POLITECNICO DI TORINO

Master's Degree in Nanotechnologies for ICTs

Master's Degree Thesis

Design of Superconducting Nanowire-Based Neurons and Synapses for Power-Efficient Spiking Neural Networks

Supervisors

Prof. Renato GONNELLI

Prof. Karl K. BERGGREN

Candidate

Matteo CASTELLANI

October 2020

Abstract

Information processing with very low power consumption and innovative computing paradigms are required for the development of modern technologies in which real-time elaboration of data is needed. Natural spiking neural networks are being explored for their speed and energy-efficiency. In these systems, spikes generated by neurons take the information, and synapses act as local memories and connections between neurons, allowing networks to learn and adapt to external stimuli.

Superconducting electronics for its intrinsic low energy dissipation is the perfect candidate for building bio-inspired neuronal systems. It has already been proposed a structure that mimics the spiking behavior of the neuron and can be based on two different superconducting devices, which are able to generate low-power pulses: (1) Josephson junctions [1]; and (2) NbN nanowires [2]. The former are widely used for their high operation speed and low power consumption. The latter are typically exploited for single-photon detectors (SNSPDs), but recently are emerging as a platform for new electronics, thanks to their ability to interface with high-impedance environments.

This work mainly focuses on the design, optimization, and characterization of the nanowire-based elements necessary for the realization of a spiking neural network. The spiking behavior of nanowire neurons has been demonstrated experimentally, but further work is needed to improve the controllability of their properties. An artificial synapse has not yet been fabricated and tested, but it has been designed, exploiting the presence of kinetic inductance, a particular effect of NbN nanowires (inductive synapse). It is able to reproduce some characteristics of its biological counterpart, like the variable connection strength, but still presents some lacks for the creation of large and versatile networks. A new structure developed to improves the performances of the inductive synapse is here proposed (nTron synapse), introducing the nano-cryotrons (nTron [3] and hTron [4]): nanowire-based comparators with tunable gain, that use the formation of a localized Joule-heated hotspot to modulate the current flow in a superconducting channel.

SPICE models of all the exploited superconducting devices were created, starting from experimental data and the existing model of SNSPDs, to facilitate a correct design of the nTron synapse and find limitations of the network. Moreover, fundamental elements of the neurons and nTron synapses like (1) large kinetic inductors, (2) shunted nanowires, and (3) nTrons, were fabricated and tested.

Electrical simulations were also performed to study in depth a possible integration of nanowire neurons with Josephson junction neurons. Merging the two technologies could be useful to increase the overall performances of the network, but it generates also some problems, that are here analyzed.

Acknowledgements

Many people have helped me during this journey, but in particular I would like to thank:

Prof. Karl K. Berggren for his guidance and for giving me the opportunity and freedom to constantly learn new things.

Emily Toomey for inspiring me with her work, and for her invaluable mentorship even in difficult periods.

Marco Colangelo for teaching me what it means to do research always with passion and fun, and for helping me to realize my ideas.

Owen Medieros for contributing to the completion of the project, and for being present when I needed help.

Ken Segall, Adam N. McCaughan, Jason Allmaras and Andrew Wagner for sharing their knowledge with me.

Dorothy Fleischer and Lara Ranieri for keeping all of us organized during these hard times.

All members of the QNN group who shared their knowledge with me and make me enjoy this journey.

All Nanotech students who shared with me the joys and pains of this master.

My roomates and friends Gabriel Giribaldi, Fabio Bersano, Dario Cattozzo Mor, Matilde Pavese and Marco Abrate with whom I had adventures and face problems in this period of pandemic. They helped me to live the lock-down in a better way.

Most of all to my parents, my grandmother and my brother, without whom I could never have finished this journey, for the support they always gave to me.

Table of Contents

List of Tables

List of Figures

Acronyms

FI

Fan-in

FO

Fan-out

hTron

Heater nano-cryotron

JJ

Josephson junction

JJ neuron

Josephson junction-based neuron

nTron

Nano-cryotron

NW

Superconducting nanowire

NW neuron

Superconducting nanowire-based neuron

RCSJ

Resistivity and Capacity Shunted Junction model

RSFQ

Rapid single flux quantum

SNN

Spiking neural network

SNSPD

Superconducting nanowire single photon detector

WTA

Winner-Takes-All

Chapter 1

Introduction

The human brain has long been studied for its capability to perform complex operations with robustness to errors and very low power consumption (10 times lower than a normal computer). Even if it executes less than 1000 operations/sec. [5] with a precision of 1 in 100, it is able to adapt to external stimuli and make decisions in real time, with low latency, thanks to a massively parallel computation paradigm. Traditional computers based on von Neumann architectures with a serial computation paradigm can perform about 10 billion operations/sec. with a precision of 1 in 4.2 billion (32-bit) [5], but struggle with tasks that require real-time elaboration of data and decision-making. The structure of the brain and its ability of learning has already inspired the creation of software-based artificial neural networks, that nowadays are intensively used for applications like patterns recognition, autonomous vehicles, and Internet of Things. They are very powerful and adaptable, but they are not energy-efficient, because implemented in standard hardware structures. The latter have intensively been improved over the years, but now are reaching their limits in terms of operation frequency, because of physical limitations and the bottleneck introduced by the access time of memories. Therefore it is necessary to drastically revolutionize the computing paradigm in order to reach the same performances of the brain. An alternative architecture that can allow the realization of power-efficient neural systems, is based on the concept of neuromorphic computing.

1.1 Neuromorphic computing

Neuromorphic computing is an innovative concept first proposed by Carver Mead in the 1980s, that revolutionizes the structure of standard hardware architectures, removing the net distinction between processing elements and memory. Recently, the interest in neuromorphic hardware has been renewed by the artificial intelligence

industry, for the growing demand of efficient neural networks and deep learning.

Multiple types of neuromorphic systems that mimic different aspects of the brain, were proposed. Among these systems, the artificial spiking neural networks (SNNs) follow the most bio-realistic approach, trying to replicate the spiking behavior of the neurons. In the following section the fundamental elements and concepts of biological networks, necessary to realize an artificial SNN, are introduced. Different technologies have already been explored to create SNNs, but the most promising in terms of energy-efficiency is based on superconducting electronics. Therefore in section 1.2 the advantages and standard devices of superconducting electronics are introduced. Superconducting SNNs have recently been designed, but their characteristics still need to be improved. Section 1.3 explains how they work and what are their pros and cons.

1.1.1 Biological spiking neural network

The biological spiking neural networks that form the brain are complex systems, composed by neurons connected to each other through synapses, that exploit trains of voltage spikes as carriers of information. The neuron is the central element of these systems and is formed by different parts: dendrites, soma and axon (see figure 1.1). The dendrites collect the input signal from other neurons and transmit it to the soma, that elaborates the signal generating voltage pulses (action potentials). These pulses propagate through the axon, which introduces a certain time delay, and reach the downstream neurons through the synapses [6].

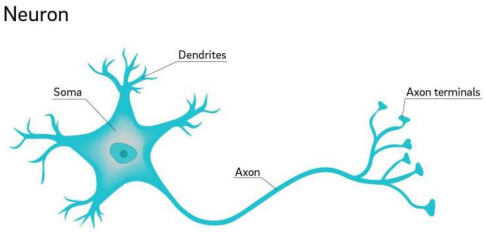

Figure 1.1: Structure of the neuron.

The spatial distribution of neurons, their connections and their behavior in the time domain define together the characteristic operation of a network [7]. The delay introduced by the axon is important to preserve time domain information. A too fast propagation could result in a lost of information [8].

Modeling of biological neurons

The critical part of the neuron is the soma, because it elaborates the signals. Indeed from the dendrites it receives the action potentials, that interacts with the ionic liquid enclosed in the cell membrane of the soma, and in turn generates others. In order to create a system that mimics a biological network is important to understand how the soma can be modeled.
The simplest phenomenological model for the soma is the 'Leaky integrate-and-fire' model. To a first and rough approximation the soma performs an operation of summation (integraton) of the input signals coming from the dendrites, and generates action potentials when a certain critical voltage is reached (firing). The information is not encoded in the shape of potentials, but in the presence or absence of them [6]. In the Leaky integrate-and-fire model, the cell membrane that surrounds the neuron is modeled by a characteristic capacitance and leaky resistance in parallel (RC-circuit). The incoming current generated by an action potentials charges the membrane with a time constant $\tau = R_m C_m$. If its voltage reaches the threshold, the membrane is reset to the resting potential u_{rest} generating a spike. If the input current is still present the dynamic is repeated, and the neuron continuously fires. Figure 1.2A shows the structure of the cell membrane and the circuit schematic of the model, while the 1.2B shows the cell membrane reaction to a step current.
The leaky integrate-and-fire model is useful to understand the passive integration behavior of the cell membrane and the general characteristics of the neuron, but it is highly simplified and can not describe many dynamics [6].

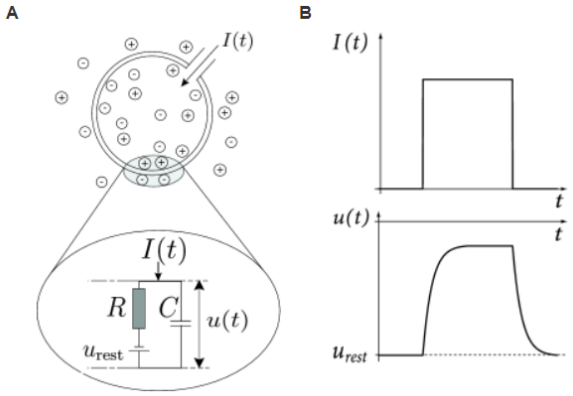

Figure 1.2: Passive membrane. **(A)** Cell membrane of the neuron, with a positive input current $I(t)$ which increases the electrical charge inside the cell. The membrane can be modeled as a capacitor with a parallel resistor in sseries a a battery with voltage u_{rest} (resting potential). **(B)** If a step current charges the membrane (top) a smooth voltage trace (bottom) is generated. Figure from [6].

