



DOI 10.22363/2312-8143-2022-23-1-7-14
УДК 621.382:538.91

Научная статья / Research article


Критерий оценки нейросетевой модели гетероструктурных нанoeлектронных устройств для прогнозирования их электрических параметров

Н.А. Ветрова^{a,b} , А.А. Филяев^{a,c}  

^aМосковский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Москва, Российская Федерация

^bРоссийский университет дружбы народов, Москва, Российская Федерация

^cНациональный исследовательский технологический университет «МИСиС», Москва, Российская Федерация

 alex.filyaev.98@gmail.com

История статьи

Поступила в редакцию: 22 февраля 2022 г.

Доработана: 15 апреля 2022 г.

Принята к публикации: 29 апреля 2022 г.

Ключевые слова:

нанoeлектроника, критерий, искусственные нейронные сети, гетероструктурные устройства

Аннотация. Работа посвящена нейросетевому подходу, который предлагается использовать для прогнозирования эксплуатационных параметров гетероструктурных наноразмерных устройств различного назначения. Его преимуществом является эффективная методика оценки весовых коэффициентов в составе обучаемой искусственной нейронной сети, что позволяет решать задачу для устройств с произвольной структурой. Обучение представляет собой сложный итерационный процесс, по окончании которого важно производить оценку работы нейросетевой модели. Поэтому после построения такой модели необходимо определить достигаемую точность, а также выявить негативные эффекты, которые могут возникнуть в процессе обучения, в частности переобучение и недообучение сети. Представлен критерий оценки качества обучения нейросетевой модели гетероструктурных нанoeлектронных устройств для прогнозирования их электрических параметров. Основное преимущество данного критерия – его чувствительность к негативным эффектам, возникающим в процессе обучения, что было продемонстрировано на примере с двумя входными обучающими параметрами и подтверждено визуальным контролем 3D-поверхностей. Доказана применимость разработанного критерия при выборе нейронных сетей с произвольной архитектурой для решения конструкторских задач при проектировании полупроводниковых приборов.

Для цитирования

Ветрова Н.А., Филяев А.А. Критерий оценки нейросетевой модели гетероструктурных нанoeлектронных устройств для прогнозирования их электрических параметров // Вестник Российского университета дружбы народов. Серия: Инженерные исследования. 2022. Т. 23. № 1. С. 7–14. <http://doi.org/10.22363/2312-8143-2022-23-1-7-14>




Evaluation criterion of the neural network model of heterostructural nanoelectronic devices for predicting their electrical parameters

Natalia A. Vetrova^{a,b} , Alexandr A. Filyaev^{a,c}  

^aBauman Moscow State Technical University (National Research University of Technology), Moscow, Russian Federation

^bPeoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation

^cNational University of Science and Technology "MISIS," Moscow, Russian Federation

 alex.filyaev.98@gmail.com

Article history

Received: February 22, 2022

Revised: April 15, 2022

Accepted: April 29, 2022

Keywords:

nanoelectronics, criterion, artificial neural networks, heterostructural devices

Abstract. The paper is devoted to the neural network approach, which is proposed to be used to predict the operational parameters of heterostructural nanoscale devices. The advantage of this approach is a clear methodology for evaluating the weighting coefficients as part of a trained artificial neural network, which makes it possible to solve the problem for devices with an arbitrary structure. Learning is a complex iterative process, at the end of which it is important to evaluate the functioning of the neural network model. Therefore, it is necessary to determine the achieved accuracy and to identify negative effects that may occur during the learning process, when such a model is being developed. The project presents a criterion for evaluation the training quality of the neural network model of heterostructural nanoelectronic devices for predicting their electrical parameters. The main advantage of this criterion is its sensitivity to negative effects arising in the learning process, which was demonstrated by an example with two input training parameters and confirmed by visual control of 3D surfaces. The applicability of the developed criterion in the selection of neural networks with arbitrary architecture for solving design problems in the development of semi-conductor devices has been proved.

For citation

Vetrova NA, Filyaev AA. Evaluation criterion of the neural network model of heterostructural nanoelectronic devices for predicting their electrical parameters. *RUDN Journal of Engineering Research*. 2022;23(1):7–14. (In Russ.) <http://doi.org/10.22363/2312-8143-2022-23-1-7-14>

Введение

Многообразие перспективных сфер применения ставит перед разработчиками задачу проектирования широкого класса наноразмерных гетероструктурных устройств с различной сложностью низкоразмерного канала в их составе [1–6]. При проектировании гетероструктурных наноприборов очевидна необходимость моделирования их электрических параметров. Применяемые в мировой практике модели носят полуэмпирический характер с рядом допущений и набором поправочных коэффициентов без четкой методики их оценки, что не позволяет решить задачу для произвольной структуры [7–8]. В качестве решения проблемы моделирования электрических параметров разнообразных гетероструктурных устройств предлагается использовать нейросетевой подход [9–10].

