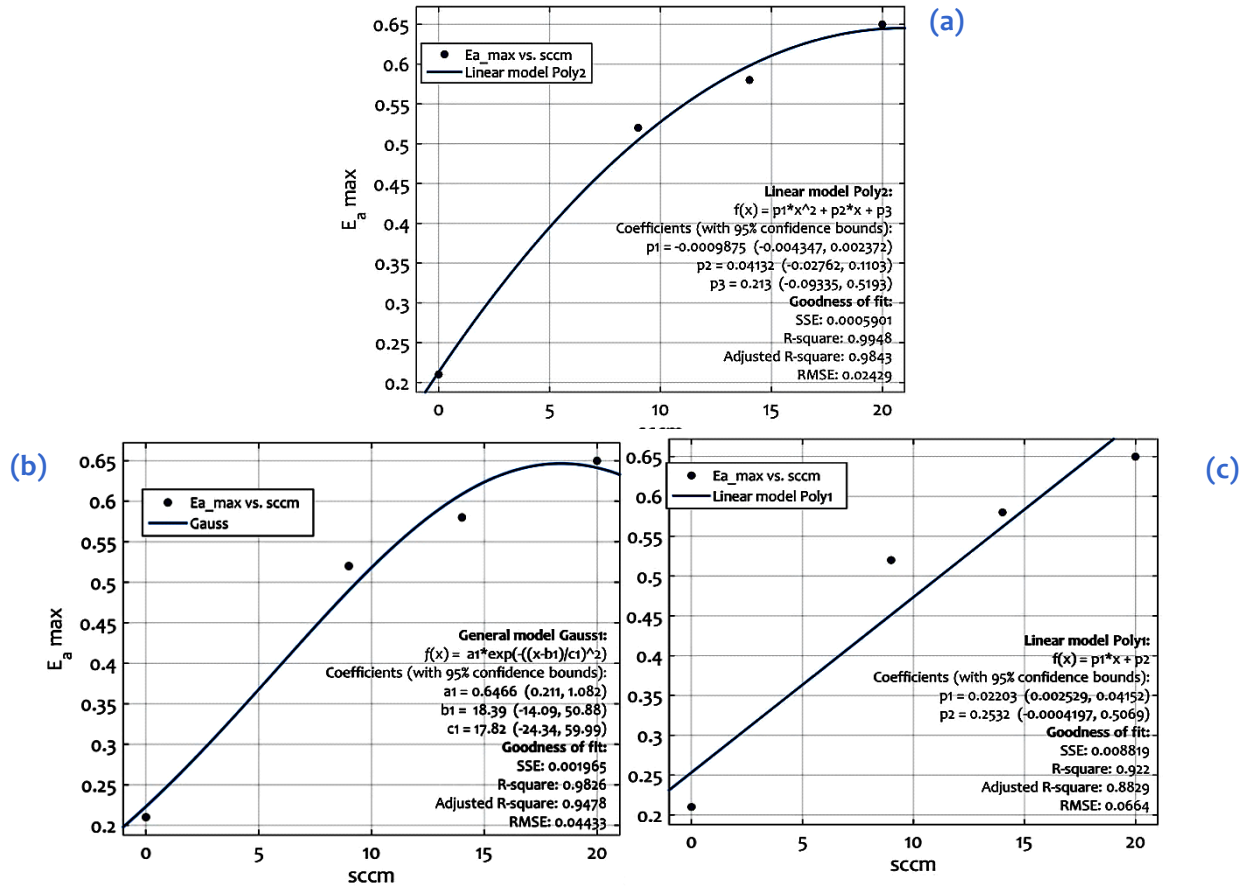

Fig. 6. (a) Quadratic (Poly2), (b) Gaussian (Gauss1), and (c) linear (Poly1) fitting curves applied to the a-SiC:H maximum conductivity activation energy ($E_{a,max}$) dataset that is expressed as a function of hydrogen flows (H_f) (color online)

