




Impact of neural cyberattacks on a realistic neuronal topology from the primary visual cortex of mice

Victoria Magdalena López Madejska¹ · Sergio López Bernal¹  · Gregorio Martínez Pérez¹ · Alberto Huertas Celdrán²

Accepted: 22 December 2023 / Published online: 25 January 2024
© The Author(s) 2024

Abstract

Brain-computer interfaces (BCIs) are widely used in medical scenarios to treat neurological conditions, such as Parkinson's disease or epilepsy, when a pharmacological approach is ineffective. Despite their advantages, these BCIs target relatively large areas of the brain, causing side effects. In this context, projects such as Neuralink aim to stimulate and inhibit neural activity with single-neuron resolution, expand their usage to other sectors, and thus democratize access to neurotechnology. However, these initiatives present vulnerabilities in their designs that cyberattackers can exploit to cause brain damage. Specifically, the literature has documented the applicability of neural cyberattacks, threats capable of stimulating or inhibiting individual neurons to alter spontaneous neural activity. However, these works were limited by a lack of realistic neuronal topologies to test the cyberattacks. Surpassed this limitation, this work considers a realistic neuronal representation of the primary visual cortex of mice to evaluate the impact of neural cyberattacks more realistically. For that, this publication evaluates two existing cyberattacks, Neuronal Flooding and Neuronal Jamming, assessing the impact that different voltages on a particular set of neurons and the number of neurons simultaneously under attack have on the amount of neural activity produced. As a result, both cyberattacks increased the number of neural activations, propagating their impact for approximately 600 ms, where the activity converged into spontaneous behavior. These results align with current evidence about the brain, highlighting that neurons will tend to their baseline behavior after the attack.

Keywords Brain-computer interfaces · Cybersecurity · Safety · Neuroscience · Neural cyberattacks

1 Introduction

Brain-computer interfaces (BCIs) are bidirectional systems able to interact with the brain, allowing the acquisition of neurological data and the stimulation of neurons. BCIs can be classified according to their level of invasiveness in invasive and non-invasive, depending if they require surgery to implant electrodes in the brain or use electrodes over the scalp respectively. The origin of these interfaces dates from the 1970 decade and from then, their use has mainly focused on medical scenarios, being employed for medical diagnosis and neurostimulation. The first one is useful for identifying a large variety of neurological conditions, such as epilepsy [1] or sleep disorders [2], being neuroimaging techniques essential to identify brain anomalies like tumors. Regarding neurostimulation, they are used to treat diseases for which a pharmacology-based

✉ Sergio López Bernal
slopez@um.es

Victoria Magdalena López Madejska
victoriamaagdalenalopezm@um.es

Gregorio Martínez Pérez
gregorio@um.es

Alberto Huertas Celdrán
huertas@ifi.uzh.ch

¹ Department of Information and Communications Engineering, University of Murcia, Campus de Espinardo, 30100 Murcia, Spain

² Communication Systems Group CSG, Department of Informatics IfI, University of Zurich UZH, 8050 Zürich, Switzerland

approach is not effective, highlighting disorders such as Parkinson's disease or obsessive-compulsive disorder [3].

BCIs have gained popularity in the last few years, increasing their usage in other economic sectors such as entertainment. This situation is determined by reducing the cost and size of non-invasive technologies, being more accessible to the general public. Invasive BCIs are also undergoing a technological revolution, reducing the size of the implanted electrodes and, therefore, the risk to patient's safety during the process. In this context, companies such as Neuralink are researching the creation of systems able to read and stimulate the brain with a resolution of individual neurons, aiming to democratize neurotechnology and bring it closer to final users [4].

Despite the advantages introduced by BCIs, they present different security problems. The literature has documented the possibility of attacks against these interfaces, affecting the integrity and availability of data and services, and the access to sensitive data, damaging information confidentiality [5]. Nevertheless, the most alarming risk lies in users' safety since attackers could take control of neurostimulation devices to overstimulate the neurons or prevent their stimulation, performing a denial of service attack over the device. In this context, the literature has defined the concept of neural cyberattacks, which can take advantage of vulnerabilities of next-generation neurostimulation systems to alter spontaneous neural activity, either by performing stimulation or inhibition actions.

However, these works present the limitation of evaluating the impact of these cyberattacks in non-realistic neuronal topologies based on a first approximation to the complexity of the brain using a CNN to translate the trained model into a neuronal topology [6–8]. Although the results and conclusions from these works were a first attempt to measure the impact of neural cyberattacks, they lack the necessary realism to fully understand the impact on real neural tissue in terms of neuronal activation.

Intending to improving the previous challenges, the present publication explores the implementation of neural cyberattacks in realistic neuronal models, allowed by recent advances in neuroscience and the availability of realistic neuronal topologies. Based on that, this work presents the following main contributions:

- The implementation of two existing neural cyberattacks, Neuronal Flooding (FLO) and Neuronal Jamming (JAM), in a realistic neuronal simulation. The scenario consists in a simulation of a realistic reconstruction of the primary visual cortex of mice, simplified to 450 neurons, based on a point-neuron network from Arkhipov et al. [9]. This simple topology serves as a first approximation to measure the impact of neural cyberattacks on realistic neurons, both in terms of their

connections and inner dynamics, accurately simulating the behavior of the mammalian visual cortex. Thus, the present work advances the literature by testing neural cyberattacks on a simplified but realistic neuronal topology for the first time. It represents a new step to evaluate them in human full-scale brain reconstructions in the future.

- The study of the impact caused by these cyberattacks on individual neuronal activity, using the number of spikes as an impact metric, which indicates the number of activations (action potentials or spikes) per neuron. For this purpose, this work defines two different experiments for FLO. The former examined the impact of different voltages over a specific number of neurons, while the latter evaluated the impact of different numbers of neurons attacked simultaneously. The experimentation of JAM only considered the impact of the number of neurons simultaneously under attack due to its inner behavior.
- The evaluation of the results obtained from different cyberattack configurations. This publication shows that FLO and JAM cyberattacks can increase the number of spikes of the affected neurons, maintaining their impact for 500–600 ms after the end of the attack, then converging with spontaneous neural activity. This work also compares between previous work focused on experimental neuronal topologies and the findings provided by this work on realistic models.

This article is organized as follows. Section 2 documents the state of the art of cybersecurity in BCI and the literature focused on realistic neuronal models. Section 3 highlights concerns related to neurostimulation systems, specifically those with single-neuron resolution, and the risks of neural cyberattacks. After that, Sect. 4 shows the design and implementation of neural cyberattacks on a realistic neuronal topology, presenting the results obtained and a comparison with the existing literature. Finally, Sect. 5 describes conclusions and future work.

2 Related work

This section first presents the state of the art of cybersecurity in the field of BCI. After that, it details the most relevant and recent models focused on reconstructing neuronal networks in a specific brain region.

2.1 Cybersecurity of BCIs

The literature focused on the cybersecurity of BCIs has gained popularity in recent years, where López Bernal et al. [10] analyzed the cybersecurity aspects of each phase of

the BCI life cycle and the most common architectural designs. The authors documented that the field of BCI offers tremendous opportunities for cybersecurity research, highlighting a reduced number of works in this domain. In a more specific way, and focusing on confidentiality, the most widely studied dimension, Martinovic et al. [11] demonstrated the impact that the presentation of malicious visual stimuli could have on users, being capable of obtaining sensitive information from brain waves obtained as a response to the stimulus, such as bank data, residence area, emotions, religious beliefs, or sexual orientation. Frank et al. [12] and Quiles Pérez et al. [13] followed a similar approach, studying the duration of the stimuli presented, aiming to ensure that the subjects were not aware of the attack, thus studying the subliminal threshold and demonstrating the effectiveness of stimuli almost invisible to the human eye.

Takabi et al. [18] studied the most common BCI applications, identifying that most of this software can access brain data without restrictions, affecting the confidentiality of its users' sensitive data. Finally, Bonaci et al. [14] presented the concept of BCI Anonymizer as a layer added to these devices, able to anonymize the signals transmitted from the BCI to external systems, guaranteeing the confidentiality of information against BCI applications offering unlimited access to neural data. Focusing on data integrity, Li et al. [5] documented the possibility of gathering the brain waves for later impersonating a subject's legitimate signals. However, cyberattacks could also affect brain data, as Martínez Beltrán et al. [22] documented, which demonstrated the possibility of introducing noise into brain signals to confuse the classifiers while identifying relevant aspects of these signals. Furthermore, Ienca et al. [19] and Li et al. [5] highlighted that an attacker could aim to disrupt the acquisition process by using different attack vectors, affecting service availability. Finally, Landau et al. [21] indicated that the alteration of the results of diagnostic tests based on BCI could impact patients' safety, inducing the application of incorrect or unnecessary treatments.