Важнейшим этапом при построении модели на основе искусственных нейронных сетей является оценка работы нейронной сети с целью определения достигаемой точности и выявления негативных эффектов недообучения и переобучения нейронной сети [11–13]. При использовании большого числа входных параметров обучения нейросетевой модели проведение такой оценки усложняется, что ставит перед разработчиками задачу формирования универсального критерия выбора нейронной сети.

1. Процесс обучения нейронной сети

Формирование критерия оценки нейронной сети невозможно без однозначного понимания процедуры ее обучения и природы возникновения негативных эффектов при данном процессе. Рассмотрим его подробнее.

Перед началом обучения формируется архитектура нейронной сети, которая методически подбирается для решения конкретной задачи. Затем начинается непосредственно процесс обучения. Существует два подхода к обучению нейронных сетей – неконтролируемый и контролируемый [14].

При неконтролируемом обучении разработчик предоставляет лишь входные данные для обучения, без желаемых выходных параметров. Затем нейросетевая система должна самостоятельно выявить зависимости (процесс самоорганизации), которые она будет в дальнейшем использовать для обработки этого набора данных.

Такая система применяется для анализа и группировки наборов данных в рамках решения трех основных задач: кластеризации, ассоциации и снижения размерности.

Контролируемое обучение подразумевает предоставление разработчиком набора как входных, так и выходных данных для обучения. Далее нейросетевая система обрабатывает векторы входных параметров и сопоставляет их с выходными параметрами, постепенно обновляя весовые коэффициенты и значение смещения. Итерационный процесс контролируемого обучения представлен на рис. 1.