The results of this analysis are shown in Fig.7, where four best possible fittings are detected, the *linear* (Poly1), with $R^2_{Poly1,adj}$ = 0.955, the *power* (Power1), with $R^2_{Power1,adj}$ = 0.934, the *exponential* (Exp1), with $R^2_{Poly2,adj}$ = 0.990, and the *quadratic* (Poly2), with $R^2_{Poly2,adj}$ = 0.998. As it can be observed, all these fittings illustrate a decaying pattern of the temperature of maximum conductivity activation energy ($T_{\max}$) as a function of hydrogen flow. This implies that the supply of hydrogen flow to the semiconductor body lowers the temperature at which the maximum conductivity activation energy appears and therefore better semiconductor structures are succeeded at lower temperatures. The shape of all fittings except the linear also implies that the temperature gain for hydrogen flows greater than 20sccm (H_f > 20sccm) is expected to be relatively lower than this of the previous states, which implies that better structures are unprofitable to achieve at high (H_f > 20sccm) hydrogen flows. In conjunction with the previous consideration, this observation describes the existence of optimum semiconductor structures at the level of 17-20sccm where the maxima of the bell-shaped curves appear.

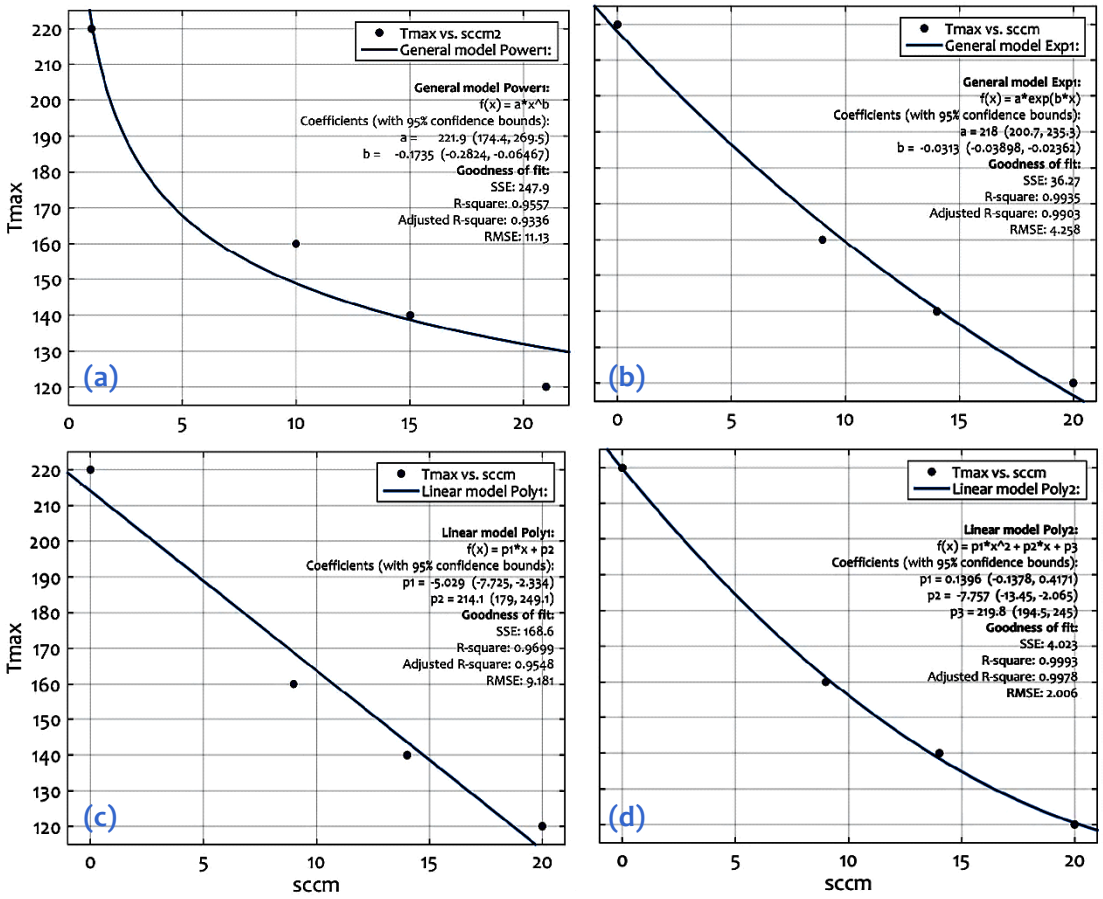

Fig. 7. (a) Power (Power1), (b) exponential (Exp1), (c) linear (Poly1), and (d) quadratic (Poly2) fitting curves applied to the temperature of a-SiC:H maximum conductivity activation energy (T_{max}=$T(E_{a,max})$) dataset, at which the maximum conductivity activation energy appears, expressed as a function of hydrogen flows (H_f) (color online)