In a more complete description the action potential of biological neurons is generated by two voltage-gated ionic currents that flow between the intracellular and extracellular medium of the soma. The inward Na^+ current accumulates ions into the neuron, rising the membrane potential, while the outward K^+ current removes ions, restoring the membrane to its resting potential (see 1.3A). If the neuron if continuously electrically stimulated, the combined action of these ionic currents generates repeatedly voltage spikes. This behavior is described by the Hodgkin-Huxley model, in which the membrane still has a characteristic capacitance and ions channels are modeled by voltage-controlled resistances as shown in figure 1.3B.

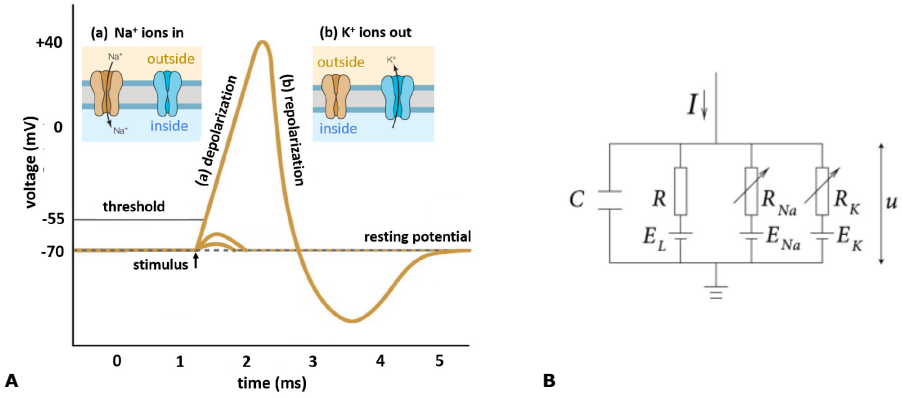

Figure 1.3: (A) Action potential of a biological neuron. The Na^+ ion channel influx of the cell membrane controls the rising edge (a), while the K^+ ion channel outflux controls the falling edge, restoring the membrane potential to its resting potential, so that the neuron can fire again. Figure from [9]. **(B)** Circuit schematic of the Hodgkin-Huxley model. C is the capacitor of the membrane and R is the leak resistor; R_K and R_{Na} are the voltage-controlled resistors that model the ion currents of the membrane; E_K, E_{Na} and E_L are battery voltages associated with the ion concentration differences between the inside and outside of the cell. u is the potential across the membrane, and $I(t)$ is the current injected into the neuron. Figure from [6].

Synapses

The synapses in a biological neuronal networks, can be direct electrical or chemical. The former permits current to flow passively through intercellular channels of a gap junction [10]. The latter is more complex: the action potential generated by the presynaptic neuron, activates the slow release of neurotransmitter molecules, which diffuse across the synaptic cleft and induce output current, binding to receptors of the postsynaptic membrane [1], as shown in figure 1.4. The chemical synapse

is able to amplify the input signal, while the electrical one can not. Moreover the chemical synapse can have an excitatory behavior if an input voltage pulse generates a positive action potential on the output, and an inhibitory behavior if it generates a negative one.

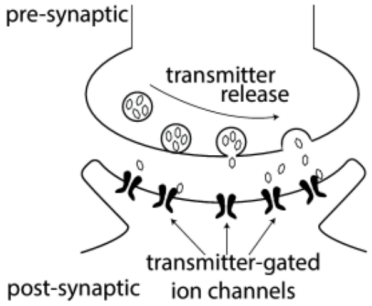

Figure 1.4: Illustration of a biological synaptic connection. When a pre-synaptic action potential arrives, the neurotransmitter is released into the synaptic cleft and captured by the post-synaptic receptors. Figure from [6].

The synapses, in addition to simply connect neurons, also have a key role in the capability of the network to learn and adapt to the external stimuli. Indeed the strength of the synaptic connection is influenced by the stimulation history. Incoming actions potentials can modify the number of presynaptic sites releasing neurotransmitter, changing the characteristic conductance of the synapse. This dynamic is called 'plasticity', and can be at short and long-term. In the short-term case (order of milliseconds) if there are two incoming spikes, the response to the second spike is larger or smaller than that to the first spike [6]. The long-term plasticity modifies the strength of the synapse for longer times, so it is related to the memory capability of the network.

Network structure

In a biological neural network external stimuli are elaborated in parallel by multiple neurons that are extremely interconnected. Indeed a single neuron can communicate with thousands of others. These characteristics are reproduced in artificial SNNs introducing a simplified network structure in which the neurons are grouped in different layers. The first layer receives the external signals, while the last one generate the output response. Each neuron in the intermediate layers collects signals from a certain number of neurons of the previous layer (fan-in), and can drive another number of neurons of the subsequent one (fan-out) through synapses that have different strengths (or weights). Figure 1.5a shows the typical structure of a neural network in which every connection between two layers is present (complete

network). A real network can have a more complex structure with only some connections active and eventually feedback signals. Figure 1.5b shows schematically the concepts of fan-in and fan-out.

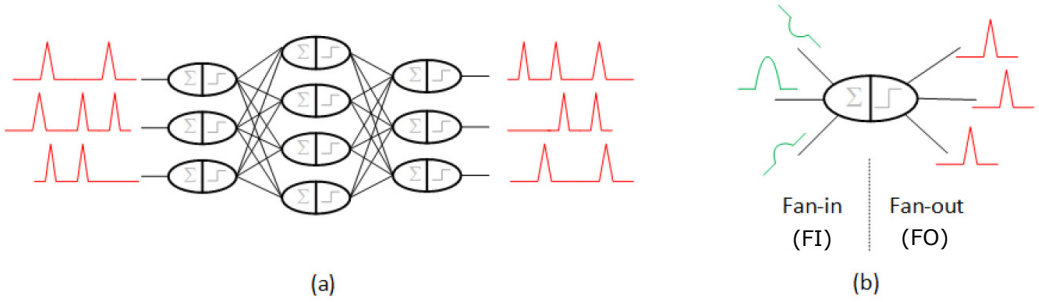

Figure 1.5: (a) Simplified schematic of a SNN. The neurons are indicated as ovals and the connecting lines can be associated with the synapses. The neurons can be divided into two parts like in the integrate-and -fire model: the input half, which sums the inputs (integration), and the output half, which generates a spike if the threshold is exceeded (firing). Both the input and output signals of the network are patterns of spikes in time. **(b)** The fan-in (FI) is the number of signals summed at the input of the neuron, and the fan-out (FO) is number of output lines that are driven by the neuron. The input signals can have different weight, and can be positive or negative, while the output signals have all the same amplitude and shape. Modified figure from [11].

In order to realize an efficient artificial SNN with the same level of parallelism of the brain, it is critical to maximise the levels of fan-in and fan-out.

1.1.2 Codification of information

In biological spiking neural networks the information is encoded in the action potentials. This is the main aspect that implemented in artificial SNNs, could allow to revolutionize the computing paradigm and drastically decrease the power consumption. In standard computers the information is encoded in bits, and processors continuously need a active clock signal to work properly, while in SNNs the computation is event-driven, so the network does not consume much power when inactive. The main problem of using spikes is that the learning algorithm to train the network are much more complex, and still need to be improved.

It has been observed that biological systems can use different type of codification, according to the tasks they have to perform. Here the most relevant typology of encoding that could be result useful in artificial SNNs are listed:

- **Rate coding:** This is the simplest kind of codification in which the information of a stimulus is encoded in the spiking frequency of the neurons. It ignores every data possibly encoded in the temporal pattern, and only considers the average firing rate, resulting to be very robust to noise. This technique is mainly used by neurons connected to sensory systems, and can be easily implemented in artificial SNNs but has high latency.

- **Temporal coding:** Most of the neural code is identified as a temporal code, in which the information stays in the precise spiking time of the action potentials, and the relative time intervals between subsequent spikes. The intensity of a stimulus is translated in the firing time of the neuron: higher is the signal, earlier the neuron fires. This technique is less robust to noise, so it is difficult o implement it in hardware, but its latency is much lower and more information, that would be lost with rate coding, are encoded. It was demonstrated by simulations in [12] that a network can be trained for image recognition on the MNIST dataset encoded in time, reproducing the results of conventional software-based networks.

- **Population coding:** With this method the stimuli are represented by the joint activities of certain number of neurons. Each neuron has a specific response to a set of inputs and the combination of all the responses defines the values associated with the inputs. The population coding is mainly in sensor and motor areas of the brain [13], and can be well described mathematically.

- **Sparse coding:** In this method an item is encoded in the strong activation of a certain area of the brain. Sparse code algorithms are used to find representative patterns in a set of input patterns. The language can be encoded with this method.

1.2 Superconducting electronics

Superconductivity is a physical phenomena observed is certain materials, for which electrical resistance vanishes and magnetic fields are expelled from the material, when the temperature is below a characteristic critical value T_c [14].