Concerning neurostimulation, the literature has identified the risk of maliciously using these technologies, where an attacker could damage users' brains irreparably, highlighting that these attacks do not need to be too complex to generate a considerable impact [16]. In the field of Implantable Medical Devices (IMD), works such as Pycroft et al. [17] indicate that overstimulation actions could cause damage to cerebral tissue, rebound effect, or denial of stimulation, altering the legitimate neurostimulation treatment. Marin et al. [20] highlighted that attacks over the neurostimulation process could affect speech or movement, causing brain damage that could even impact patients' life. Moreover, Cámara et al. [15] documented

attacks focused on draining the battery of stimulation systems, affecting the availability of the system. They also proposed countermeasures to avoid attacks, such as using external authorization systems or anomaly detection techniques.

López Bernal et al. [7] identified vulnerabilities in the design of new-generation neurostimulation systems (e.g., Neuralink [4]), which intend to offer neurostimulation and inhibition actions with single-neuron resolution. Moreover, this work introduced the concept of neural cyberattacks as attacks that exploit the previously indicated vulnerabilities to disrupt neural activity. Finally, this work implemented two neural cyberattacks, Neuronal Flooding (FLO) and Neuronal Scanning (SCA), focused on overstimulating the neurons the attack targets differently. After that, these authors presented in [6] Neuronal Jamming (JAM) as a novel cyberattack capable of inducing neuronal inhibition maliciously. Finally, López Bernal et al. [8] presented a taxonomy of eight neural cyberattacks (including the previous three) with different behaviors. They studied the impact of these cyberattacks on neural activity using a neuronal simulation based on the model obtained from training a CNN due to limitations in the existing neuronal models at that moment.

Table 1 summarizes the works previously highlighted, comparing their focus (neural data acquisition or neurostimulation), the dimension that existing threats can affect (confidentiality, integrity, availability, or safety), and a brief description of these threats.

2.2 Realistic neuronal models

Although there is a large number of models representing the reconstruction of neural networks, most of them lack the required realism. This section presents a summary of the most relevant neuronal models existing in the literature.

The model proposed by Chadderdon et al. [23] was developed as a computational proposal based on studies focused on mapping the cerebral activity of the primary motor cortex (M1) in mice, including excitatory and inhibitory neurons. After conducting two types of analysis, one static and one dynamic, the results showed that static brain maps could be related to the mapping of cerebral activity. Ferguson et al. [24] provided a better understanding of the mammalian brain by creating a model focused on excitatory neurons in the CA1 region of the hippocampus with theta frequency bursts (electromagnetic oscillations in the human brain associated with the early stages of sleep) between 3–12 Hz. They concluded that cellular adaptation in pyramidal cells could be an important aspect of the hippocampus as a starting point for including inhibitory cells in future models.

Table 1 Summary of the state of the art of cybersecurity in BCI

Reference	Approach	Impact	Threat
Martinovic et al. [11]	Acquisition	Confidentiality	Malicious visual stimuli
Bonaci et al. [14]	Acquisition	Confidentiality	Unrestricted applications
Camara et al. [15]	Stimulation	Safety	Brain damage
Li et al. [5]	Acquisition	Integrity availability	Forge legitimate signals Stop the BCI
Ienca et al. [16]	Stimulation	Safety	Brain damage
Pycroft et al. [17]	Stimulation	Safety	Brain damage
Takabi et al. [18]	Acquisition	Confidentiality	Unrestricted applications
Frank et al. [12]	Acquisition	Confidentiality	Malicious visual stimuli
Ienca et al. [19]	Acquisition	Availability	Stop the BCI
Marin et al. [20]	Stimulation	Safety	Brain damage
Landau et al. [21]	Acquisition	Safety	Disrupt diagnostic tests
López Bernal et al. [7]	Stimulation	Safety	Neuronal overstimulation
Quiles Pérez et al. [13]	Acquisition	Confidentiality	Malicious visual stimuli
López Bernal et al. [6]	Stimulation	Safety	Neuronal inhibition
Martínez Beltrán et al. [22]	Acquisition	Integrity availability	Disrupt brain waves confuse classifiers
López Bernal et al. [8]	Stimulation	Safety	Taxonomy of neural cyberattacks

After this work, the study of Markram et al. [25] documented a first approximation of the digital reconstruction of a neocortical microcircuit based on the somatosensory cortex of rats, using both excitatory and inhibitory neurons. The authors used cellular and synaptic organizing principles to reconstruct the anatomy and physiology of the neurons algorithmically. The simulation recreated a series of results that allowed further investigation of the underlying cellular and synaptic mechanisms, also providing experiments that were not possible either *in vitro* or *in vivo*. Bezaire et al. [26] recreated a full-size CA1 area of the hippocampus from rodents based on the integration of supercomputing data and data obtained after studying interneurons during theta oscillations. This work revealed new knowledge about the organization of the CA1 area during this kind of oscillation.

Arkhipov et al. [9] proposed a realistic reconstruction of the fourth layer (L4) of the primary visual cortex (V1) from mice, exploring the neuronal behavior of this cortical layer on 45,000 neurons, containing excitatory and inhibitory cells. In particular, this publication studied five different

morpho-electrical neuronal models, providing realism to the simulation by including thalamocortical stimuli, representing different categories of realistic visual stimuli. Moreover, these authors provide a lightweight topology of 450 neurons to allow simplified experimentation and analysis of their behavior. These simulations can be performed using either NEURON or NEST simulations, where the first allows obtaining a biophysically realistic behavior. At the same time, the latter offers an approximation to their real behavior but maintains low computational resources.

After that, Bittner et al. [27] analyzed the activity of the neurons according to whether they are excitatory or inhibitory, applying factorial analysis to their spontaneous activity. The authors demonstrated the importance of studying the type of neuron, allowing more robust statistics. Crone et al. [29] evaluated the computational performance of GENESIS and PGENESIS neural simulators for large-scale simulations. The authors focused on high-fidelity neuronal models representing 50–74 compartments per neuron instead of LIF neurons. They subsequently analyzed the simulation performance and scalability using

a modified version of PGENESIS and a thalamocortical network model.

Billeh et al. [30] created a biological model of V1 from mice, offering a reconstruction of a complete microcircuit, representing both excitatory and inhibitory neurons in a close way to the behavior of the neurons in the brain. In particular, this reconstruction offers a topology of around 230,000 neurons, containing models with differentiated behaviors according to the morpho-electrical characteristics of each neuronal group and the cortex layer in which they have been detected. This work offers two levels of granularity, using the first alternative biophysically detailed neurons. In contrast, in the second, they employed Gated Leaky Integrate-and-Fire (GLIF) neurons following a point-neuron approach. Regarding neural simulations, the biophysical model uses NEST for low-level simulation, while the second alternative uses NEST.

Table 2 summarizes the models used by the previous publications, comparing the simulators used, the neuronal model in which they focus, and the region of the brain they study. Based on them, the present publication selected the model of Arkhipov et al. for testing neural cyberattacks, presenting more details in Sect. 4.1.

3 Concerns of invasive BCIs and neural cyberattacks

This section presents the problems existing in the designs of prospecting neurostimulation systems, highlighting the feasibility of performing neural cyberattacks to affect users' safety. Moreover, this section describes the limitations existing in the literature regarding these cyberattacks, being an opportunity to advance the state of the art.

Invasive neurostimulation techniques are widely used in the medical sector. One of these technologies is Deep Brain

Stimulation (DBS), which consists in placing electrodes on targeted areas affected by a neurological condition, such as Parkinson's disease, reducing the effects caused by these neurological alterations [31]. Responsive Neurostimulation (RNS) is also an essential technology used nowadays, which consists in a closed-loop neurostimulation system able to record neural activity to predict abnormal behavior. Once it is detected, the implanted device automatically stimulates the brain to stop de seizure. This technology is widely used in epilepsy when a drug-based treatment is ineffective [32]. Despite their advantages, these technologies cover relatively large areas of the brain. To improve the spatial resolution of these interfaces, new-generation neurostimulation BCIs aim to miniaturize the size of the electrodes used to the nanoscale, being able to cover individual neurons, thus reducing the side effects of affecting wide cerebral areas. Examples of prototypes of these novel BCIs are Neuralink [4], Synchron [33], and wireless optogenetic nanonetworks based on neural dust [34].

However, some of these recent alternatives present vulnerabilities in their designs that cyberattackers could exploit [7]. In particular, the literature has detected issues in the architecture of Neuralink and optogenetic nanonetworks, highlighting vulnerabilities in using smartphones to control neurostimulation devices, which are relatively easy to attack, being malware a common example. Moreover, the communication between the smartphone and the BCI device is not robust. In Neuralink, the communication is based on Bluetooth, which has many known vulnerabilities, while optogenetic nanonetworks do not implement protection mechanisms. Based on that, an attacker could take control of the BCI to perform malicious actions, such as stimulating or inhibiting neural activity.