■ Network analysis

In the third part, we apply complex network analysis to the graph models generated by the visibility graph (NVG) algorithm [32]. Although the NVG is defined within a time series context, this analysis broadens its applicability and applies the algorithm to discrete conductivity activation energy instead of time-series datasets. At the first step, the a-SiC:H conductivity activation energy data for each level of hydrogen flows are transformed into graph models, as shown in Fig.8. We apply a hub detection analysis to these graph models, conducted by node-degree (for the case of each separate layer) and node-strength (weighted-degree, for the aggregate case). Node labels in the visibility graphs correspond to temperature values (T_s). As is evident, the network-hubs (maximum-degree nodes) vary across different levels of hydrogen flows, where the temperatures T_s=110 and 220ºC are hubs for the 0sccm level, the temperature T_s=180ºC is a hub for the 9sccm level, the temperature T_s=140ºC is a hub for the 14sccm level, and the temperature T_s=120ºC is a hub for the 20sccm level. In the aggregate (overlaid) visibility graph, this perplex picture configures a rich-club [26,49,50] structure consisting of the temperatures T_s={120, 140, 160, 180}ºC, which represents a highest activation-energies zone. This rich-club configuration illustrates the temperature range of [120,180]ºC where better a-SiC:H structures can be developed regardless of the level of hydrogen flows.

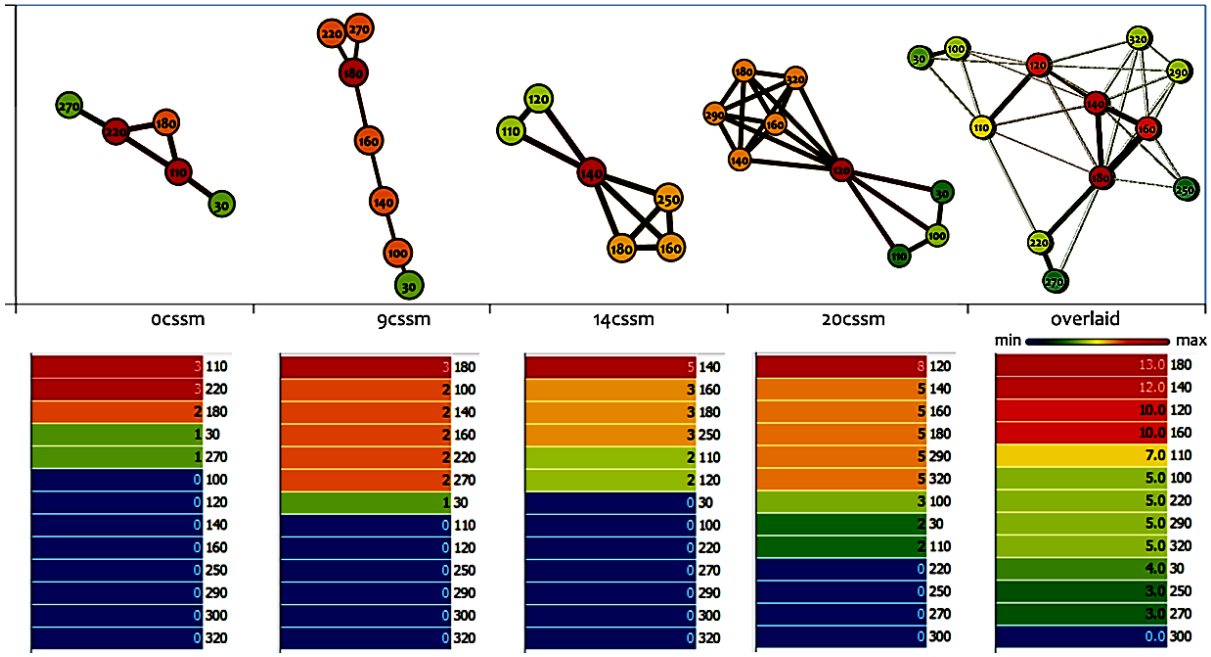

Fig. 8. Visibility graphs that are associated with the a-SiC:H conductivity activation energy (E_a) datasets for each level of hydrogen flows (0, 9, 14, and 20sccm) and the overlaid (aggregate) dataset including all hydrogen flow levels. Node labels correspond to temperature values (T_s) (color online)