In a superconductor electrons can be attracted to each other due to the electron-phonon interaction, generating the Cooper pairs. The Cooper pairs act as charge carriers and can be considered as bosons that are described by the collective wavefunction $\Psi = \Psi_0 e^{i\phi(r,t)}$, where ϕ is the spatially and temporally dependent phase.

Superconductors are used in digital and analog electronics for their very low power consumption and their unique non-linear behavior. The most prominent

device is the Josephson junction, which is the building block of rapid single flux quantum electronics (RSFQ) [15].

1.2.1 Josephson junction

The Josephson junction is composed by two superconducting layers separated from each other by a normal or insulating layer. In 1962, B.D Josephson predicted that the Cooper pairs of the two superconducting layers could tunnel through the central barrier without generating voltage drop on the junction, if the current is below its critical value I_c (zero-voltage state). When the current exceed I_c, a voltage is developed and the current starts oscillating with a voltage-dependent frequency (voltage-state)[16]. The voltage-dependence of the frequency allowed to exploit the JJs as voltage standard [17]. Figure 1.6 shows the structure of the Josepshon junction, and its characteristic I-V curve.

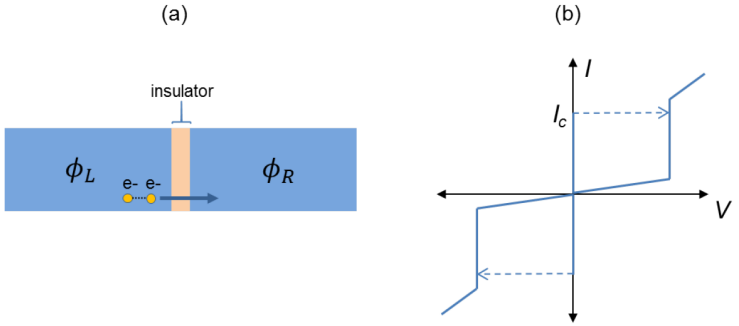

Figure 1.6: **(a)** Simplified schema of a Josepshon junction. The tunneling of a Cooper pair through the insulator is shown. Φ_R and Φ_L are the superconducting phases. **(b)** Typical I-V curve of a Josephson junction. Figure from [18].

Due to the presence of the barrier, a phase difference ($\varphi = \Phi_L - \Phi_R$) is maintained between the left and right superconducting layers, if a bias is applied to the junction. The two basic equations that determine the behavior of the superconducting current I_s in the junction are:

$$I_s = I_c \sin \varphi \tag{1.1}$$

and

$$\frac{d\varphi}{dt} = \frac{2\pi V}{\Phi_0} \tag{1.2}$$

where V is the voltage across the junction and $\Phi_0 = h/2e$ is the magnetic flux quantum.

In the voltage-state, a normal current I_n is generated due to the creation of quasiparticles. This current is a resistive current: $I_n = V/R(V)$. When $|V|$ is lower

than the characteristic voltage gap V_g of the material the current is generated due to thermal excitation of quasiparticles, and $R(V)$ is equal to the sub-gap resistance R_{sg}. If the voltage exceeds V_g, the quasiparticles are able to pass through the junction and R becomes the normal resistance R_N. When dV/dt is non-zero, also the displacement current $I_d = CdV/dt$ has to be considered, where C is the junction capacitance. Introducing I_n and I_d, it is possible to define the "Resistivity and Capacity Shunted Junction" model (RCSJ), whose current is described by the following equation [19]:

$$I = I_c \sin\varphi + \frac{V}{R} + C\frac{dV}{dt} \tag{1.3}$$

that can be rewritten as:

$$i = \sin\varphi + \Gamma\frac{d\varphi}{d\tau} + \frac{d^2\varphi}{d\tau^2} \tag{1.4}$$

where $i = I/I_c$, and $\tau = t\sqrt{\Phi_0 C/2\pi I_c}$ are respectively the normalized current and time. $\Gamma = \sqrt{\Phi_0/2\pi I_c R^2 C}$ is the normalized damping factor of the junction [1]. If $\Gamma \gg 1$, the junction is overdamped, this means that when $I > I_c$ and the current is decreased until $I < I_c$, the zero-voltage state is reached almost instantly. If $\Gamma \ll 1$, the junction is underdamped, so the voltage oscillates before reaching the zero-voltage state after a very long time.

1.2.2 Superconducting nanowire

The Josephson junction is a powerful technology, which is widely used in superconducting circuits, but its fabrication process generally is not simple. Moreover it is difficult to integrate Josephson junctions with standard electronics. These problems can be overcame by using superconducting nanowires, monolithically patterned as strip on thin films (usually in NbN), which do not exploit the tunneling of Cooper pairs like Josephson junctions. In a nanowire the supercoductivity can be broken down when the temperature, magnetic field or current exceed their critical values. The breakdown of superconductivity is not a coherent phenomenon as it would be in a Josephson junction, so the nanowire can not be used for quantum applications. During the breakdown a Joule heated region of hot quasiparticles, named hotspot [20], is generated in the nanowire. The boundary between normal and superconducting state of the hotspot can expand through the entire length of the nanowire, causing it to become normal and latch. Considering that the normal impedance of the nanowires is usually in the order of $\sim 1 - 10\ k\Omega$, they are ideal for amplification and fan-out.

If the nanowire is current biased slightly below its critical current, The hotspot can be created by a photon impinging on the nanowire, that generates a locally

increase of temperature, and consequently a decrease of the critical current. Exploiting this phenomenon, superconducting nanowires with planar meandered structures (see figure 1.7a) can be used as single photon detectors (SNSPD), as demonstrated in 2001 by Gol'tsman et al [21].

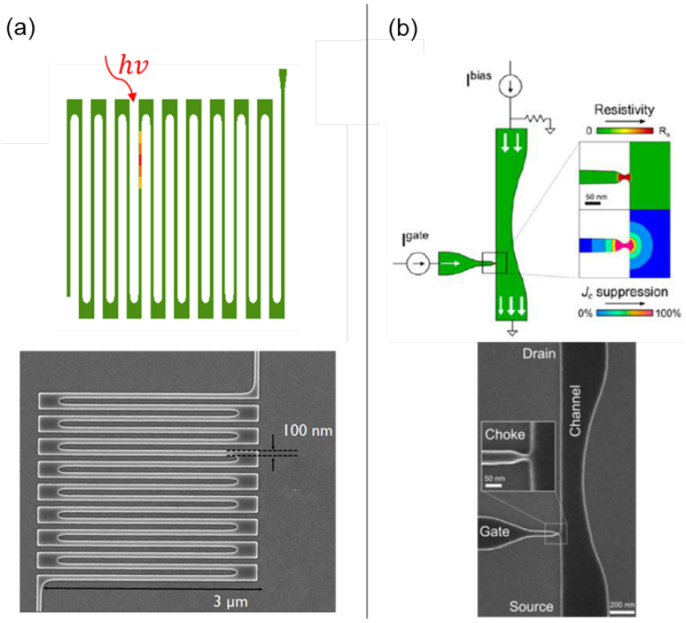

Figure 1.7: Examples of superconducting nanowire devices. **(a)** (top) Schematic of an SNSPD, formed by a meandered nanowire on which a phonon impinges generating a local increase of temperature (hotspot). (bottom) SEM image of a fabricated SNSPD. **(b)** (top) Schematic of the nano-cryotron (nTron). Injecting current into the gate choke the vertical nanowire (channel) can be made switch for Joule heating (explained in section 2.1). (bottom) SEM image of the nTron with small inset showing the gaet choke. Figure from [18].

The nanowires can be also used in superconducting electronics to perform logic operations, through Nanowire-based cryotrons like the Nano cryotron [3] and the Heater nano cryotron [4] (explained in section 2). They exploit and manipulate the hotspot formation in order to act as switches or amplifiers for superconducting digital circuits (see figure 1.7b).

Kinetic inductance

A particular characteristic of nanowire devices, is the contribution of the kinetic inductance to the total impedance. The phenomenon of the kinetic inductance does not depends on magnetic phenomenona, but derives from the intrinsic inertia

of electrons in the material. It is present in every conductors, and can be described by the Drude model, that defines a complex conductivity, in which the imaginary part contains the inductive contribute:

$$\sigma = \frac{ne^2\tau}{m(1+i\omega\tau)} = \frac{ne^2\tau}{m(1+\omega^2\tau^2)} - i\frac{ne^2\omega\tau^2}{m(1+\omega^2\tau^2)} \tag{1.5}$$

where n is the carrier number density, m is the mass of the charge carrier, and τ is the main collision time of charge carries. In normal metals the collision time is very low, so $\omega\tau$ is much lower than 1. As consequence the imaginary part is negligible. In superconducting materials, $\tau \to \infty$ so the term of the kinetic inductance becomes dominant. The easiest way to compute the kinetic inductance in a superconductor is to relate the kinetic energy of Cooper pairs to the inductive energy:

$$\frac{1}{2}(2m_e v^2)(n_s l A) = \frac{1}{2}L_K I^2 \tag{1.6}$$

where $2m_e$ is the mass of a Cooper pair, n_s is the density of Cooper pairs, l i the length of the superconductor and A is the area of the cross section of the superconductor. Considering that $I = (2e)n_s v A$ it is possible to obtain the expression of the kinetic inductance:

$$L_K = \left(\frac{m_e}{2n_s e^2}\right)\left(\frac{l}{A}\right) \tag{1.7}$$

In general, the kinetic inductance is two orders of magnitude larger than the magnetic counterpart. This allows to realize scaled down inductive loops in superconducting electronics. It allows also to control the electrothermal feedback when large loads are connected to the nanowires, because the time constant of the feedback L_k/R depends on the inductance.