As briefly documented in Sect. 2, the literature has addressed the concept of neural cyberattacks as

Table 2 Neuronal models studied from the literature

Reference	Simulator	Neuronal model	Brain region
Chadderdon et al. [23]	NEURON	Personalized	Motor cortex M1 (excitatory and inhibitory)
Ferguson et al. [24]	Brian	Abstract Izhikevich	Hippocampus CA1 (excitatory)
Markram et al. [25]	N/A	Personalized	Somatosensory cortex
Bezair et al. [26]	NEURON	Personalized	Hippocampus CA1
Bittner et al. [27]	Julia	Abstract leaky integrate-and-fire	Not specified
Schmidt et al. [28]	NEST	Abstract leaky integrate-and-fire	Visual cortex
Arkhipov et al. [9]	NEURON, NEST	Personalized	L4 of the visual cortex V1 (excitatory and inhibitory)
Crone et al. [29]	PGENESIS, GENESIS	Personalized	Neocortex (excitatory and inhibitory)
Billeh et al. [30]	NEURON, NEST	Personalized	Microcircuit from visual cortex V1 (excitatory and inhibitory)

mechanisms to alter neural activity. As an introduction, Fig. 1 represents how an attacker could execute neural cyberattacks by exploiting vulnerabilities existing in new-generation neural implants. That way, these cyberattacks impact the brain by applying stimulation or inhibition, causing an alteration of neural activity.

Although the works published to date are promising, they have certain limitations that must be considered. When conducting the investigation described in previous articles, there was a lack of realistic neuronal topologies to experiment with. In this way, these works opted to build a topology as realistic as possible while keeping in mind certain assumptions and approximations of reality. In particular, these works trained a CNN to solve the particular problem of a mouse that must exit a determined maze, based on similarities between the architecture and functioning of CNNs with the visual cortex in mammals. The resulting model from the training process was translated to a neuronal simulation in the Brian2 simulator, using the Izhikevich neuronal model and keeping the same distribution of nodes and weights between nodes.

Moreover, these works only considered the behavior of excitatory neurons, not including the dynamics of neurons capable of inhibiting the ability of other neurons to spike, a situation that occurs in the brain naturally. Finally, the visual stimuli considered as input to study the neuronal activity were a simplification of the vision process. Based on the above, there is a lack in the literature for works implementing and evaluating neural cyberattacks on realistic neuronal conditions.

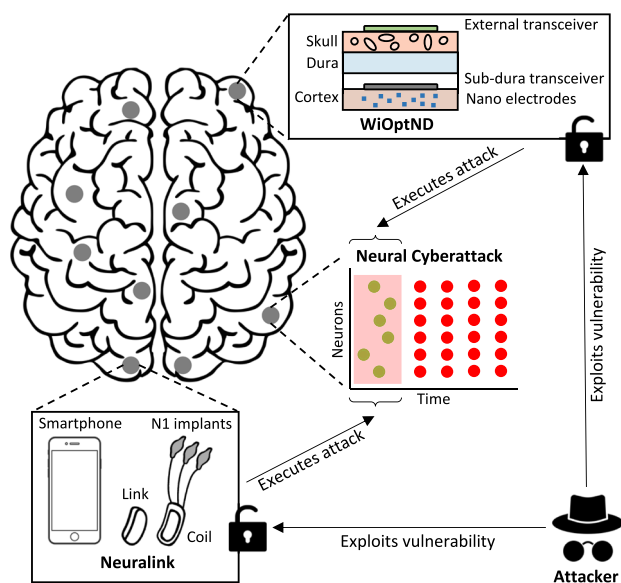


Fig. 1 Attacker executing the proposed neural cyberattacks able to exploit vulnerabilities in BCI systems of neural neuromodulation and generate particular impacts in BCIs

The presence of vulnerabilities is not exclusive to the neurostimulation field, being common in IMD sectors. Despite these limitations, companies manufacturing IMDs focus most efforts on developing the functionality of the devices, not taking into consideration the criticality of implanting robust security mechanisms [35]. In conclusion, many studies have identified vulnerabilities in a wide range of IMDs that cyberattackers could exploit to affect users' safety [36, 37]. Nevertheless, several standards and guidelines aim to help improve IMD security [35]. First, the NIST Cybersecurity Framework contains recommendations to mitigate the risks in any organization, providing a comprehensive set of actions to respond to attacks [38]. Furthermore, the FDA published a series of premarket and postmarket recommendations to manage the security of medical devices [39–41]. Finally, several standards are also applicable to the security of IMDs. UNE-EN ISO 14971:2020 documents how to manage the risks associated with medical devices [42]. At the same time, UL 2900 is a cybersecurity standard for products connected to the network, also accepted by the FDL for their use in medical devices and software [43].

4 Neural cyberattacks over realistic topologies

This section presents the initial work toward applying neural cyberattacks over realistic neuronal topologies to analyze the impact these threats can have on the brain. First, it describes the scenario used to validate neural cyberattacks, later describing relevant aspects of the design and the implementation. Finally, this section presents the results obtained and a discussion that compares these results with those existing in the state of the art.

4.1 Scenario description

Among the taxonomy of eight neural cyberattacks existing in the literature [8], this section presents two of them, Neuronal Flooding (FLO) and Neuronal Jamming (JAM), as representative examples of opposite behaviors. First, FLO consists in overstimulating multiple neurons in a particular time instant, where attackers trigger the activation of individual neurons controlled by the BCI at will, causing an action potential that is then propagated to adjacent neurons. This cyberattack does not require previous knowledge about the status of the target neurons, having a lower complexity compared to other neural cyberattacks from the literature. This cyberattack is influenced by flooding cyberattacks from computer networks, where an attacker intends to collapse a network by

transmitting a high number of data packets, having typically specific systems within the network as targets [7].

In contrast, cyberattacks based on jamming can introduce interference to the physical medium, disrupting legitimate communication between devices. In a neurological context, the literature presents JAM as a neural cyberattack able to inhibit the activity of a set of neurons during a determined temporal window. In this context, cyberattackers can completely suppress the activity of the affected neurons, preventing the brain from having its normal behavior. This attack also presents a low execution complexity, only requiring configuring the duration of the attack and the list of neurons to be inhibited [6].

To better understand the behavior of FLO and JAM cyberattacks, Fig. 2 presents an example of the temporal evolution of neural activity in a neuronal topology synthetically created with three layers, having 200 neurons in the first, 72 in the second, and four in the third. In particular, it represents a simplified simulation of 215 ms to facilitate its visualization, presenting the spontaneous behavior and the dynamics of FLO and JAM when executed at instant 10 ms. Focusing on FLO, Fig. 2 depicts that this neural cyberattack can anticipate the occurrence of spikes in the affected neurons (see instant 10 ms) compared to spontaneous behavior. That way, after applying FLO, those neurons affected by the attack repeatedly anticipate their spikes, causing an unbalance in neural activity periodicity. In contrast, the JAM cyberattack applied between instants 10 and 60 ms causes a complete inhibition of the targeted neurons, which resume their behavior after the finalization of the attack. As in FLO, JAM causes a variation in neural activity synchronization, as the affected neurons resume their activity as soon as the attack ends. It is relevant to note that this figure does not intend to present a detailed study of the impact of neural cyberattacks but to illustrate their mechanisms of action in a simplified way.

To implement FLO and JAM neural cyberattacks, this work used the neuronal topology published by Arkhipov

et al. [9]. These authors modeled the behavior of a portion of L4 of V1 using input stimuli presented to the eye in a simulated way. These inputs (static images, videos with multiple images, and particular visual effects) arrive at the Lateral Geniculate Nucleus (LGN), which sends information to V1, making the simulation more realistic. In the same direction, Fig. 3a represents the reconstruction performed by Arkhipov et al. in a visual way, highlighting the five biophysiological neuronal models employed, containing around 45,000 neurons from the L4 of the V1 region.

As shown in Fig. 3b, this representation offers two different resolution levels. First, the biophysiological approach characterizes neurons at the molecular level, reconstructed from a V1 portion of 400 μm . On the contrary, the Leaky Integrate-and-Fire (LIF) reconstruction, obtained from an area of 845 μm , simplifies neuronal complexity while keeping neuronal characteristic behavior. Based on them, Arkhipov et al. [9] offer two simulation approaches. The first one can recreate the biophysiological behavior of neurons in detail at the molecular level using the NEURON simulator. In contrast, the second option offers a point-neuron approach that is a computationally lighter alternative, although sufficient for most purposes, using NEST as simulation software. In both cases, the creation of the network and the management of the simulation at a high level is facilitated by the Brain Modeling Toolkit (BMTK) simulator developed by Allen Institute [44]. Finally, Table 3 highlights the number of cells per neuronal population, and the number of morpho-electrical models identified.