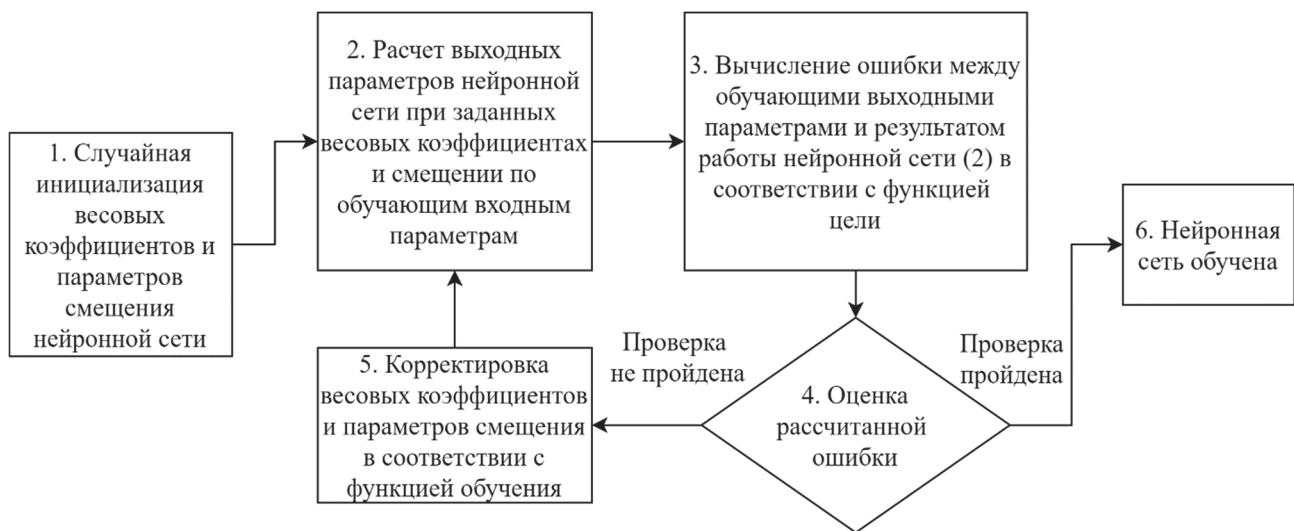


Рис. 1. Итерационный процесс контролируемого обучения нейронной сети

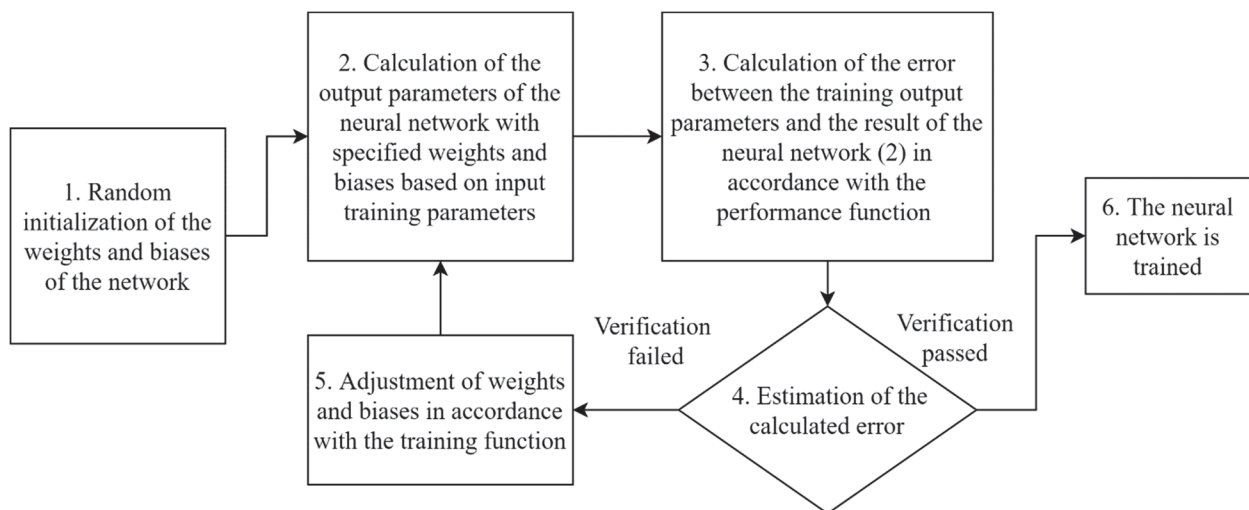


Figure 1. Iterative process of supervised learning of a neural network

Такой подход позволяет решать задачу прогнозирования, что применимо для построения предикторов электрических параметров устройств нанoeлектроники.

В результате неудачного выбора архитектуры нейронной сети (количество скрытых слоев, число нейронов в каждом скрытом слое) можно столкнуться с негативными эффектами, проявляющимися после ее обучения: недообучения и переобучения. При недообучении нейросетевая модель не способна точно отразить взаимосвязь между входными и выходными переменными, что приводит к высокой частоте ошибок между обучающими выходными параметрами и результатом работы нейронной сети. Это происходит, когда модель обладает слишком простой архитектурой. Переобучение представляет собой противоположный эффект, возникающий при использовании излишне усложненной нейросетевой архитектуры. В таком случае модель не приобретает способности к обобщению – возможности распространять обнаруженные на обучающем множестве зависимости и закономерности на новые данные.

2. Критерий оценки нейросетевой модели

Для оценки точности нейронной сети, а также выявления негативных эффектов при ее обучении предлагается введение расчетного параметра σ , который вычисляется как

$$\sigma(X, Y) = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \cdot 100, \quad (1)$$

где X – массив данных, полученных в результате работы обученной нейронной сети, с элементами x_i в своем составе; Y – массив выходных данных, предназначенных для обучения (контроля) нейронной сети, с элементами y_i в своем составе; n – количество элементов в массивах X , Y .

Для выявления процессов недообучения, исходя из природы данного эффекта, целесообразно проводить сравнение результатов работы нейросетевой модели с обучающими выходными параметрами. В таком случае значение параметра σ более 1 % будет индикатором выявления данного негативного эффекта.

Для выявления процессов переобучения следует заранее сформировать отдельную выборку контрольных выходных значений (например, промежуточные точки относительно обучающей вы-

ходной выборки), так как очевидно, что значение параметра σ относительно обучающей выходной выборки при переобучении примет минимальное значение, а это не позволит выявить данный эффект. В таком случае значение параметра σ относительно контрольной выборки выходных значений более 1 % будет индикатором выявления переобучения нейросетевой модели.