It was demonstrated that the kinetic inductance depends on the bias current of the nanowire, because the density of Copper pairs in the nanowire depends on the current passing through it [22]. Experimentally the inductance can be almost 20% higher than the zero-bias inductance, when the current approaches the switching current I_{sw}. The relation between the kinetic inductance and the current i_D is [23]:

$$L_k(i_D) = \frac{L_0}{2\cos\left(2\arccos\left(0.6 i_D/I_{sw}\right)/3\right) - 1} \tag{1.8}$$

Simplified physical model

A simplified model, that does not introduce a physical description of the hotspot formation, can be introduced to understand the behavior of the nanowire [23].

Figure 1.8 displays a schematic I-V curve of the nanowire, in which three different states can be defined: (1) the superconducting state; (2) the hot-spot state; (3) the normal state.

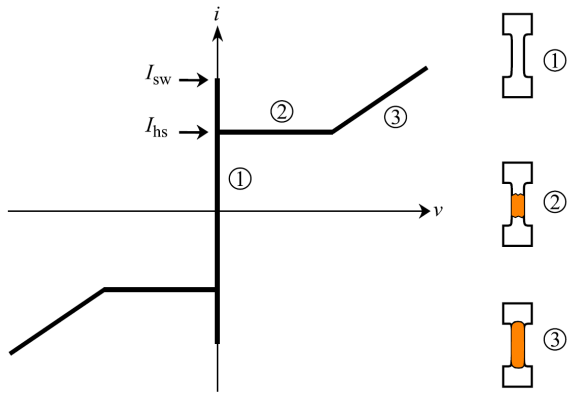

Figure 1.8: Schematic I–V curve of a superconducting nanowire The characteristic operation regions are shown: (1) superconducting state; (2) hotspot state; and (3) normal state. Figure from [23]

When the bias current is lower than the switching current the superconducting state occours, so the entire nanowire is superconductive, and the voltage drop is zero. In this state the nanowire electrically behaves like a kinetic inductor.

The nanowire enters the hotspot state, when an extended and sustained region becomes normal. After the formation of the hotspot the voltage drop on the nanowire is not more zero and the normal/supercondcucting boundary expands through the entire nanowire, if additional voltage continues to be applied. In this state the current remains almost constant, due to the balance between Joule heating and cooling through the substrate at the boundary.

When the entire nanowire is normal it enters the normal state, behaving like a resistor. The expansion of the hotspot along the length of the nanowire can be described by a phenomenological model developed by Kerman et al [24], that defines a current-dependent velocity of the superconducting/normal-phase boundary.

Usually the width of the nanowire is in the order of magnitute of the diffusion length of the superconductor ($\sim$ 100 nm). This means that the hotspot instantly expands along the width of the nanowire, so only the expansion in time along the length is considered. When $i > I_{sw}$ the length of the hotspot l varies with the following velocity:

$$v_{hs} = \frac{dl}{dt} = 2v_0 \frac{\Psi i^2/I_{sw}^2 - 2}{\sqrt{\Psi i^2/I_{sw}^2 - 1}} \tag{1.9}$$

where v_0 is a characteristic velocity that depends on the thermal parameters and the thickness of the film, and Ψ is the Stekly parameter. Ψ characterizes the ratio between the heat generated through Joule effect by the hotspot and the heat dissipated through the substrate. The factor 2 is introduced to consider that the hotspot propagates in both the directions if the nanowire is symmetric. This equation shows that the hotspot can also stay in a steady state in which v_{hs} is zero, and the nanowire remains in the hotspot or normal state. This phenomena happens for a certain value of current named I_{ss}, for which the local power density becomes equal to a fixed value. If it is greater the hotspot grows, if it is lower it contracts [24].

Shunting and relaxation oscillations

It is relatively easy to fabricate superconducting nanowires, and their operation is very simple, but they are limited by Joule heating. After the formation of the hotspot, for a correct operation usually it is required that the nanowire returns quickly to the superconducting state, with a certain reset time [18]. When the nanowire becomes normal, if the bias current is maintained constant, the Joule heating tends to keep the nanowire in the normal state if the heat dissipation is not strong enough. As consequence a slow and thermally dependent reset time is introduced. The heating also generates an hysteretic behavior, in fact after the formation of the hotspot, the current must be reduced below the retrapping current I_r instead of I_{sw}, in order to counteract the heating and reset the superconducting state. These problems can be solved shunting the nanowire with a resistor, which allows the bias current to divert from the hotspot, drastically decreasing the Joule heating. In this case the reset time depends on the discharging of the inductance between the nanowire and the shunt resistance, according to the formula L/R_s, where L is mainly composed by the kinetic inductance and R_s (or R_{sh}) is the shunt resistor.

Due to electrothermal feedback [24], the shunted nanowire can act in different ways after the switching event: (1) if R_s is too high, the timescale $\tau_e = L/R_s$ to restore the current in the nanowire after it is diverted to the shunt resistor becomes shorter than the timescale of the cooling down of the device. Therefore enough current (I_{ss}) flows into the nanowire to generate Joule heating before it is completely cooled down reaching the steady state, so that the nanowire latches; (2) if L/R_s is on the order of ~ 100 ps a single contolled pulse is generated. (3) If L is sufficiently large and the nanowire is biased with a current higher than I_{sw}, relaxation oscillations are obtained [2], as shown in figure 1.9 .

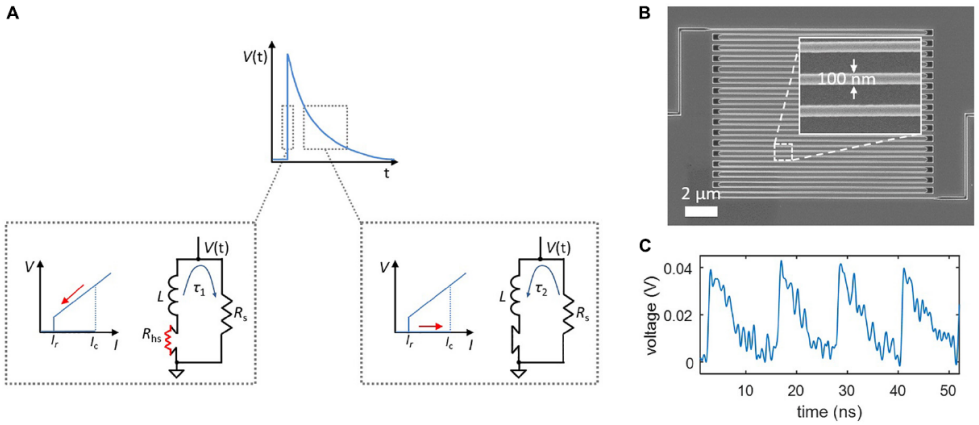

Figure 1.9: Relaxation oscillations in shunted superconducting nanowires. **(A)** Simplified model of a spike generated during relaxation oscillations. When the nanowire switches into the normal state ($R_{hs} > 0$), the bias current is diverted to the shunt resistor, so the rising edge occurs. When the nanowire return to be superconductive ($R_{hs} = 0$), the bias current flows back into it, so the falling edge takes place. **(B)** SEM image of a meandered nanowire designed to have high kinetic inductance. **(C)** Experimental relaxation oscillations in a meandered superconducting nanowire shunted by 50 Ω .Figure from [2]

In the relaxation oscillations, the rising time of each spike is much lower than the falling time, because just after the hotspot formation the total series resistance of L is $R_{hs} + R_s$, where R_{hs} is the hotspot resistance of the nanowire. As consequence the time constant of the rising edge is equal to $\tau_1 = L/(R_{hs} + R_s)$. After the rising edge, the current in the nanowire falls below the retrapping current, so that the nanowire can becomes superconductive again. The time constant of the falling edge is $\tau_2 = L/R_s$ because the hotspot is not longer present.

The total period of this kind of oscillations can be approximated as:

$$T = -\tau_1 \ln\left(\frac{I_r}{I_{sw}}\right) - \tau_2 \ln\left(\frac{I_{bias} - I_{sw}}{I_{bias} - I_r}\right) \tag{1.10}$$

In section 3 it is explained how a shunted nanowire can be modeled in LTspice, and the similarities with shunted Josephson junctions are shown. With well defined I_{sw} and L, the level of damping is determined by the shunt resistance. As for Josephson junctions, a lower resistance leads to an higher damping, so a lower spiking frequency.

1.3 Superconducting spiking neural network

Many hardware approaches for the spiking neural networks have been explored, including CMOS technology, magnetic materials and memristors. The CMOS circuits are the best option for large-scale integration, but their power consumption is much higher than the one of the human brain. Magnetic materials are not well integrated in a complete system with synapses and memristors have to be paired with CMOS technology [2]. Considering that low power dissipation both in the dynamic firing state and static state are strongly required, superconducting spiking neural networks, with their energy efficiency, have been introduced. The non-linear characteristics of superconducting devices allow to mimic the behavior of biological neurons with very few components. Superconducting neurons composed by Josephson junctions or shunted nanowires have been proposed. Both the devices are able to reproduce the behavior of the Hodgkin-Huxley model of biological neurons [6].