Despite the richness of this reconstruction, its simulation is computationally expensive, even when using point-neuron models. Based on that, Arkhipov et al. [31] also offer a simplified neuronal topology for testing, having 450 LIF neurons, that model the change of neuronal membrane voltage over time as differential equations, containing excitatory and inhibitory cells. This reduced topology is used in the present publication as a first approach for implementing neural cyberattacks in realistic scenarios.

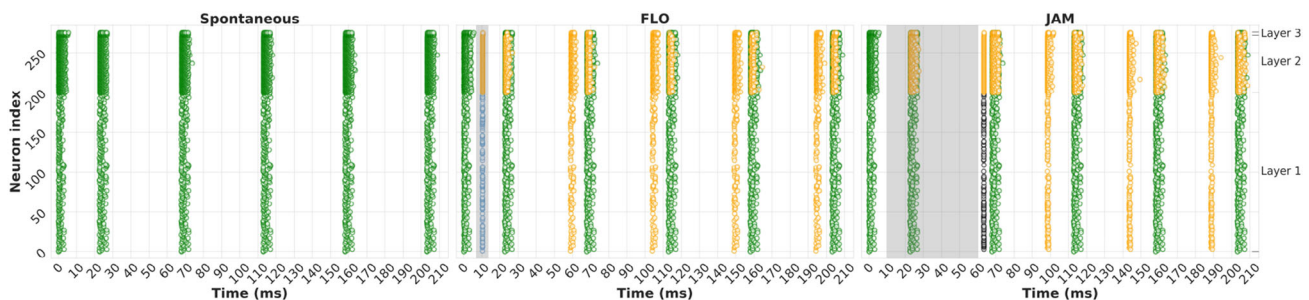
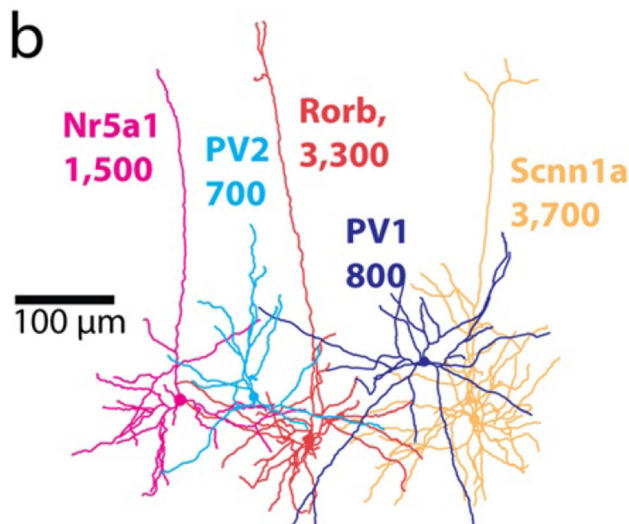
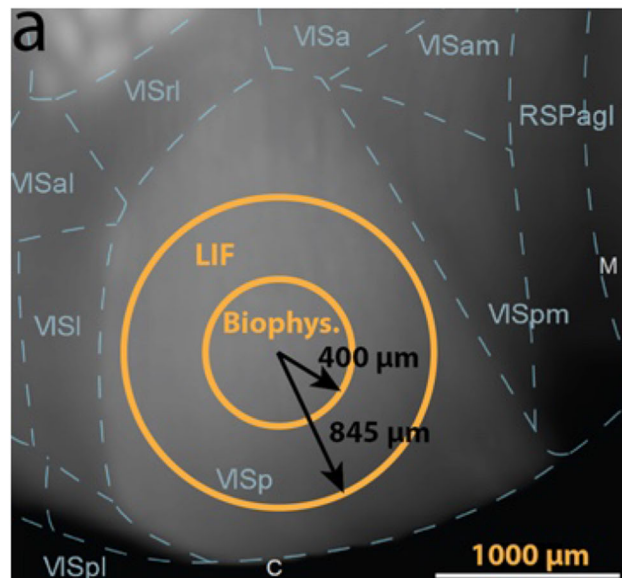


Fig. 2 Visual representation of the behavior of two neural cyberattacks: FLO and JAM. Green points represent neuronal spikes of spontaneous behavior, blue points indicate stimulated neurons, black

points inhibited neurons, and orange points highlight spikes produced due to the attack. The grey background indicates the attack duration



(a) Neuronal models employed by the biophysical reconstruction.



(b) Location of the neurons used in the neural reconstruction.

Fig. 3 Overview of the neuronal reconstruction provided by Arkhipov et al. [9]

4.2 Experimental design

Figure 4 depicts the solution proposed, highlighting the use of a neuronal topology from the primary visual cortex of mice to test FLO and JAM neural cyberattacks. In particular, this work uses the BMKT simulator to orchestrate the multiple simulations performed, including the parameters used for each attack. Then, BMKT relies on NEST to perform low-level simulations of the point-neuron topology, exporting the resulting data (spike trains) to CSV.

Table 3 Description of the neuronal populations reconstructed by Arkhipov et al. [9] indicating, for each neuronal population, the number of neurons and their behavior

Reconstruction	Population	Behavior	Cells
Biophysical	Scnn1a	Excitatory	3700
Biophysical	Rorb	Excitatory	3300
Biophysical	Nr5a1	Excitatory	1500
Biophysical	PV1	Inhibitory	800
Biophysical	PV2	Inhibitory	700
LIF	Excitatory LIF	Excitatory	29,750
LIF	Inhibitory LIF	Inhibitory	5250

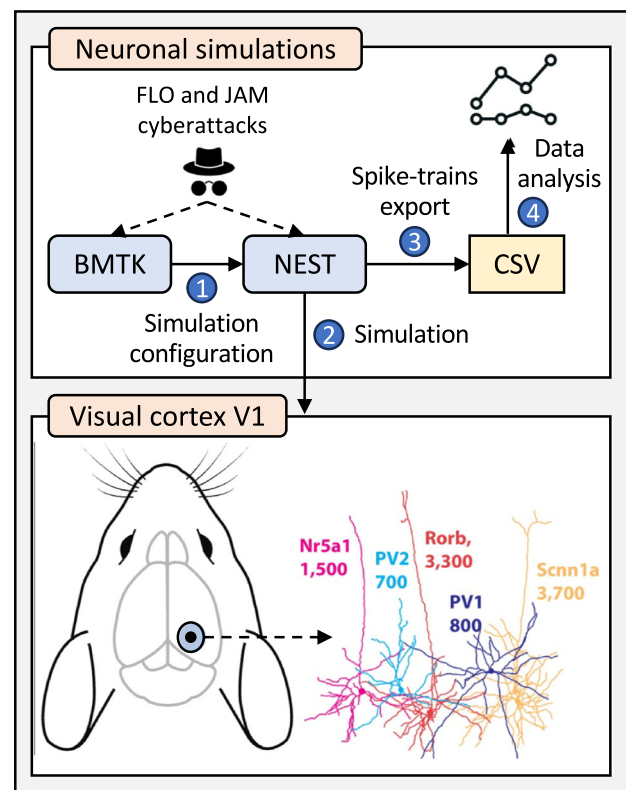


Fig. 4 Graphical representation of the solution proposed, indicating the main steps to simulate neural cyberattacks and the analysis of the results obtained

Finally, a data analysis process measures the impact caused by cyberattacks in a quantitative way.

In terms of neural cyberattacks, this work evaluates different experiments and configurations for each neural cyberattack to measure their impact, supported by the decisions done in existing literature [6, 7]. Focusing first on FLO, this work performed two experiments. The first experiment studied the impact caused by different voltages when attacking a particular number of neurons, in this case, the complete topology of 450 neurons. The second

experiment evaluated the impact of attacking different numbers of neurons simultaneously, using as fixed voltage the threshold voltage of the membrane, which guarantees that the attacked neurons will be activated, thus generating a spike. In contrast, the experimentation for JAM only studied the impact of simultaneously attacking different numbers of neurons. This decision was motivated by forcing each affected neuron to be in its particular resting voltage (V_{reset}) while the attack was active, which completely inhibited its neuronal activity, thus excluding the possibility of testing different voltages.