3. Результаты и обсуждение

Для моделирования резонансно-туннельного диода (РТД) в составе умножителя сигнала предложены три нейросетевые модели различной архитектуры (таблица). С применением упрощенной квантово-механической модели (КММ) на основе метода неравновесных функций Грина [15–16] получены обучающие выборки, включающие в себя матрицу входных значений, где в качестве входных параметров выступают ширина потенциальной ямы и величина прикладываемого напряжения, а в качестве выходного параметра выступает рассчитанное значение плотности тока. Диапазон значений входного обучающего вектора напряжения обусловлен интересом к начальному участку вольт-амперной характеристики РТД при дальнейшей интеграции РТД в состав умножителя сигнала. Диапазон значений входного обучающего вектора ширины ямы выбирался для типичных конструкций РТД, шаг обусловлен шириной монослоя. Формирование контрольных входных векторов производилось на промежуточных точках по напряжению и ширине потенциальной ямы.

Для оценки обученных нейросетевых моделей применим критерий, описанный в разделе 2 (рис. 2–3). У нейронной сети № 2 выявлен эффект недообучения ввиду превышающего 1 % значения параметра σ относительно обучающей выходной выборки, что можно наблюдать на поверхности, представленной на рис. 2, б. Действительно, эта нейронная сеть обладает слишком простой архитектурой, что предполагает проявление данного негативного эффекта при обучении.

У нейронной сети № 3, напротив, параметр σ относительно контрольной выходной выборки превышает 1 %, что индицирует процесс переобучения (рис. 3, б), который также подтверждается чрезмерной сложностью архитектуры этой нейросетевой модели для данной регрессионной задачи. Нейронная сеть № 1 обладает приемлемыми (менее 1 %) значениями параметров σ , что говорит о ее успешном обучении (рис. 2, а и 3, а).

Архитектура рассматриваемых нейросетевых моделей

Номер нейронной сети	Число скрытых слоев	Функция активации в каждом скрытом слое	Число нейронов в каждом скрытом слое
1	2	Гиперболический тангенс	[10,5]
2	2	Гиперболический тангенс	[1,3]
3	2	Гиперболический тангенс	[100,25]

Architecture of the considered neural network models

No. of neural network	Number of hidden layers	Activation function in each hidden layer	Number of neurons in each hidden layer
1	2	Hyperbolic tangent	[10,5]
2	2	Hyperbolic tangent	[1,3]
3	2	Hyperbolic tangent	[100,25]