1.3.1 JJ Neuron

A structure for the soma based on Josephson junctions was proposed in 2010 by [1]. As shown in figure 1.10, it is composed by two Josephson junctions that are connected in a loop, and have a similar behavior of ion channels in biological neurons.

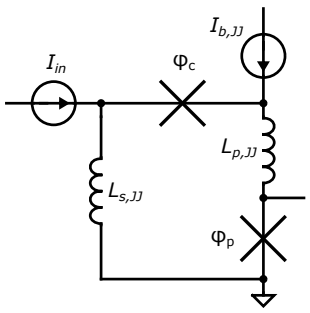

Figure 1.10: Circuit schematic of a JJ neuron

This structure derives from the superconducting logic circuit called "dc-to-SFQ converter" [25], and can mimic some neuronal behaviors like: action potential, firing threshold and refractory period [1]. A Josephson junction on its own can generate spikes similar to action potentials, but the voltage accumulates an offset if the bias current is increased. Considering that in a biological neuron the resting potential remains constant, the structure of the "dc-to-SFQ" turns out to be more suitable because it allows to keep an almost zero resting potential.

The junction on the right branch of the loop is named pulse junction (p), while the one on the left branch is the control junction (c). The inductances $L_{s,JJ}$ and $L_{p,JJ}$ are equal in order to perfectly split the bias current between the two junctions. Applying the current conservation and fluxoid quantization to the loop, it is possible to obtain the two equations that describe the behavior of the two junctions:

$$\sin\varphi_p + \Gamma\frac{d\varphi_p}{d\tau} + \frac{d^2\varphi_p}{d\tau^2} = -\lambda(\varphi_p + \varphi_c) + \frac{i_{in}}{2} + \frac{i_b}{2} \tag{1.11}$$

$$\sin\varphi_c + \Gamma\frac{d\varphi_c}{d\tau} + \frac{d^2\varphi_c}{d\tau^2} = -\lambda(\varphi_p + \varphi_c) + \frac{i_{in}}{2} - \frac{i_b}{2} \tag{1.12}$$

where $\lambda = \Phi_0/2\pi(L_{s,JJ} + L_{p,JJ})I_c$ is the parameter of the coupling between the phases of the two junctions, i_{in} and i_b are respectively the normalized input and bias currents. The loop is in equilibrium when $i_{in} = 0$, and the bias current equally splits between the two junctions biasing them just below their critical current. When i_{in} reaches the threshold, the control oscillator does not fire since the input current opposes the direction of the bias, while the current through the pulse junction exceeds I_c, generating a voltage pulse across it, that creates a magnetic flux in the loop. The flux induces current into the control junction, which starts whirling and make the pulse junction stop whirling. At this point, the system is ready to fire again, and this process can repeat continuously, if the input current is held above the threshold. The setting of L_s and L_p is important to define the coupling parameter. Lower is the total inductance of the loop, more coupled are the two junctions, so more current is generated in the other junction when one of them whirls. A too high coupling parameter could results in an unstable firing of the neuron.

The flux of the JJ neuron can be associated with the membrane potential of a biological neuron, and the voltages across the junctions corresponds to the ionic currents. Figure 1.11 shows the time profile of the flux in the JJ neurons, compared to the time profile of the action potential in the Hodgkin-Huxley model of biological neurons. The flux (defined as the time integral of the voltage) generated by the pulse junction for each spike is always equal to Φ_0.

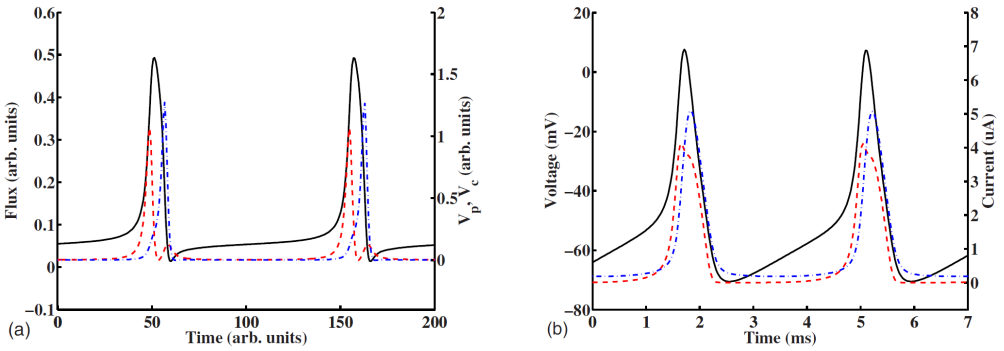

Figure 1.11: **(a)** Time-domain profile of action potentials in the JJ Neuron (simulation). The damping factor is $\Gamma = 1.5$. The input DC current is $i_{in} = 0.21$. **(b)** Time-domain profile of action potentials in the Hodgkin-Huxley model. The flux (black solid) in the JJ Neuron is associated to the membrane potential (black solid) of the Hodgkin-Huxley model. The voltages v_p (red dashed) and $-v_c$ (blue dot-dashed) in the JJ Neuron model are associated to the ion currents I_K (blue dot-dashed) $-I_{Na}$ (red dashed). Figure from [1]

In biological systems there exist two types of neurons that differ in their response to varying signal strength: Class I neurons have a spiking frequency that increases with increasing input current, while Class II neurons keep a constant firing rate [26]. In JJ neurons both these behaviors can be reproduced setting the value of the normalized dumping factor Γ: If $\Gamma > 1$ the JJ neuron is Class I, while for $\Gamma < 1$ it is Class II.

In order to completely reproduce the behavior of a neuron, it is required to introduce also the axon. For JJ neurons it is reproduced by a Josephson Transmission Line (JTL), that isolates the soma from the synapses and introduces a time delay [27].

1.3.2 NW Neuron

JJ neurons are fast and energy-efficient but have low fan-in/fan-out performances, so in large networks amplification components are needed. The action potential of JJ neurons is also difficult to be measured directly with standard electronics. For these reasons, the same structure of JJ neurons was implemented exploiting relaxation oscillators made by shunted nanowires instead of Josephson junctions [2]. The schematic is shown in figure 1.12. It is possible to use the same structure , because the shunted nanowires have I-V curves and dynamic behavior very similar to the ones of the junctions, as demonstrated by simulations in section 3. Even if the electrical behaviors are very similar, the two devices are based on different working principles as explained in previous sections.

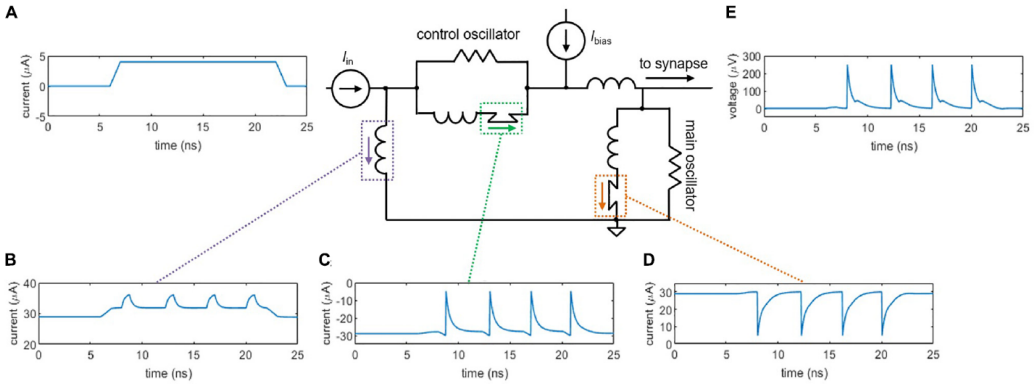

Figure 1.12: Circuit schematic of the NW neuron with results of simulations showing its time-domain behavior. **(A)** Input current pulse, $I_{in} = 4\ \mu A$. **(B)** Current through the loop inductor. **(C)** Current through the control oscillator, which reduces the amount of counterclockwise current in the loop, making the main oscillator fire. **(D)** Current through the main oscillator. **(E)** Output voltage that is sent to a synapse. Simulation parameters: $I_{sw} = 30\ \mu A$, $I_{bias} = 58.6\ \mu A$. Figure from [2].

In the NW neuron, as soon as the input current reaches the threshold, the main oscillator switches generating a voltage pulse. The current through it is diverted to the control oscillator in counterclockwise direction, summing with its bias current, so that the control oscillator can switch and the main one can stop firing. At this point the main oscillator is ready to fire again (see figure 1.12).

The NW neurons can reproduce the same characteristics of the biological neuron of the JJ neuron: action potential, firing threshold and refractory period. The shape of the action potential is slightly different from that of the JJ neurons, and the maximum spiking frequency is more than two order of magnitude lower. The spiking frequency is limited by the long reset time of shunted nanowires. It was observed a dependence of the spiking frequency on the bias current of the neuron, so the NW neuron is a Class I neuron [2].

The relaxations oscillators do not exploit a coherent phenomenon as in the case of Josephson junctions, so the flux released for each voltage spike depends on the switching current of the nanowires and the bias current of the loop. With $I_{sw} = 30\ \mu A$ the neuron generates a flux of $\sim 70\Phi_0$ per spike.

For NW neurons the behavior of the axon is reproduced by a long NbN nanowire, that is designed to act like a transmission line, thanks to its high kinetic inductance.

1.3.3 Inductive synapse

The inductive synapse (shown in figure 1.13A) introduced for the NW neurons, reproduces the slow release of neurotransmitters of a chemical synapse explained in section 1.1.1, with the slow charging of a large inductor [2].