Table 4 summarizes the parameters considered for each implemented neural cyberattack. In particular, it is relevant to note that FLO is executed in the instant 10 ms, while JAM is active during 500 ms, from 1000 to 1500 ms. Moreover, each combination of parameters per experiment was executed ten times to provide variability in the results by selecting random neurons in each execution and to better understand the impact of these threats. This variability allows to validate the results obtained by studying the stability of the attacks when changing the neurons targeted in each execution. These parameters are inspired by those already used in the literature [6, 7]. The simulation duration has been set to three seconds as it is sufficient to evaluate the impact and propagation of the attacks while simplifying the experimentation complexity. Moreover, the instants under attack and the number of executions have been considered based on the experimentation and results existing in the literature. The number of neurons under attack is homogeneously distributed to analyze the differences of this parameter on neural activation, from a small set of neurons to the maximum possible number. Finally, the voltages selected for FLO are homogeneously distributed in the natural range of neuronal voltages, reaching V_{th} , corresponding to the threshold voltage in which the neuron fires. In contrast, JAM sets the voltage to the minimum possible value to prevent neurons from having spikes (V_{reset}). It is essential to note that these parameters were selected following an incremental process in which the complexity of the attacks and the parameters used for each experiment were evaluated and tuned until reaching the values presented above.

4.3 Implementation of the solution

To implement the neural cyberattacks, it was necessary to perform adaptations to the source code of BMTK and NEST simulators. Focusing first on BMTK, Fig. 5 presents the structure of the model, where grey boxes correspond to directories and white boxes are files or scripts. In particular, the file *Run_pointnet.py* is highlighted in red as it contains all necessary changes to the simulator to be compatible with the execution of neural cyberattacks. The first modification consists in parameterizing the attacks to prevent recompiling the source code after each test. For that, this script accesses three new external files that can be read and modified by *Run_pointnet.py*. In particular, the file *Type_attack.txt* indicates whether the attacker wants to run a FLO or a JAM cyberattack. The files *FLO_attributes.txt* and *JAM_attributes.txt* contain the attack parameters for FLO and JAM, respectively. The configuration file for FLO contains the time instant when the simulator must execute the attack, as well as two lists, one indicating the voltages that must be evaluated and another with the number of neurons. In the case of JAM, the configuration file indicates the start and end of the attack, in addition to the list of numbers of neurons. All these parameters are stored in Python dictionaries for their access.

Using these dictionaries, BMTK can now test each combination of parameters. It requires communicating the configuration files to NEST, which is responsible for building the network and performing the simulation. This functionality has also been included in *Run_pointnet.py* in a specific function. Once an execution ends, NEST exports the results in CSV format. However, this file lacks essential information for its subsequent processing. For that, it has been adapted to create a more extensive CSV file, indicating the attack performed, the instant or temporal period of the attack, the voltage used to attack (only for FLO), the number of neurons under attack, the number of execution (from one to ten), and the spikes of all the neurons. This last field is represented in two different columns. The first one contains timestamps, representing a list of all simulation instants in which the simulator registered spikes. The second contains a list of neuron IDs indicating the spiked neurons at each specific timestamp. After testing all

Table 4 Parameters used for each implemented neural cyberattack

Attack	Simulation duration	Instant(s) under attack	Number of executions	Number of neurons under attack	Voltage (mV) used to attack
FLO	3 s	10 ms	10	[50, 100, 200, 300, 400, 450]	[5, 10, 20, 30, V_{th}]
JAM	3 s	1000–1500 ms	10	[50, 100, 200, 300, 400, 450]	V_{reset}

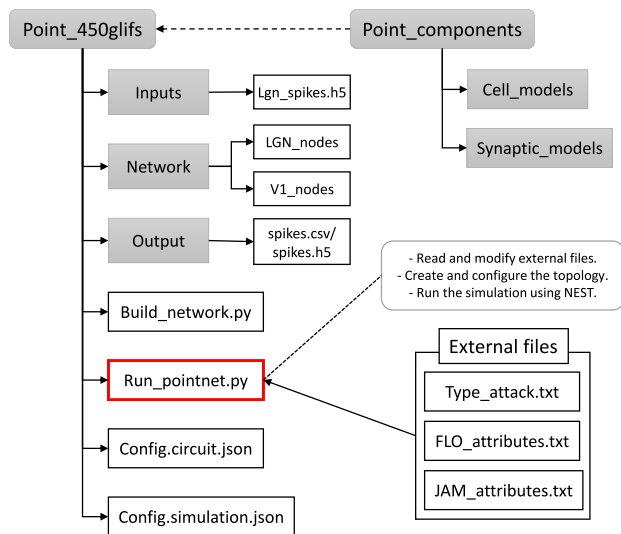


Fig. 5 Representation of the folders (grey) and files (white) defining the neuronal topology used, highlighting in red where the changes have been performed

parameters, the results are integrated into a common CSV file to ease the subsequent data analysis process.

Moving to the changes performed in NEST, Fig. 6 first presents the tree of function calls, starting from the invocation of NEST within BMTK until reaching the function modified in NEST to run neural cyberattacks (*Update()*). Specifically, white boxes represent the simulators used, blue blocks indicate software functions, and orange figures are software classes.

The invocation of the *Nest.Simulate(t)* method from BMTK triggers the execution of the *Run(t)* function in NEST, which subsequently calls the *Simulation_Manager* class. This class is key to implementing and executing neural cyberattacks, providing functionalities to control

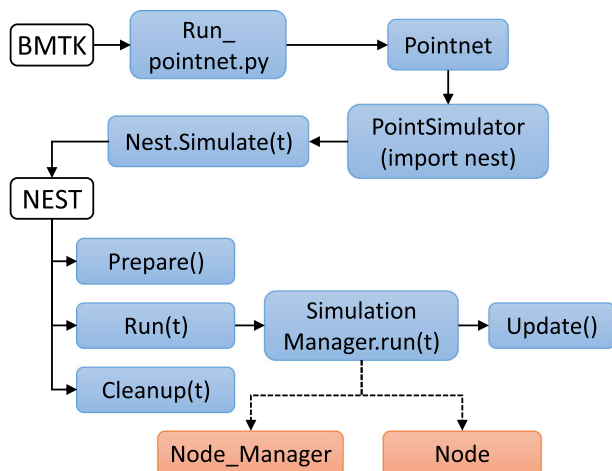


Fig. 6 Tree of function invocations from BMTK and NEST to run neural cyberattacks

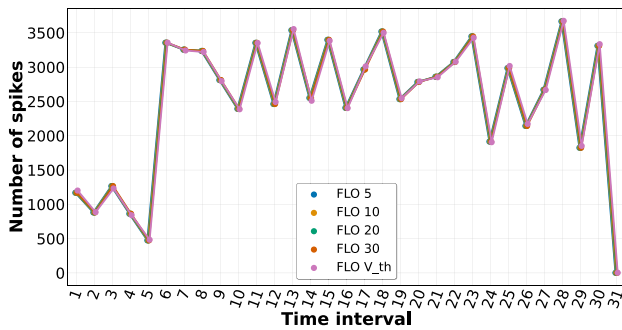
and manage the execution of simulations. The code of this class is complex, containing many neuronal models and relationships with other classes. In particular, this class relies on the classes *Node_Manager* and *Node* to obtain all the parameters of neurons of a specific type. Since most low-level functionality is implemented in this class, this is the desired place to include the functionality of neural cyberattacks.

To correctly implement FLO and JAM, creating a function capable of storing the parameters of the attacks included in external files into global dictionaries in the *Simulation_Manager* class was essential. After that, the authors defined a method for each cyberattack. Focusing first on FLO, the added function iterates over the nodes that will be attacked, obtaining their membrane potential. The implementation considers several possibilities. First, if the number of neurons is lower than the total number of neurons available, it will choose a random set from them. If not, all neurons are attacked. Then, the voltage resulting after the attack will be the sum of the current membrane potential and the voltage used to attack. However, if the result is greater or equal to the specific threshold of a neuron, then the membrane potential is set to the threshold value. The implementation of JAM follows the same approach for the random selection of neurons. Once having a set of neurons to attack, their voltages are set to the resting voltage of the neuron (V_{reset}).

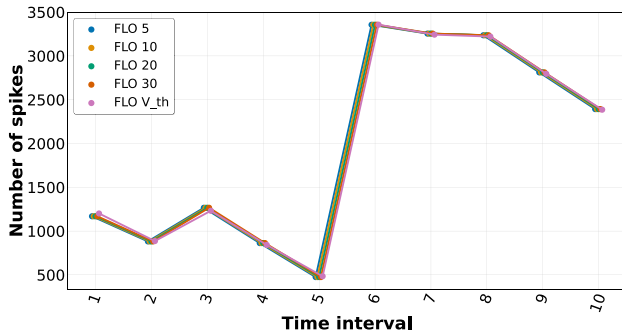
Once the attacking functions are implemented, they must be included where the simulation status is updated. This process is performed in different threads within the *Update()* function, where the most relevant is the *master* thread, which updates the current time of the simulation. This location is selected for running the attacking methods. In particular, the modifications performed first verify if the time instant of the simulation is included within the parameters of the current attack. If so, the simulation runs the attacking function.