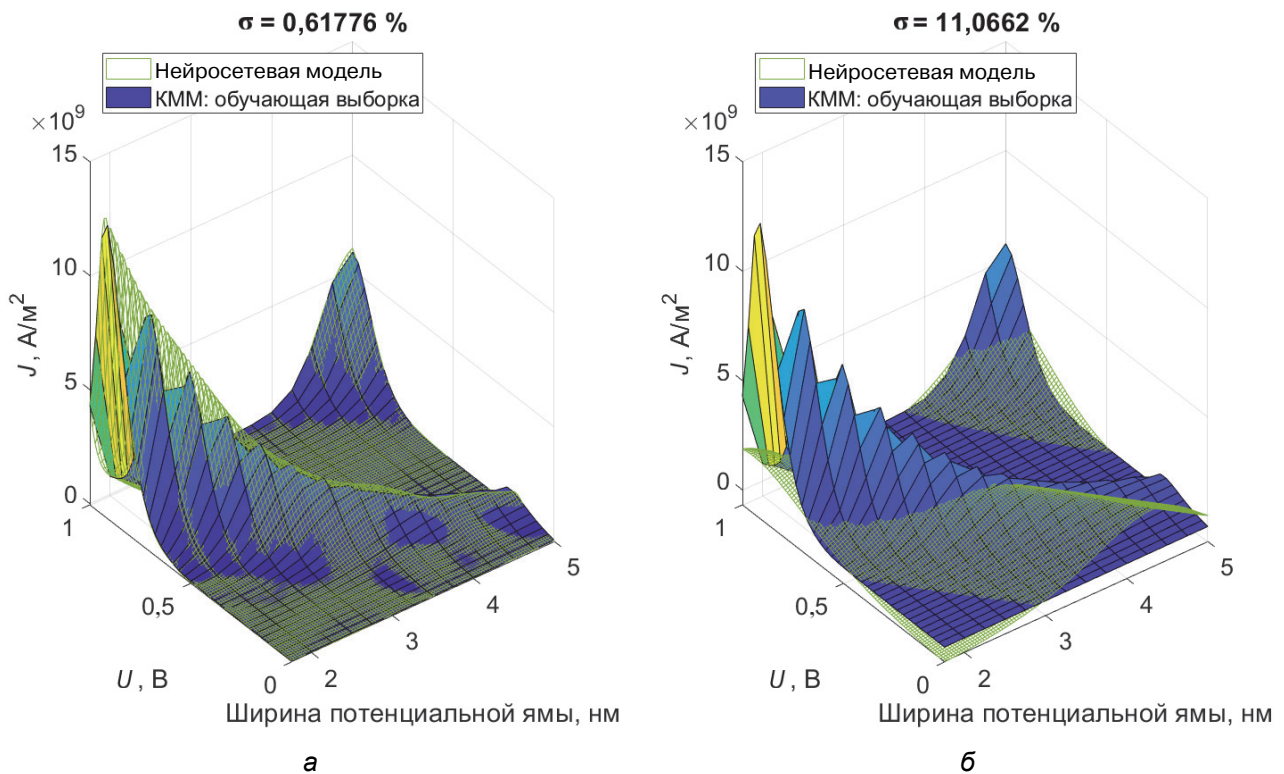


Рис. 2. Выявление эффекта недообучения нейросетевой модели:
а – нейронная сеть № 1; б – нейронная сеть № 2

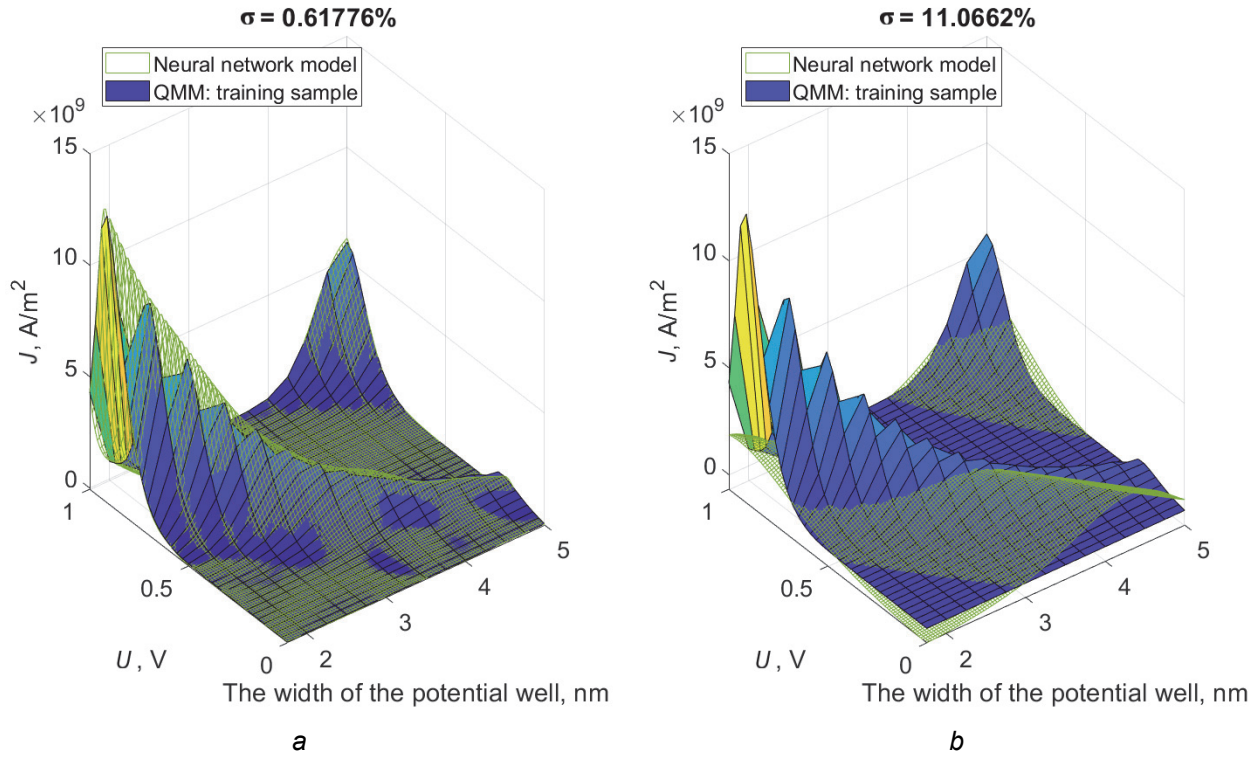


Figure 2. The underfitting detection of the neural network model:
a – neural network No. 1; b – neural network No. 2