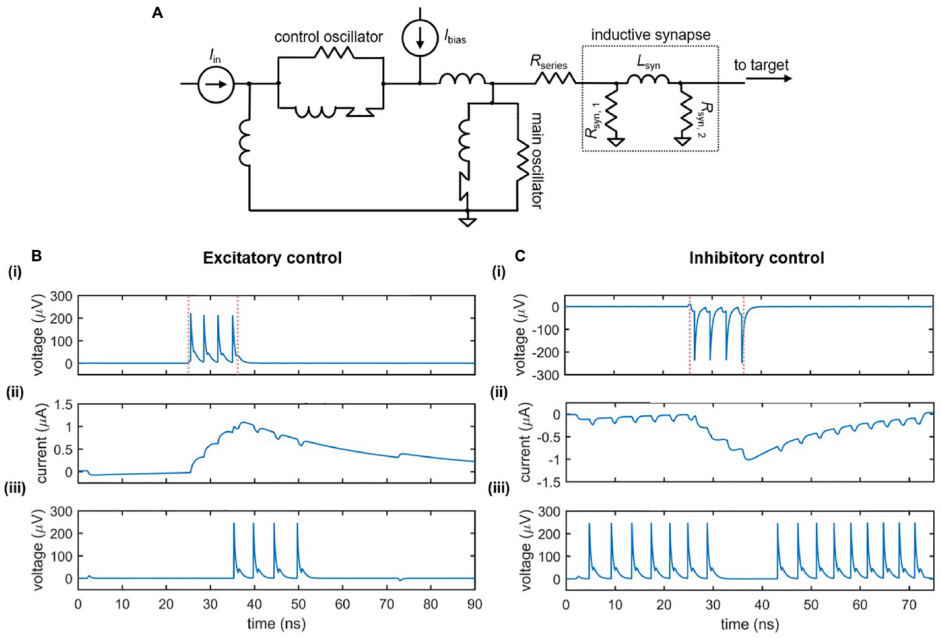

Figure 1.13: NW neuron with an inductive synapse. **(A)** Circuit schematic. The output current of the neuron charges the large synaptic inductor L_{syn}, that discharges providing current to the input of the target neuron, eventually making it fire. R_{series} is in placed to reduce the back-propagation of signal from the target to the main neuron **(B)** Excitatory control. Parameters: $I_{bias,main} = 59\ \mu A$, $R_{series} = 14\ \Omega$, $R_{syn,1} = 40\ \Omega$, $L_{syn} = 265\ nH$, $R_{syn,2} = 40\ \Omega$, $I_{bias,target} = 57.17\ \mu A$, $I_{in} = 4.6\ \mu A$. **(C)** Inhibitory control. Parameters: $I_{bias,main} = -58.6\ \mu A$, $R_{series} = 24\ \Omega$, $R_{syn,1} = 40\ \Omega$, $L_{syn} = 230\ nH$, $R_{syn,2} = 40\ \Omega$, $I_{bias,target} = 57.68\ \mu A$, $I_{in} = 4.6\ \mu A$. For both cases: Panel **(i)** displays the output voltage of the main neuron and the red dashed lines indicate the rising and falling edge of the input current signal; Panel **(ii)** displays the current through L_{syn}; Panel **(iii)** displays the output voltage of the target neuron. Figure from [2].

The large inductor can be obtained by a long NbN nanowire with high kinetic inductance. The inductor is charged by the spike generated by the upstream neuron and slowly discharged, providing a broaden pulse of current to the downstream neuron. If the presynaptic neuron is firing multiple times, and the time constant of the inductor is larger than the spiking period, the pulses can add up generating an higher output current. This kind of dynamic, shown in figure 1.13B, reproduce the integrating behavior of the integrating-and-fire model introduced in section 1.1.1. In a biological neuron the action potentials add up charging the membrane capacitance, while in NW neurons the spikes charge the synaptic inductors, and the output pulses of all the synapses electrically add up on the input of the downstream soma.

In biological systems the chemical synapse can have an excitatory or inhibitory control on the downstream neuron. Since the control depends only on the synapse, the spikes generated by the same neuron can contemporary control more target neurons with inhibitory and excitatory behavior. This characteristic is not reproducible by the inductive synapse. In order to achieve the inhibition, the main neuron needs to be biased with a negative current, so the control depends on the soma, instead of the synapse. Figure 1.13 shows the dynamic of two coupled NW neurons. The inductive synapse can be used also with JJ neurons, but the value of the inductance has to be reduced, since the Josephson junctions generates only one fluxon per spike, so a too small output current would be released.

Variable strength synapse

The ability to modulate the control by tuning the synaptic strength is a fundamental property in artificial neural networks. It allows to implement algorithms for unsupervised learning with long or short-term plasticity if the strength is internally controlled by action potentials, or just adapt the network to different tasks if it is externally controlled (supervised learning) [28]. The strength of the inductive synapse can be varied by changing the value of the kinetic inductance. As explained in section 1.2.2, the kinetic inductance can have an enhancement of almost 20%, if the current through the inductor is increased. This can be done posing an ideal current source in parallel with the synapse inductor as shown in figure 1.14. Ideally the value of the current source might be stored in integrated programmable memory cells based on superconducting loops [2].

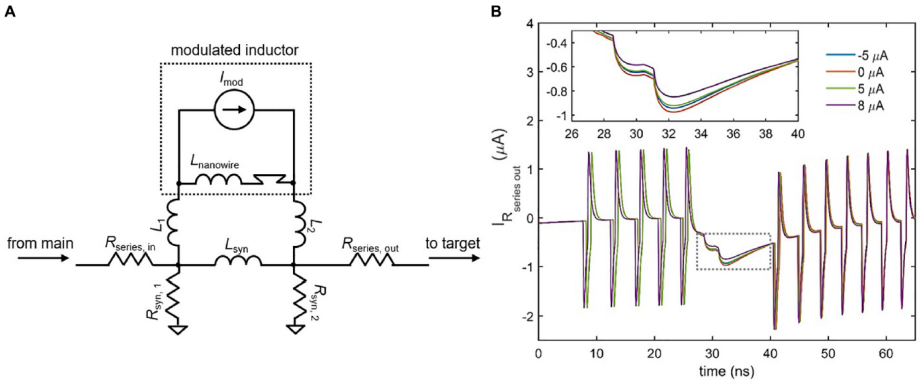

Figure 1.14: Variable synaptic strength. **(A)** Circuit schematic of the complete inductive synapse. A high-inductance nanowire and ideal current source are placed in parallel. L_1, $L_2 \ll L_{nanowire}$, L_{syn}. $R_{series,out}$ has been added to limit back-propagation from the target neuron. **(B)** Time-domain profile (simulation) of the current through $R_{series,out}$ in a synapse with inhibitory control as a function of the modulation current L_{mod}. The spikes are generated by the backaction from the firing target neuron. (Inset) Enlarged view of the boxed area. Increasing I_{mod} the overall synaptic inductance rises, so the current sent to the target decreases. Since with $I_{mod} = \pm 5\ \mu A$ the effect is nearly the same, the modulation is not polarity-dependent. Parameters: $R_{series,in} = 25\ \Omega$, $R_{syn,1} = 39\ \Omega$, $R_{syn,2} = 40\ \Omega$, $R_{series,out} = 0.1\ \Omega$, $L_{syn} = 450\ nH$, $L1 = L2 = 50\ pH$, $I_{bias,target} = 57.65\ \mu A$, $I_{bias,main} = -59.5\ \mu A$. Figure from [2].

Thanks to the high and controllable number of fluxons generated by the NW neurons and the tunability of the inductive synapses, it has been demonstrated by simulations that a NW neuron could contemporary control four different target neurons through inductive synapses with different synaptic strengths. The maximum fan-out (number of neurons controllable by a single one) is a fundamental figure of merit for the realization of a complete spiking neural network as will be explained in section 4.2.2, and NW neurons are more suitable for reaching high fan-out, respect to JJ neurons [2].

The energy dissipation of the neuron based on nanowires is about 50 aJ per action potential, of which the synapse contributes less than 5 aJ. However it is important to notice that in large systems if there are $O(N)$ neurons, there will be $O(N^2)$ synapses, so the power dissipation will be dominated by them.

It was recently demonstrated by simulations that using NW neurons and inductive synapses (with inductive coupling to introduce inhibitory control, as will be explained in section 4.1) it is possible to realize a neural network that recognizes

simple patterns on 9-pixel images [29] (as already done with JJ neurons [30]). In this kind of network the information is encoded in a deterministic way: if the pixel is black the associated input signal has a spike. If it is white no spike is generated. However biological neurons also present a stochastic behavior that is normally exploited in the brain: the activation of a neuron is not deterministic but has a certain probability to happens, that depends on the strength of the incoming action potential. This characteristic was observed experimentally also in NW neurons, indeed it was also designed and simulated a network made of NW neurons and inductive synapse [29], that applies the winner-takes-all (WTA) theory [31]. According to WTA the brain develops selectivity through competition between excitatory neurons, with firing probabilities, which share inhibitory connections. This functionality has been used also in artificial networks for image recognition, filtering and decision making.
The stochastic behavior of the NW neuron comes from the noise injected into the system by thermal and quantum fluctuations. In this work for sake of simplicity the noise effect is not taken into account, so the neurons and synapses are handled by a deterministic point of view.