4.4 Results analysis

Regarding the results, Figs. 7 and 8 show, on the one hand, the impact of applying the FLO neural cyberattack on the realistic topology at 10 ms using different attack configurations. As the resolution of the simulator is 0.1 ms, the figure would have 30,000 timestamps on the X-axis. The axis data have been aggregated into 30 intervals of 100 ms to improve its visualization. Based on the results obtained in Fig. 7 concerning the first experiment, all the voltages except the threshold generate a similar number of spikes. This situation is because, as mentioned above, the threshold value forces neurons to generate activity, whereas, with the rest, not all the targeted neurons will generate activity. On the other hand, Fig. 8 shows the results of the second



(a) Complete simulation.



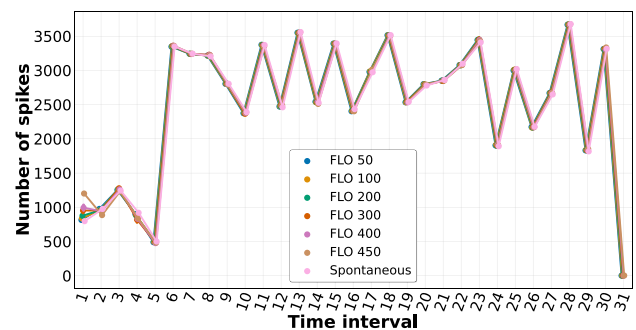
(b) Representation of the first ten intervals.

Fig. 7 Impact of FLO cyberattacks evaluating different voltages on all neurons of the topology (experiment 1), executed at instant 10 ms (interval 1)

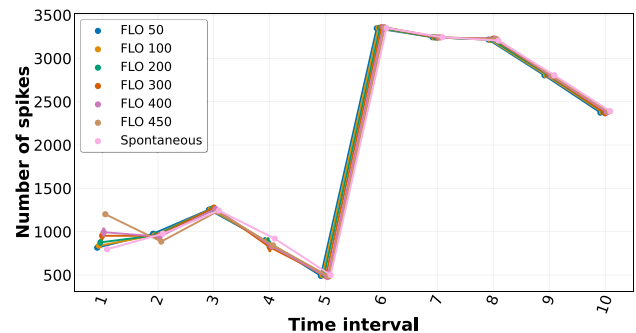
experiment. It can be observed that the greater the number of neurons attacked, the greater will be the number of spikes. Furthermore, by conducting the cyberattack on all neurons, the impact is significantly higher than in the other configurations.

In both experiments, neurons suffer a variation in their temporal evolution during the cyberattack. However, over time, the activity converges to its spontaneous state. This trend is explained by the inner behavior of realistic neuronal networks, which tend to their “normal” state. Specifically, this dynamic is explained by the tendency of neurons in the brain to return to their resting state after transmitting an action potential. Therefore, the figures of these experiments (Fig. 7 and Fig. 8) depict that neurons require about five intervals (500 ms) to get stabilized towards the spontaneous behavior after applying the attack.

The impact caused by JAM cyberattacks is presented in Fig. 9, representing attacks conducted for 500 ms (intervals ten to 15), inhibiting all targeted neurons simultaneously. Thus, during this interval, all neurons attacked are prevented from generating any activity. However, even if the number of spikes gets reduced, the neural activity presents a similar trend over time, where the intervals with higher and lower spikes are maintained. This figure also indicates that the greater the number of neurons affected,



(a) Complete simulation.



(b) Representation of the first ten intervals.

Fig. 8 Impact of FLO cyberattacks evaluating different numbers of simultaneously attacked neurons fixing the voltage to the maximum threshold (experiment 2), executed at instant 10 ms (interval 1)

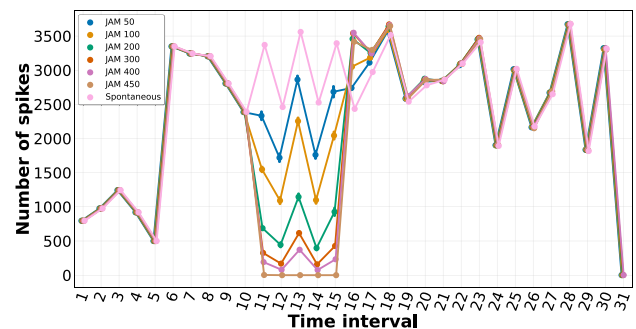


Fig. 9 Impact caused by different numbers of simultaneously attacked neurons in JAM, executed during the temporal window between 1000 and 1500 ms (intervals 10 to 15)

the greater the impact on the number of spikes. When the action of the attack stops, since all neurons are at their lower limit within their natural voltage range, all neurons get momentarily synchronized to generate an elevated spikes peak, as seen in interval 16. As a result, from the end of the attack until past 600 ms (intervals 15 to 21), the activity significantly differs from the spontaneous state, getting stabilized after that instant, although with small variations maintained until the end of the simulation.

As can be seen, both cyberattacks can affect neuronal activity in a realistic environment. Regarding FLO, this

attack does not generate large variations. Focusing on JAM, there is a considerable decrease during the attack. However, in both cases, the neuronal activity gets synchronized, stabilizing after around 500 ms or 600 ms, respectively.

4.5 Discussion

This section compares the results from the literature with those obtained in this publication. However, it is essential first to conduct a comparison between scenarios. In the case of the works documented in Sect. 3, the topology used contained 276 neurons based on a model trained from a CNN to solve the problem of a mouse that must exit a certain maze. Thus, there were large dependencies between neurons that influenced the results. In contrast, the topology used in this paper represents a simplified, realistic recreation of the primary visual cortex composed of 450 neurons. This way, the results obtained present a higher realism when considering the activation and inhibition of neurons induced by neural cyberattacks. Moreover, the inputs in each case are different. In the related work, the neuronal model employed received the current status of the maze, thus introducing restrictions on realism. However, the topology from Arkhipov et al. [9] can use realistic visual stimuli, although the present publication has simplified these inputs due to the computational overload they introduce. Particularly, the implemented models have as input realistic static values represented as LGN spikes provided by the BMTK simulator. Furthermore, the simulators in both scenarios are different, where the first use Brian2, and the second employs NEST.

Regarding the comparison of results, only the number of spikes will be considered since it is the only metric considered in the study of the realistic topology. In the first experiment of FLO, the differences between voltages are minimal in both scenarios. Besides, in both the experimental and realistic scenarios, FLO produces a greater impact on the number of spikes when a higher voltage is used. However, in the literature, the number of spikes decreases, while in the current publication, it increases depending on the magnitude of the voltage used.

In the second experiment of FLO, specifically in the state of the art, attacking more neurons resulted in a greater decrease in neuronal activity. In contrast, in the realistic scenario, an increase in the number of affected neurons implies an increase in the activity of these neurons. Additionally, this work analyzed the ability of neurons to converge toward spontaneous behavior. While this publication observed a notable impact during 500 ms since the neural activity gets synchronized with the spontaneous activity after this period, the literature shows a propagation of the impact over time. According to neuroscience

literature, the brain can stabilize neuronal activity when it identifies an abnormal activity, using compensatory mechanisms such as inhibiting neurons or using brain plasticity. Based on that, the results shown in Sect. 4.4 are coherent and aligned with the current evidence on neuroscience [45].

The experimentation of JAM concluded that there is a greater impact when increasing the number of neurons affected in both scenarios. Besides, both scenarios presented a reduction in the number of spikes. However, the literature did not study the propagation of JAM cyberattacks over time. In contrast, in the realistic scenario, the neuronal activity stabilizes about 600 ms after the attack. Finally, Table 5 summarizes the comparison between the literature and this publication, highlighting the differences in key aspects of the experimental scenario and the results obtained.

It is essential to note that although this paper introduces a next step in realism compared to the existing literature, it still focuses on analyzing the impact of neural cyberattacks on neural activity in a quantitative way. This work quantifies the variability in neural activity caused by these threats without attending to the real-world impact they could cause on neural functions, such as vision. Thus, these limitations highlight the need for further research using large and complex topologies, considering their relationship with the external medium, such as visual stimuli.