Provided that the aggregate visibility graph can deal with a higher level of complexity, the second step of network analysis examines the node-distribution for network topological measures of degree, betweenness centrality, closeness centrality, clustering coefficient, strength, and modularity classification. The results of the analysis are shown in Fig.9. Within a degree-driven consideration expressed by the measures of degree (Fig.9a) and strength (Fig.9e), the rich-club configuration of the temperatures {120, 140, 160, 180}ºC implies a zone of better a-SiC:H structures regardless of the level of hydrogen flows. According to a path-defined consideration, as it is expressed by the measure of betweenness centrality (CB, Fig.9b), the temperature of 180ºC is a crucial node included in most of the network paths of the aggregated visibility graph. This observation implies that the temperature of 180ºC can suggest a threshold in the aggregate network's structure and secondly the temperature of 120ºC. This consideration supports the rich-club configuration that was previously observed and it further supports the division of the available dataset into the temperature groups T_A={30, 100, 110}ºC, T_B={120, 140, 160, 180}ºC, and T_C={220, 250, 270, 290, 320}ºC. This grouping can also be supported by the results of closeness centrality (Fig.9c) and clustering coefficient (Fig.9d).

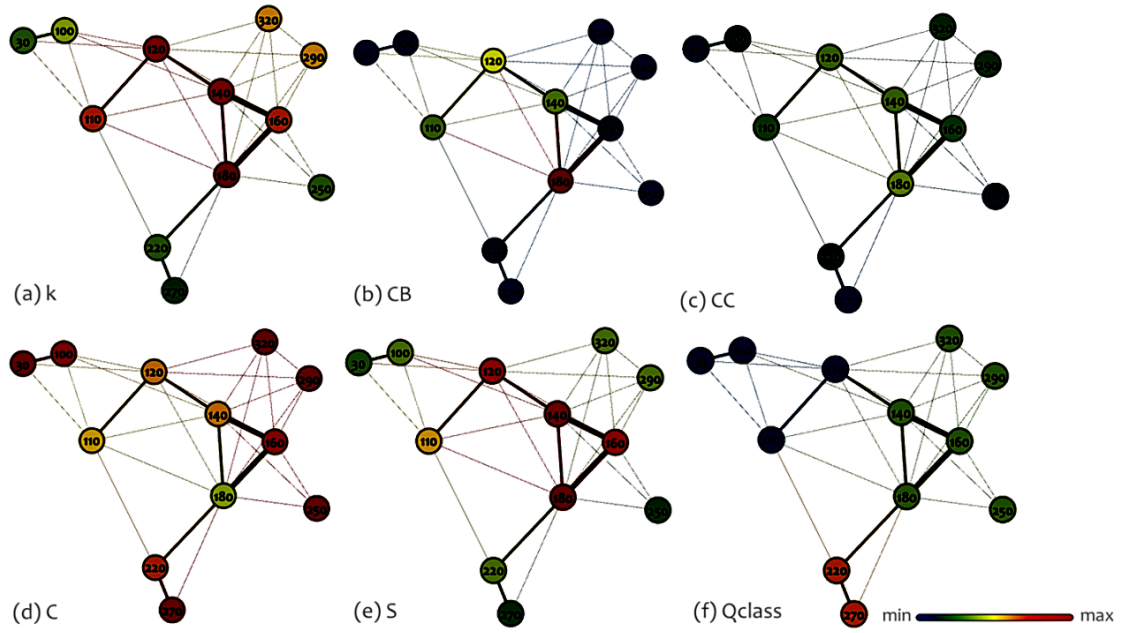

Fig. 9. Network measures of the aggregate (overlaid) conductivity activation energy (Ea) Visibility Graph (color online)

Finally, the modularity classification node-distribution (Fig.9f) illustrates the existence of three connected communities, namely the community Q_1={30, 100, 110, 120}ºC, Q_2={140, 160, 180, 250, 290, 320}ºC, and Q_3={220, 270}ºC. The configuration of the third (Q_3) community is probably related to the availability of measurements at the 220 and 270ºC temperatures, and for the 0sccm and 9cssm hydrogen flow levels; therefore is not that insightful. However, the division of the aggregate network into the communities Q_1, Q_2, and Q_3 is insightful in revealing the "gate" (or transitive) role of the temperatures 120ºC and 180ºC in the T_B={120, 140, 160, 180}ºC group's configuration, which implies that these temperatures can operate as gates (or thresholds) of different states in the levels of conductivity activation energy. The overall network analysis approach illustrates the existence of three distinct conductivity activation energy states (defined by the T_A, T_B, and T_C groups), which correspond to different structural (semiconductor) behaviors of the a-SiC:H thin-film alloys, and instructs seeking for optimum structure within the temperature range of {120, 140, 160, 180}ºC.