1.3.4 Synapse alternatives

The inductive synapse has a structure that has been designed for NW neurons and allows to reproduces the basic behavior of a biological synapse, but it has a low tunability, and it is externally controlled. Alternative types of synapses have been proposed for JJ neurons. Studying the working principle of the following alternative synapses is useful as inspiration for a possible performances improvement of the inductive synapse.

Capacitive synapse

The first synapse architecture for JJ neurons was introduced by [1]. It reproduces the broadening of the action potential and the integrating behavior, exploiting a RLC resonant circuit.

The spikes generated by the upstream neuron charge the capacitor, that slowly releases current to the downstream neuron. As for the inductive synapse, the inhibitory control can be obtained only reversing the bias current of the upstream neuron. The strength of the synaptic connection can not be modified, once the synapse has been fabricated, so this architecture is not appropriate to create a complex adaptable network. The capacitive synapse can be used also with NW neurons, but the inductive version is easier to fabricate and integrate with nanowires electronics. The parameters of the inductive synapse were derived setting the time constant equal to the one of the capacitive synapse [9].

SQUID synapse

An improvement of the capacitive synapse for JJ neurons was proposed by [27]. It exploits a superconducting interference device (SQUID) to change the amplitude and the delay of the action potential before it starts charging the capacitor of the aforementioned capacitive synapse. The SQUID is a superconducting loop formed by two Josephson junctions, which is coupled to a magnet in this configuration. The amplitude of the output current of the SQUID, and the delay of the signal depend on the bias current of the loop, and the magnetic field applied to it. The magnet that generates the field is controlled by an on-chip current. The SQUID allows to introduce the variability of the connection strength, that can be externally controlled setting two current sources, into the capacitive synapse.

Magnetic JJ synapse

An interesting alternative that simplifies the structure of the synapse for JJ neurons is the magnetic JJ synapse, presented by [32]. Essentially the synapse is composed by a Josephson junction with magnetic nanoclusters of spins in the barrier. When there is an incoming action potential, the critical current is exceeded and the junction generates a spike, whose strength depends on the value of I_c. The critical current is strongly dependent on the nanoclusters configuration. In presence of an external magnetic field, the input action potentials can increase order of the nanoclusters, decreasing I_c. Without field, the order is decreased and I_c increased. This means that the weight of the synapse can be controlled internally by spikes, ideally making possible the implementation of unsupervised learning. The limitation of this structure is the need to have an external applied field that makes a real system not really unsupervised, and more complex to be practically realized.

Optoelectronic synapse

[7] proposed an optoelectronic artificial neural network based on JJ neurons, in which the signal is transferred through waveguides from a neuron to another, so that it is possible to reach very high fan-out. The synapse is designed to sense the optical incoming signal with an SNSPD. The spike generated by the sensor activates a Josephson junction biased slightly below its critical current, that fires and injects current in a superconducting loop in which it remains stored. The loop is inductively coupled with the JJ neuron so that it can discharges activating it. The number of fluxons generated by the firing JJ of the synapse depends on its bias current, therefore the strength of the synapse can be tuned. The bias current can be controlled by the incoming potentials through a dedicated additional optoelectronic circuit. This means that both supervised and unsupervised learning

can be implemented. The main problem is that this system is very complex and needs integration of different technologies.

1.4 Thesis goal

Superconducting electronics for its intrinsic low energy dissipation is the perfect candidate for building bio-inspired neuronal systems. The structure of superconducting neurons, introduced in section 1.3.1, mimics the spiking behavior of the neuron, and can be realized with Josephson junctions (JJ neuron) or shunted NbN nanowires (NW neuron). In this work the NW neuron is chosen as principal device for the realization of superconducting spiking neural networks. NW neurons have already been fabricated and tested [29], but further work is still necessary to completely characterize and optimize them, considering also that shunted nanowires are only recently being used for electronics. Here the first characterization results of new fabricated shunted nanowires and somas are presented.

The synapses are key components of a spiking neural networks, so it is fundamental to optimize them. The inductive synapse is able to reproduce some behaviors of the biological synapse, but present different problems that might be solved: (1) it is a passive component, so the output current depends on the structure of the network, and can be very low; (2) It is not possible to obtain an inhibitory control regardless of the bias current of the main neuron. The only way would be to introduce a transformer with negative coupling, as done for the synapse of [7]. Even with this modification it would not be possible to modify the type of control after the fabrication of the system; (3) The 20% tunability of the synaptic strength is low, if compared with the value obtained with alternatives synapses.

In this work a new structure that improves the performances, by solving the aforementioned problems, of the inductive synapse is proposed and in part characterized, in order to ideally make possible the realization of larger and more flexible networks. This is done introducing the Nano-cryotron (nTron), device explained in details in section 2.1, that acts as amplifier with tunable gain for the spikes generated by neurons. A similar device named Heater Nano-cryotron (hTron), in which the hotspot is generated through thermal coupling, is analyzed as possible substitute of the nTron, for its lower leakage currents. The final aim is to demonstrate that NW neurons with the improved synapse can be used as building blocks for large superconducting spiking neural networks. Therefore their electrical characteristics and limitations in terms of connectivity were studied and solutions to some of the encountered problems are here proposed.

This thesis presents also a deep electrical analysis, based on SPICE simulations, of a possible integration between the JJ-based and nanowire-based systems. The goal is: (1) to demonstrate that merging the two technologies can be useful to

increase the overall performances of the network, like the driving capability of a single neuron, and also (2) to analyze the limits and the drawbacks of this integration.

The thesis will be organized as follows:

Chapter 2 - Nanowire-based cryotrons: In this chapter the working principle and the key parameters of two innovative nanowire-based devices taht could be exploited for SNNs are described: the nano-cryotron (nTron) and the heater nano-cryotron (hTron).

Chapter 3 - Modeling: In order to demonstrate the functionalities of the new synapse and perform electrical simulations on more complex systems, SPICE models of all the considered superconducting devices were created, starting from experimental data and the existing model of the SNSPD [23]. In this chapter the models of Josephson junctions, nanowires, nTron and hTron are presented and explained in details.

Chapter 4 - Nanocryotron-based synapse: Design and electrical analysis: In this chapter the new synapse based on the nTron is introduced, showing its behavior through simulations results. All the choices made during the design are justified. Moreover electrical analysis are performed to show the fan-out/fan-in limitation, and possible solutions to the problems are proposed.

Chapter 5 - Nanowire-based neurons and synapses: fabrication and characterization Here the fabrication processes of all the devices necessary to create a NW-based SNN, with relative encountered issues, are described. The characterization techniques applied to shunted nanowires, NW neurons, nTrons and inductive synapses are explained, and some results are shown.

Chapter 6 - Integration of JJ and NW neurons: electrical analysis: This chapter shows the results of the SPICE simulations used to analyze pros and cons of the integration between JJ and NW neurons. Firstly it is demonstrated that the two different neurons can be coupled with different synapses. Then the fan-in/fan-out limitations of an hybrid system are studied.

Chapter 7 - Conclusion and outlook In this chapter the main results presented in this thesis are reviewed and possible future works are proposed.

Chapter 2

Nanowire-based cryotrons

The cryotron is a four-terminals superconducting device invented more than 50 years ago by Dudley Buck [33]. It is composed of two intertwined superconducting wires: The current passing through one of the two nanowire (the gate wire) induces a magnetic field that suppresses the superconductivity in the other nanowire (the channel), so that it is possible to control its resistance. Superconducting electronics based on RSFQ pulses is very powerful for its speed and low power consumption, but Josephson junctions are not able to drive large impedances, or generate digital signals and can not operate in noisy magnetic environments [3]. A superconducting electronics based on cryotron realized at the nanoscale could solve these problems. Two different types of nano-cryotrons that exploits non-magnetic phenomena to control the resistivity of the channel were proposed: the nTron and the hTron.

2.1 Nano-cryotron (nTron)

The nTron is a three-terminal thin-film NbN nanowire-based device formed by a gate, a source and a drain terminal, all directly connected without junctions. As shown in figure 2.1B, the gate nanowire perpendicularly intersects one side of the channel (the channel connects the drain and source) through a narrow bottleneck called choke. In the choke a localized, Joule-heated hotspot [20] can be formed if the critical current density J_c of the film is locally exceeded. The hotspot modulates the resistivity of the channel inducing a non-linear suppression of its critical current $I_{c,ch}$, so that a sharp transition to the normal state occurs. The hotspot effect can be produced in all superconductors [20], so the nTron could be realized also with different materials.

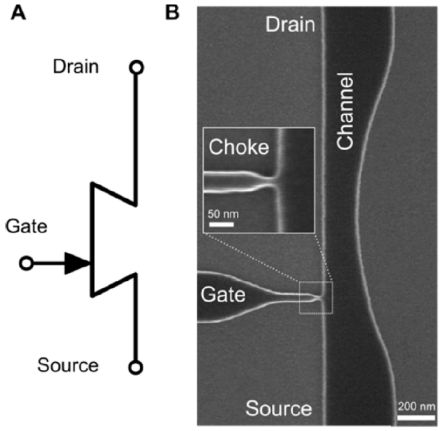

Figure 2.1: (A) Three-terminal circuit symbol of the nTron. The gate arrow is positioned to define the location of the choke respect to the channel. **(B)** SEM image of a nTron, the inset shows the choke, the region where the hotspot is generated. Figure from [3].