5 Conclusion

This work presents the implementation of two neural cyberattacks, FLO and JAM, using a realistic neuronal topology of the primary visual cortex, quantifying the impact these attacks could have on vision. Focusing on the results for FLO, the first experiment analyzed the effect of different voltages in a specific set of neurons, attacking all

Table 5 Comparison between the literature and this publication

Category	Literature	This work
Number of neurons	276	450
Neuronal model	Extracted from a CNN	Realistic from V1
Model input	Status of the maze	Inputs from LGN
Simulator	Brian2	BMTK, NEST
FLO—voltage	↑ voltage, ↑ impact	↑ voltage, ↑ impact
FLO—neurons	↑ neurons, ↓ spikes	↑ neurons, ↑ spikes
FLO—propagation	Yes, incremental	Yes, 500 ms
JAM—neurons	↑ neurons, ↑ impact	↑ neurons, ↑ impact
JAM—propagation	N/A	Yes, 600 ms

neurons. The results indicate that the threshold voltage is the most appropriate value to induce spikes in the targeted neurons. The second experiment studied the impact of attacking different neurons simultaneously, using a single voltage, particularly the threshold value. This experiment demonstrates that the greater the number of neurons attacked, the greater the number of spikes.

This work also studied for JAM the impact of attacking different numbers of neurons during a temporal window, in contrast to FLO, which is conducted at a certain time instant. This experiment demonstrates that the higher the number of neurons attacked, the lower the number of spikes. The results also highlight that, for both threads, the activity tends to revert to spontaneous behavior after the attack, aligned with neuroscience evidence. These results provide new approaches within the BCI cybersecurity field and could improve treatments for neurodegenerative diseases, specifically in the vision field.

Future work could implement the remaining neural cyberattacks to measure their impact, using the impact metric already defined in the literature and new ones to quantify how these attacks can affect spontaneous neuronal activity. Also, after its validation in the topology of 450 neurons, the cyberattacks and their configurations could be applied to a larger point-neuron topology, providing more realistic results in terms of the size of the neuronal sample. A natural evolution from that would be to evaluate this taxonomy of cyberattacks on a biophysically detailed simulation. Finally, future research could study the qualitative effect that these attacks could have on vision, based on the type of visual stimulus presented. Analyzing these stimuli is essential because if an attacker can capture neuronal activity when a person is in total darkness, it could recreate this situation in other contexts, causing temporal blindness to the person for the duration of the attack.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work has been partially funded by (a) the strategic project "Development of Professionals and Researchers in Cybersecurity, Cyberdefense and Data Science (CDL-TALENTUM)" from the Spanish National Institute of Cybersecurity (INCIBE) and by the Recovery, Transformation and Resilience Plan, Next Generation EU, (b) the Swiss Federal Office for Defense Procurement (armasuisse) with the CyberTracer (CYD-C-2020003) project, and (c) the University of Zürich UZH.

Availability of data and materials Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Improvements compared to the conference version The present publication includes an analysis of the current concerns and

limitations in new-generation neurostimulation systems able to stimulate and inhibit neural activity with single-neuron resolution. Moreover, this work improves the experiments performed in the previous conference paper, extending the design and implementation of the neural cyberattacks with a richer analysis. Finally, this publication includes a discussion of the results, highlighting the main differences with the state of the art.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Lebedev, M. A., & Nicolelis, M. A. L. (2017). Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation. *Physiological Reviews*, 97(2), 767–837. <https://doi.org/10.1152/physrev.00027.2016>
2. Zhao, W., Van Someren, E. J. W., Li, C., Chen, X., Gui, W., Tian, Y., Liu, Y., & Lei, X. (2021). EEG spectral analysis in insomnia disorder: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 59, 101457. <https://doi.org/10.1016/j.smrv.2021.101457>
3. Edwards, C. A., Kouzani, A., Lee, K. H., & Ross, E. K. (2017). Neurostimulation devices for the treatment of neurologic disorders. *Mayo Clinic Proceedings*, 92(9), 1427–1444. <https://doi.org/10.1016/j.mayocp.2017.05.005>
4. Musk, E. (2019). An integrated brain-machine interface platform with thousands of channels. *Journal of Medical Internet Research*, 21(10), 16194. <https://doi.org/10.2196/16194>
5. Li, Q., Ding, D., & Conti, M. (2015). Brain-computer interface applications: Security and privacy challenges. In *2015 IEEE conference on communications and network security (CNS)* (pp. 663–666). IEEE. <https://doi.org/10.1109/CNS.2015.7346884>
6. López Bernal, S., Huertas Celdrán, A., & Martínez Pérez, G. (2022). Neuronal jamming cyberattack over invasive BCIs affecting the resolution of tasks requiring visual capabilities. *Computers & Security*, 112, 102534. <https://doi.org/10.1016/j.cose.2021.102534>
7. López Bernal, S., Huertas Celdrán, A., Fernández Maimó, L., Barros, M. T., Balasubramaniam, S., & Martínez Pérez, G. (2020). Cyberattacks on miniature brain implants to disrupt spontaneous neural signaling. *IEEE Access*, 8, 152204–152222. <https://doi.org/10.1109/ACCESS.2020.3017394>
8. López Bernal, S., Huertas Celdrán, A., & Martínez Pérez, G. (2023). Eight reasons to prioritize brain-computer interface cybersecurity. *Communications of the ACM*, 66(4), 68–78. <https://doi.org/10.1145/3535509>
9. ...Arkipov, A., Gouwens, N. W., Billeh, Y. N., Gratiy, S., Iyer, R., Wei, Z., Xu, Z., Abbasi-Asl, R., Berg, J., Buice, M., Cain, N., da Costa, N., de Vries, S., Denman, D., Durand, S., Feng, D., Jarsky, T., Lecoq, J., Lee, B., ... Koch, C. (2018). Visual physiology of the layer 4 cortical circuit in silico. *PLOS*