4. Conclusions

This paper applied a multilevel analysis using methods of statistical physics to detect structures of semiconductor a-SiC:H thin-film alloys with the best possible electrical performance expressed as a function of the temperature and hydrogen flow. The analysis is built on statistical physics and particularly on statistical and complex network analysis. On the one hand, the statistical inference analysis revealed that: at zero hydrogen flow levels, the conductivity activation energy is statistically indifferent to the temperature, whereas, for non-zero hydrogen flows, temperature affects the semiconductor's structure. Also, the application of non-zero hydrogen flows can statistically change the levels of conductivity activation energy. However, variations amongst conductivity activation energy levels cannot be considered statistically significant. In a pattern recognition approach, the supply of hydrogen flow to the semiconductor body led to better semiconductor structures at lower temperatures, where a zone of 17-20sccm appeared with better activation energy levels. On the other hand, the network analysis based on the visibility graph transformation revealed a rich-club configuration at a temperature range [120,180]ºC and three distinct conductivity activation energy levels, which correspond to different structural behaviors of the semiconductor a-SiC:H thin-film alloys. The overall analysis provided insights of dealing with multivariate structural analysis, within the context of insufficient information.

References

[1] M. S. Gudiksen, J. Wang, C. M. Lieber, The Journal of Physical Chemistry B **105**(19), 4062 (2001).
[2] J. Takeya, J. Kato, K. Hara, M. Yamagishi, R. Hirahara, K. Yamada, T. Takenobu, Physical Review Letters **98**(19), 196804 (2007).
[3] O. Sneh, R. B. Clark-Phelps, A. R. Londergan, J. Winkler, T. E. Seidel, Thin Solid Films **402**(1-2), 248 (2002).
[4] L. Yu, R. S. Kokenyesi, D. A. Keszler, A. Zunger, Advanced Energy Materials **3**(1), 43 (2013).
[5] K. L. Chopra, P. D. Paulson, V. Dutta, Progress in Photovoltaics: Research and applications **12**(2-3), 69 (2004).
[6] G. Lucovsky, J. Yang, S. S. Chao, J. E. Tyler, W. Czubatyj, Physical Review B **28**(6), 3225 (1983).
[7] J. Henry, J. Livingstone, Advanced Materials **13**(12-13), 1022 (2001).
[8] E. Vallat-Sauvain, U. Kroll, J. Meier, A. Shah, J. Pohl, Journal of Applied Physics **87**(6), 3137 (2000).
[9] H. Jung, M. Park, S. H. Han, H. Lim, S. K. Joo, Solid State Communications **125**(7-8), 387 (2003).
[10] K. Tanaka, K. Shimakawa, Amorphous chalcogenide semiconductors and related materials, Springer Science & Business Media, 2011.
[11] A. Kumbhar, S. B. Patil, S. Kumar, R. A. K. E. S. H. Lal, R. O. Dusane, Thin Solid Films **395**(1-2), 244 (2001).
[12] R. Saleh, L. Munisa, W. Beyer, Thin Solid Films **426**(1-2), 117 (2003).
[13] P. Girard, Nanotechnology **12**(4), 485 (2001).
[14] J. S. Dugdale, The electrical properties of metals and alloys, Courier Dover Publications, 2016.
[15] J. P. McKelvey, Solid state and semiconductor physics, Harper & Row, 2018.
[16] M. Cardona, Y. Y. Peter, Fundamentals of semiconductors. Springer-Verlag Berlin Heidelberg, 2005.
[17] A. Miller, D. A. Miller, S. D. Smith, Advances in Physics **30**(6), 697 (1981).
[18] L. L. Bonilla, H. T. Grahn, Reports on Progress in Physics **68**(3), 577 (2005).
[19] J. Peinke, J. Parisi, O. E. Rossler, R. Stoop, Encounter with chaos: self-organized hierarchical complexity in semiconductor experiments, Springer Science & Business Media, 2012.
[20] K. A. Shore, Solid-State Electronics **30**(1), 59 (1987).
[21] M. P. Hanias, A. N. Anagnostopoulos, Physical Review B **47**(8), 4261 (1993).
[22] M. P. Hanias, J. A. Kalomiros, Ch. Karakotsou, A. N. Anagnostopoulos, J. Spyridelis, Physical Review B **49**(24), 16994 (1994).
[23] S. Kumar, P. R. Kumar, IEEE Transactions on Robotics and Automation **17**(5), 548 (2001).
[24] L. Zhang, W. Pan, L. Yan, B. Luo, X. Zou, M. Xu, IEEE Journal of Selected Topics in Quantum Electronics **25**(6), 1 (2019).
[25] U. Brandes, G. Robins, A. McCranie, S. Wasserman, Network Science **1**, 1 (2013).
[26] A. L. Barabasi, Network Science, Cambridge, UK,