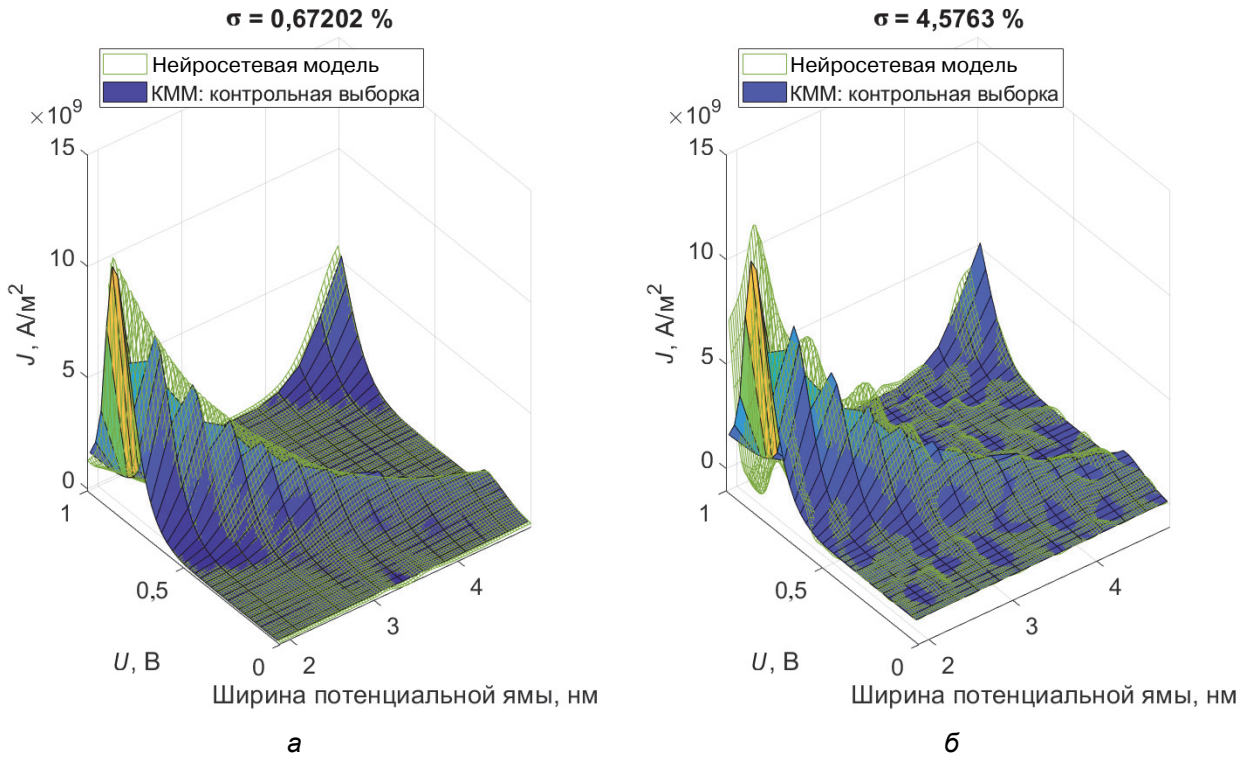


Рис. 3. Выявление эффекта переобучения нейросетевой модели:
а – нейронная сеть № 1; б – нейронная сеть № 3