- Computational Biology*, 14(11), 1–47. <https://doi.org/10.1371/journal.pcbi.1006535>
10. López Bernal, S., Huertas Celdrán, A., Martínez Pérez, G., Barros, M. T., & Balasubramaniam, S. (2021). Security in brain-computer interfaces: State-of-the-art, opportunities, and future challenges. *ACM Computing Surveys*. <https://doi.org/10.1145/3427376>
 11. Martinovic, I., Davies, D., & Frank, M. (2012). On the feasibility of side-channel attacks with brain-computer interfaces. In *Proceedings of the 21st USENIX security symposium* (pp. 143–158). USENIX
 12. Frank, M., Hwu, T., Jain, S., Knight, R.T., Martinovic, I., Mittal, P., Perito, D., Sluganovic, I., & Song, D. (2017). Using EEG-based BCI devices to subliminally probe for private information. In *Proceedings of the 2017 on workshop on privacy in the electronic society—WPES '17* (pp. 133–136). ACM Press. <https://doi.org/10.1145/3139550.3139559>
 13. Quiles Pérez, M., Martínez Beltrán, E. T., López Bernal, S., Huertas Celdrán, A., & Martínez Pérez, G. (2021). Breaching subjects' thoughts privacy: A study with visual stimuli and brain-computer interfaces. *Journal of Healthcare Engineering*, 2021, 5517637. <https://doi.org/10.1155/2021/5517637>
 14. Bonaci, T., Calo, R., & Chizeck, H. J. (2015). App Stores for the Brain: Privacy and Security in Brain-Computer Interfaces. *IEEE Technology and Society Magazine*, 34(2), 32–39. <https://doi.org/10.1109/ETHICS.2014.6893415>
 15. Camara, C., Peris-Lopez, P., & Tapiador, J. E. (2015). Security and privacy issues in implantable medical devices: A comprehensive survey. *Journal of Biomedical Informatics*, 55, 272–289. <https://doi.org/10.1016/j.jbi.2015.04.007>
 16. Ienca, M., & Haselager, P. (2016). Hacking the brain: Brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18(2), 117–129. <https://doi.org/10.1007/s10676-016-9398-9>
 17. Pycroft, L., Boccard, S. G., Owen, S. L. F., Stein, J. F., Fitzgerald, J. J., Green, A. L., & Aziz, T. Z. (2016). Brainjacking: Implant security issues in invasive neuromodulation. *World Neurosurgery*, 92, 454–462. <https://doi.org/10.1016/j.wneu.2016.05.010>
 18. Takabi, H., Bhalotiya, A., & Alohal, M. (2016). Brain computer interface (BCI) applications: Privacy threats and countermeasures. In *IEEE 2nd international conference on collaboration and internet Computing* (pp. 102–111). IEEE. <https://doi.org/10.1109/CIC.2016.026>
 19. Ienca, M., Haselager, P., & Emanuel, E. J. (2018). Brain leaks and consumer neurotechnology. *Nature Biotechnology*, 36(9), 805–810. <https://doi.org/10.1038/nbt.4240>
 20. Marin, E., Singelée, D., Yang, B., Volski, V., Vandenbosch, G.A.E., Nuttin, B., & Preneel, B. (2018). Securing wireless neurostimulators. In *Proceedings of the eighth ACM conference on data and application security and privacy. CODASPY '18* (pp. 287–298). Association for Computing Machinery. <https://doi.org/10.1145/3176258.3176310>
 21. Landau, O., Puzis, R., & Nissim, N. (2020). Mind your mind: EEG-based brain-computer interfaces and their security in cyber space. *ACM Computing Surveys*, 53(1), 1–38. <https://doi.org/10.1145/3372043>
 22. Martínez Beltrán, E. T., Quiles Pérez, M., López Bernal, S., Huertas Celdrán, A., & Martínez Pérez, G. (2022). Noise-based cyberattacks generating fake p300 waves in brain-computer interfaces. *Cluster Computing*, 25(1), 33–48. <https://doi.org/10.1007/s10586-021-03326-z>
 23. Chadderdon, G. L., Mohan, A., Suter, B. A., Neymotin, S. A., Kerr, C. C., Francis, J. T., Shepherd, G. M. G., & Lytton, W. W. (2014). Motor cortex microcircuit simulation based on brain activity mapping. *Neural Computation*, 26(7), 1239–1262. https://doi.org/10.1162/NECO_a_00602
 24. Ferguson, K. A., Njap, F., Nicola, W., Skinner, F. K., & Campbell, S. A. (2015). Examining the limits of cellular adaptation bursting mechanisms in biologically-based excitatory networks of the hippocampus. *Journal of Computational Neuroscience*, 39(3), 289–309. <https://doi.org/10.1007/s10827-015-0577-1>
 25. Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., Bilgili, G. A. A. K., Buncic, N., Chalimourda, A., Chindemi, G., Courcol, J.-D., Delalandre, F., Delattre, V., Druckmann, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2), 456–492. <https://doi.org/10.1016/j.cell.2015.09.029>
 26. Bezaire, M. J., Raikov, I., Burk, K., Vyas, D., & Soltesz, I. (2016). Interneuronal mechanisms of hippocampal theta oscillations in a full-scale model of the rodent ca1 circuit. *eLife*, 5, 18566. <https://doi.org/10.7554/eLife.18566>
 27. Bittner, S. R., Williamson, R. C., Snyder, A. C., Litwin-Kumar, A., Doiron, B., Chase, S. M., Smith, M. A., & Yu, B. M. (2017). Population activity structure of excitatory and inhibitory neurons. *PLOS One*, 12(8), 1–27. <https://doi.org/10.1371/journal.pone.0181773>
 28. Schmidt, M., Bakker, R., Shen, K., Bezgin, G., Diesmann, M., & van Albada, S. J. (2018). A multi-scale layer-resolved spiking network model of resting-state dynamics in macaque visual cortical areas. *PLOS Computational Biology*, 14(10), 1–38. <https://doi.org/10.1371/journal.pcbi.1006359>
 29. Crone, J. C., Vindiola, M. M., Yu, A. B., Boothe, D. L., Beeman, D., Oie, K. S., & Franaszczuk, P. J. (2019). Enabling large-scale simulations with the genesis neuronal simulator. *Frontiers in Neuroinformatics*. <https://doi.org/10.3389/fninf.2019.00069>
 30. Billeh, Y. N., Cai, B., Gratiy, S. L., Dai, K., Iyer, R., Gouwens, N. W., Abbasi-Asl, R., Jia, X., Siegle, J. H., Olsen, S. R., Koch, C., Mihalas, S., & Arkhipov, A. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, 106(3), 388–40318. <https://doi.org/10.1016/j.neuron.2020.01.040>
 31. Khabarova, E., Denisova, N., Dmitriev, A., Slavin, K., & Verhagen Metman, L. (2018). Deep brain stimulation of the subthalamic nucleus in patients with Parkinson disease with prior pallidotomy or thalamotomy. *Brain Sciences*, 8(4), 66. <https://doi.org/10.3390/brainsci8040066>
 32. Simpson, H. D., Schulze-Bonhage, A., Cascino, G. D., Fisher, R. S., Jobst, B. C., Sperling, M. R., & Lundstrom, B. N. (2022). Practical considerations in epilepsy neurostimulation. *Epilepsia*, 63(10), 2445–2460. <https://doi.org/10.1111/epi.17329>
 33. Opie, N. L., John, S. E., Rind, G. S., Ronayne, S. M., Wong, Y. T., Gerboni, G., Yoo, P. E., Lovell, T. J. H., Scordas, T. C. M., Wilson, S. L., Dornom, A., Vale, T., O'Brien, T. J., Grayden, D. B., May, C. N., & Oxley, T. J. (2018). Focal stimulation of the sheep motor cortex with a chronically implanted minimally invasive electrode array mounted on an endovascular stent. *Nature Biomedical Engineering*, 2(12), 907–914. <https://doi.org/10.1038/s41551-018-0321-z>
 34. Wirdatmadja, S. A., Barros, M. T., Koucheryavy, Y., Jornet, J. M., & Balasubramaniam, S. (2017). Wireless optogenetic nanonetworks for brain stimulation: Device model and charging protocols. *IEEE Transactions on NanoBioscience*, 16(8), 859–872. <https://doi.org/10.1109/TNB.2017.2781150>
 35. Hassija, V., Chamola, V., Bajpai, B. C., & Zeadally, S. (2021). Security issues in implantable medical devices: Fact or fiction? *Sustainable Cities and Society*, 66, 102552. <https://doi.org/10.1016/j.scs.2020.102552>
 36. Yaqoob, T., Abbas, H., & Atiquzzaman, M. (2019). Security vulnerabilities, attacks, countermeasures, and regulations of

- networked medical devices: A review. *IEEE Communications Surveys & Tutorials*, 21(4), 3723–3768. <https://doi.org/10.1109/COMST.2019.2914094>
37. Das, S., Siroky, G. P., Lee, S., Mehta, D., & Suri, R. (2021). Cybersecurity: The need for data and patient safety with cardiac implantable electronic devices. *Heart Rhythm*, 18(3), 473–481. <https://doi.org/10.1016/j.hrthm.2020.10.009>
 38. National Institute of Standards and Technology. (2014). NIST Cybersecurity Framework. National Institute of Standards and Technology. <https://www.nist.gov/cyberframework>
 39. U.S. Food and Drug Administration (2018). Content of Premarket Submissions for Management of Cybersecurity in Medical Devices. Technical report, U.S. Food and Drug Administration
 40. Schwartz, S.B. (2018). Medical device cybersecurity through the FDA lens. In *27th USENIX security symposium*. USENIX Association.
 41. U.S. Food and Drug Administration. (2016). Postmarket Management of Cybersecurity in Medical Devices. Technical report, U.S. Food and Drug Administration.
 42. Spanish Association for Standardization. (2020). Medical devices—Application of risk management to medical devices (UNE-EN ISO 14971:2020). Spanish Association for Standardization. <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0064637>
 43. UL Solutions. (2018). U.S. FDA Recognizes UL 2900-2-1 for Use in Premarket Reviews. <https://www.ul.com/news/us-fda-recognizes-ul-2900-2-1-use-premarket-reviews>
 44. Dai, K., Gratiy, S. L., Billeh, Y. N., Xu, R., Cai, B., Cain, N., Rimehaug, A. E., Stasik, A. J., Einevoll, G. T., Mihalas, S., Koch, C., & Arkhipov, A. (2020). Brain modeling toolkit: An open source software suite for multiscale modeling of brain circuits. *PLOS Computational Biology*, 16(11), 1–23. <https://doi.org/10.1371/journal.pcbi.1008386>
 45. Squire, L., Berg, D., Bloom, F., Du Lac, S., Ghosh, A., & Spitzer, N. (2012). *Fundamental neuroscience* (4th Ed.) Elsevier Inc. <https://doi.org/10.1016/C2010-0-65035-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Victoria Magdalena López Madejska is a Ph.D. student in Computer Science at the University of Murcia, Spain. She received a B.Eng. degree in Computer Engineering, specializing in Information Technologies, and an M.Eng. in New Technologies in Computing. Her Ph.D. focuses on studying the impact of cyberattacks on realistic neuronal representations to define new approaches in the brain-computer interfaces cybersecurity field. Besides, her

research interests include neuroscience, data analysis, privacy, and security.



Sergio López Bernal is a post-doctoral researcher at the University of Murcia, obtaining his Ph.D. from this university. He holds a M.Sc. in Computer Engineering and a B.Sc. in Computer Engineering from the University of Murcia, and a M.Sc. in Architecture and Engineering for the Internet of Things from IMT Atlantique, France. His research interests include cybersecurity, brain-computer interfaces, artificial intelligence, and data analysis.



Gregorio Martínez Pérez received a Ph.D. degree in Computer Science at the University of Murcia, where he is Full Professor since 2014. His scientific activity is mainly devoted to cybersecurity and data science.



Alberto Huertas Celdrán is currently a Postdoctoral Researcher Associated with the Communication Systems Group at the University of Zurich, Switzerland. He received the M.Sc. and Ph.D. degrees in computer science from the University of Murcia, Spain.