Cambridge University Press, 2016.
[27] Z-K. Gao, Y-X. Yang, P-C. Fang, Y. Zou, C-Y. Xia, M. Du, Europhysics Letters 109, 30005 (2015).
[28] D. Tsiotas, A. Charakopoulos, Physica A **505**, 280 (2018).
[29] J. Zhang, M. Small, Physical review letters **96**(23), 238701 (2006).
[30] Y. Yang, H. Yang, Physica A: Statistical Mechanics and its Applications **387**(5), 1381 (2008).
[31] X. Xu, J. Zhang, M. Small, Proceedings of the National Academy of Sciences **105**(50), 19601 (2008).
[32] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. C. Nuno, Proceedings of the National Academy of Sciences **105**(13), 4972 (2008).
[33] B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Physical Review E **80**(4), 046103 (2009).
[34] L. Magafas, D. Bandekas, A. K. Boglou, A. N. Anagnostopoloulos, Journal of Non-Crystalline Solids **353**(11-12), 1065 (2007).
[35] L. Magafas, J. Kalomiros, D. Bandekas, G. Tsirigotis, Microelectronics journal **37**(11), 1352 (2006).
[36] S. J. Pennycook, P. D. Nellist (Eds.), Scanning transmission electron microscopy: imaging and analysis, Springer Science & Business Media, 2011.
[37] J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. Ritchie, J. H. J. Scott, D. C. Joy, Scanning electron microscopy and X-ray microanalysis, Springer, 2017.
[38] T. Carlson, Photoelectron and Auger spectroscopy, Springer Science & Business Media, 2013.
[39] P. B. Amama, C. Lan, B. A. Cola, X. Xu, R. G. Reifenberger, T. S. Fisher, The Journal of Physical Chemistry C **112**(49), 19727 (2008).
[40] M. Norusis, SPSS 16.0 Statistical Procedures Companion, New Jersey (USA), Prentice Hall Publications, 2008.
[41] R. E. Walpole, R. H. Myers, S. L. Myers, K. Ye, Probability & Statistics for Engineers & Scientists, ninth ed. New York, USA: Prentice Hall Publications, 2012.
[42] D. Tsiotas, A. Charakopoulos, Software X **11**, 100379 (2020).
[43] D. Koschutzki, K. Lehmann, L. Peeters, S. Richter, Centrality indices. In: Brandes, U., Erlebach, T., (eds) Network analysis, Berlin, Springer-Verlag Publications, 16 (2005).
[44] M. E. J. Newman. Networks: An Introduction, Oxford (UK), Oxford University Press, 2010.
[45] M. Barthelemy, Physics Reports **499**, 1 (2011).
[46] V. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Journal of Statistical Mechanics **10**, P10008 (2008).
[47] S. Fortunato, Physics Reports **486**, 75 (2010).
[48] D. Tsiotas, Proceedings of the National Academy of Sciences **116**(14), 6701 (2019).
[49] V. Colizza, A. Flammini, M. A. Serrano, A. Vespignani, Nature Physics **2**(2), 110 (2006).
[50] D. Tsiotas, Scientific Reports **10**(1), 10630(2020).

*Corresponding authors: tsiotas@aua.gr; lmagafas@otenet.gr