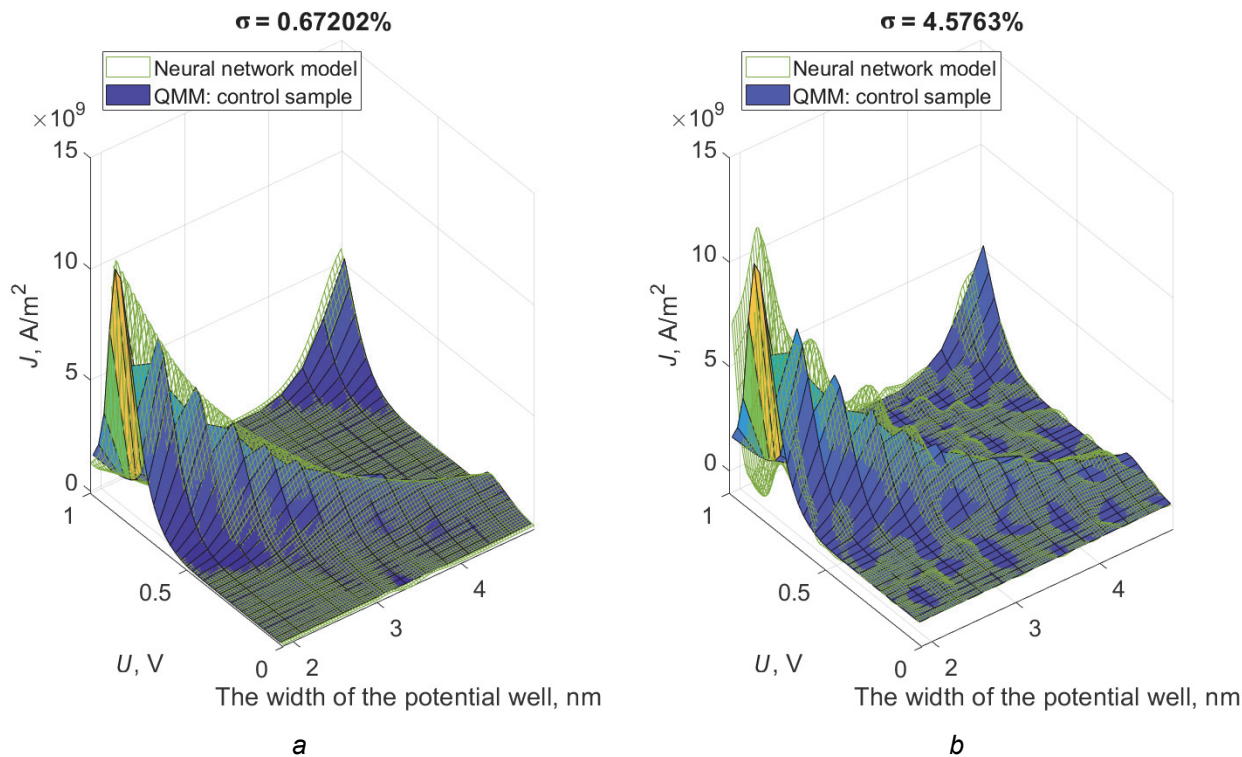


Figure 3. The overfitting detection of the neural network model:
 a – neural network No. 1; b – neural network No. 3

На представленных нейросетевых моделях продемонстрирована корректность работы критерия их оценки, что в рассмотренном примере с двумя входными обучающими параметрами возможно подтвердить визуальным контролем 3D-поверхностей.

Заключение

Разработан критерий оценки качества обучения нейросетевых моделей гетероструктурных нанoeлектронных устройств для прогнозирования их электрических параметров. Данный критерий позволяет не только определить точность таких моделей, но и выявить негативные эффекты, проявляющиеся в процессе их обучения. Проведенный анализ упрощенных нейросетевых моделей, обученных на двух входных параметрах, доказывает применимость критерия для оценки нейронных сетей с более сложной архитектурой, которые позволят решить задачи проектирования полупроводниковых нанoeлектронных устройств произвольной сложности.

Список литературы / References

1. Manh LD, Diebold S, Nishio K, Nishida Y, Kim J, Mukai T, Fujita M, Nagatsuma T. External feedback effect in terahertz resonant tunneling diode oscillators. *IEEE Trans-*

actions on Terahertz Science and Technology. 2018;8(4): 455–464. <https://doi.org/10.1109/TTHZ.2018.2842209>

2. Kasagi K, Suzuki S, Asada M. Large-scale array of resonant-tunneling-diode terahertz oscillators for high output power at 1 THz. *Journal of Applied Physics.* 2019;125(15):151601. <https://doi.org/10.1063/1.5051007>

3. Sattari-Esfahlan SM. Multi-peak negative differential resistance in silicene superlattice: Toward multi-valued silicene logic devices. *Journal of Applied Physics.* 2018;123(24):244503. <https://doi.org/10.1063/1.5032122>

4. Fujita K, Jung S, Jiang Y, Kim JH, Nakanishi A, Ito A, Hitaka M, Edamura T, Belkin MA. Recent progress in terahertz difference-frequency quantum cascade laser sources. *Nanophotonics.* 2018;7(11):1795–1817. <https://doi.org/10.1515/nanoph-2018-0093>

5. Rakić AD, Taimre T, Bertling K, Lim YL, Dean P, Valavanis A, Indjin D. Sensing and imaging using laser feedback interferometry with quantum cascade lasers. *Applied Physics Reviews.* 2019;6(2):021320. <https://doi.org/10.1063/1.5094674>

6. Bosco L, Franckić M, Scaliari G, Beck M, Wacker A, Faist J. Thermoelectrically cooled THz quantum cascade laser operating up to 210 K. *Applied Physics Letters.* 2019;115(1):010601. <https://doi.org/10.1063/1.5110305>

7. de Oliveira ERC, Pfenning A, Castro EDG, Teodoro MD, dos Santos EC, Lopez-Richard V, Marques GE, Worschech L, Hartmann F, Höfling S. Electroluminescence on-off ratio control of $n-i-n$ GaAs/AlGaAs-based

resonant tunneling structures. *Physical Review B*. 2018; 98(7):075302. <https://doi.org/10.1103/PhysRevB.98.075302>

8. Sawai S, Narahara K. Submillimeter-wave multiphase oscillation using traveling pulses in a resonant-tunneling diode-oscillator lattice. *Journal of Infrared, Millimeter, and Terahertz Waves*. 2021;42(4):426–445. <https://doi.org/10.1007/s10762-021-00780-z>

9. Pchelintsev KP, Vetrova NA, Shashurin VD. Modeling of the degradation of resonant-tunneling diodes using artificial neural networks. *Journal of Surface Investigation: X-ray, Synchrotron and Neutron Techniques*. 2022;16(1): 82–85. <https://doi.org/10.1134/S102745102201013X>

10. Vetrova NA, Pchelintsev KP, Shashurin VD. An artificial neural network as a predictor of electrical characteristics of nanoelectronic device channel based on a low-dimensional heterostructure. *Journal of Physics: Conference Series*. 2020;1695(1):012152. <https://doi.org/10.1088/1742-6596/1695/1/012152>

11. Koehrsen W. *Overfitting vs. underfitting: a complete example*. Toronto: Towards Data Science; 2018.

12. Gavrilov AD, Jordache A, Vasdani M, Deng J. Preventing model overfitting and underfitting in convolutional neural networks. *International Journal of Software Science and Computational Intelligence*. 2018;10(4):19–28. <https://doi.org/10.4018/IJSSCI.2018100102>

13. Narayan S, Tagliarini G. An analysis of underfitting in MLP networks. *2005 IEEE International Joint Conference on Neural Networks*. 2005;2:984–988. <https://doi.org/10.1109/IJCNN.2005.1555986>

14. Demuth HB, Beale MH, De Jess O, Hagan MT. *Neural network design*. Oklahoma: Martin Hagan; 2014.

15. Anantram MP, Lundstrom MS, Nikonov DE. Modeling of nanoscale devices. *Proceedings of the IEEE*. 2008; 96(9):1511–1550. <https://doi.org/10.1109/JPROC.2008.927355>

16. Datta S. *Quantum transport: atom to transistor*. Cambridge University Press; 2005.

Сведения об авторах

Ветрова Наталья Алексеевна, кандидат технических наук, доцент кафедры «Технологии приборостроения», Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1; доцент кафедры нанотехнологий и микросистемной техники, Инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 6. ORCID: 0000-0002-6218-4111, Scopus Author ID: 57197737296, eLIBRARY SPIN-код: 1555-2871. E-mail: vetrova@bmstu.ru

Филяев Александр Александрович, магистрант, кафедра «Технологии приборостроения», Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1; инженер научного проекта, лаборатория квантовых коммуникаций, Национальный исследовательский технологический университет «МИСиС», Российская Федерация, 119049, Москва, Ленинский пр-кт, д. 4. ORCID: 0000-0001-7319-8001, Scopus Author ID: 57218269710, eLIBRARY SPIN-код: 9454-4619. E-mail: alex.filyaev.98@gmail.com

About the authors

Natalia A. Vetrova, Candidate of Technical Sciences, Associate Professor of the Department of Instrument Engineering Technologies, Bauman Moscow State Technical University (National Research University of Technology), 5 2-ya Bauman-skaya St, bldg 1, Moscow, 105005, Russian Federation; Associate Professor of the Department of Nanotechnology and Microsystems Engineering, Academy of Engineering, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation. ORCID: 0000-0002-6218-4111, Scopus Author ID: 57197737296, eLIBRARY SPIN-code: 1555-2871. E-mail: vetrova@bmstu.ru

Alexandr A. Filyaev, master student, Department of Instrument Engineering Technologies, Bauman Moscow State Technical University (National Research University of Technology), 5 2-ya Baumanskaya St, bldg 1, Moscow, 105005, Russian Federation; engineer of the scientific project, National Technology Initiative Center for Quantum Communications, National University of Science and Technology "MISIS," 4 Leninskii Prospekt, Moscow, 119049, Russian Federation. ORCID: 0000-0001-7319-8001, Scopus Author ID: 57218269710, eLIBRARY SPIN-code: 9454-4619. E-mail: alex.filyaev.98@gmail.com