The operation of the nTron depends on its gate current and bias current of the channel. As shown in figure 2.2 the nTron can stay in three different states: OFF state, when $|I_{gate}| < I_{c,g}$ the channel is entirely superconducting; Transition state, when $|I_{gate}| \geq I_{c,g}$ the hotspot is formed and the superconductivity is locally suppressed; ON state, when the suppression is strong enough to lower $I_{c,ch}$ below the bias current of the channel, so that the hotspot expands along the channel and the resistive state is reached.

The expansion of the normal region is mainly caused by the out-diffusion of hot electrons from the hotspot to the surrounding material, that interact with the bath breaking Cooper pairs. This phenomenon can be described by the two-temperatures model [3], that uses two coupled heat equations to introduce the interaction between electrons and phonons [34].

The behavior of the nTron is strongly dependent on its geometry: the critical current of the gate is approximately $I_{c,g} = w_g d J_c$, for the channel it is $I_{c,ch} = w_{ch} d J_c$, where d is the thickness of the film, w_g the width of the choke and w_{ch} the width of the channel. The geometry shown in figure 2.1, designed by [3], has the channel that becomes narrower in proximity of the choke to ensure an acceptable suppression of the critical current. The hotspot tends to expand towards narrower regions in which the density of current is higher, so the choke is located on the bottom of the channel so that the hotspot can propagates to the drain and a good electrical isolation between the drain and the gate is obtained in the ON state.

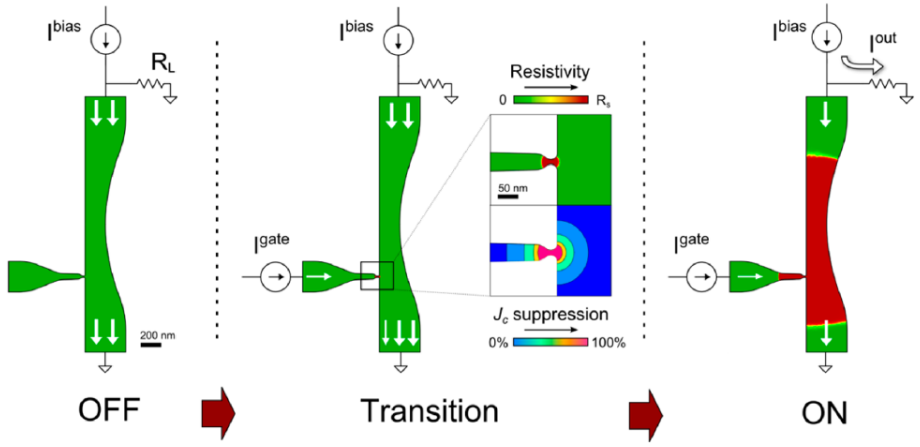

Figure 2.2: Three states of operation of the nTron. The shown dynamic was obtained with electrical simulations. OFF state: The device is fully superconducting, and all the bias current flows from the drain to the ground. Transition state: After the critical current of the gate is exceeded, a resistive hotspot that locally suppresses the superconductivity is formed. (Inset, top) Resistivity distribution in the choke. (Inset, bottom) Contour map of J_c suppression around the hotspot. From inner to outer, the bands represent reductions in J_c by 0% (blue), 25% (light blue), 50% (green), 75% (orange), and 99% (magenta). ON state: The critical current of the channel is lowered and the bias current is high enough to create a hotspot in the channel. Figure from [3].

An external load connected to the drain of the nTron acts like a shunt resistor for a nanowire, so in the ON state most of the bias current is diverted to it. Figure 2.2 shows the expansion of the normal region during the three states of the nTron, when a load R_L is present. It was demonstrated that the nTron can drive resistive loads from few Ω to 100 $k\Omega$. For resistances higher than about 50 Ω the channel works in the latching regime due to the electrothermal feedback, as explained in section 1.2.2, so the bias must be turned off and reset after each switching event, allowing the nTron to switch again. In digital circuits this can be done with a clock signal. If the circuit can not work in latching mode, it is possible to fabricate a long meandered nanowire in series with the channel, placed between the drain and the choke, so that the total inductance of the channel is increased and it is possible to drive larger loads.

Figure 2.3 from [3] shows the output current as function of the gate and channel currents, with a load of 10 $k\Omega$. The tested device has $I_{c,g} = 2.9\ \mu A$ and $I_{c,ch} = 106\ \mu A$. When $|I_{gate}| < I_{c,g}$ the the nTron switches at about $I_{ch} = I_{c,ch} - I_{gate}$ because the gate current adds up to the channel one. When $|I_{gate}| > I_{c,g}$ the channel critical current is modified by a factor that depends on $|I_{gate}| - I_{c,g}$.

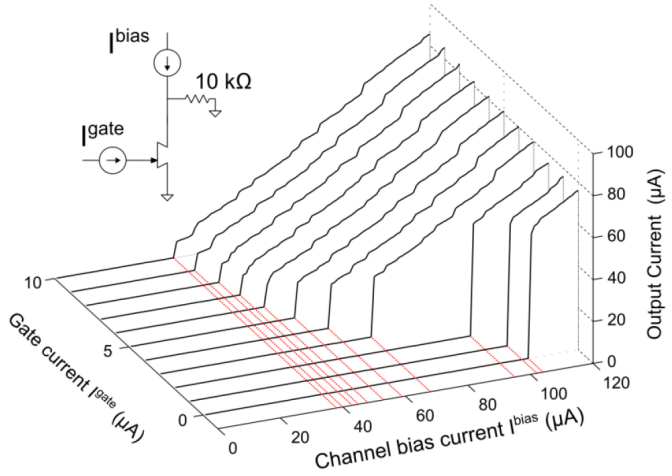

Figure 2.3: Circuit schematic of a nTron in non-inverting configuration (latching mode) and experimental graph showing the output current as a function of the bias and gate currents. I_{gate} was constant and I_{bias} was swept from 0 to 120 A. Figure from [3].

It was demonstrated that the nTron can be used as building block for superconducting digital circuits [3]. The NOT and the AND/OR gates were designed, and a possible structure for a half-adder was also proposed.

2.2 Heater nano-cryotron (hTron)

The nTron works correctly as superconducting switch, and has the advantage of a simple planar structure that can be easily fabricated on a single thin film. However it presents leakage currents since the gate is not electrically isolated from the channel, and it has poor fan-out performances [4]. It has been proposed a different kind a nano-cryotron that solves the problem of leakages and increases the fan-out capability, removing the electrical connection between gate and channel. This device is named P-hTron (planar) [35], and it is composed by two adjacent superconducting nanowires: the heater and the channel. When the critical current of the heater is exceeded, it becomes normal and heats the channel thanks to Joule effect. The increase of temperature lowers the critical current of the channel that can eventually switch if correctly biased. The behavior is very similar to the one of the nTron, but the physical phenomenon that modulate $I_{c,ch}$ is different. It has also been proposed an alternative structure that allows to drive higher impedance loads, in which a meandered channel is covered by a normal metal heater, and the two layers are isolated by a thin dielectric film [36]. With this device it is necessary to heat a very large area of superconductor and it was demonstrated only at 1 K. A different hTron that can solve the aforementioned problem is the multi-layer hTron (M-hTron) [4]. Its channel is a straight superconducting nanowire, and the heater is a normal metal nanowire that crosses the channel and it is placed over it. The two nanowires are isolated with a thin oxide film. This device has better performances than the P-hTron in terms of fan-out and thermal coupling. Figure 2.4 shows the typical structure of the M-hTron.

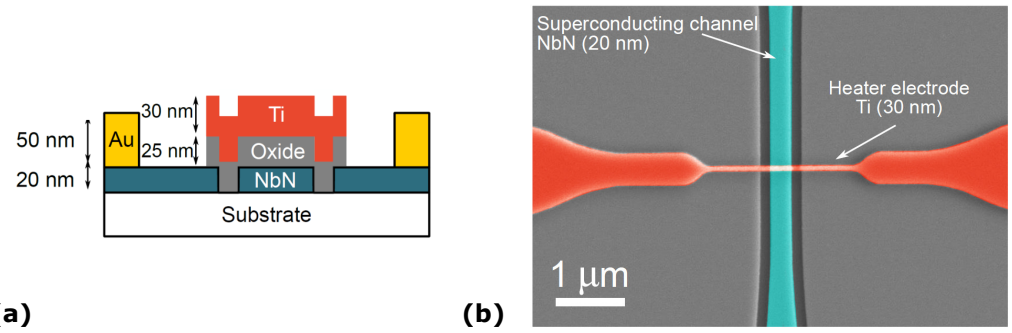

Figure 2.4: **(a)** Cross section of the M-hTron (The Au film composes the marks used for the lithography). **(b)** SEM image of the M-hTron. Figure from [4].

The superconducting channel was realized on a 20-nm thick NbN film, the dielectric layer is a 25-nm thick SiO_2 film and the nanowire of the heater is fabricated in Ti with a thickness of 30 nm. In this work, only the M-hTron shown in figure 2.4 is described and modeled.

It has been demonstrated that the hTron can be used as preamplifier for SNSPDs, and a memory cell formed by two coupled hTrons has been realized [4]. Moreover it has been shown by simulations that it can be used as oscillator with tunable frequency.

2.2.1 Heat transfer

All the following description was extracted by [4]. The heat transfer between the metal and the superconductor can be described by a quasi-equilibrium two-temperature model [4], like the hotspot formation in the nTron. Each material involved in the transfer has its own phonon and electron distributions and temperatures (T_{ph} and T_e), that exchange heat. The heat transfer in the metal heater is described by

$$C_e(T_e)\frac{\partial T_e}{\partial t} = -\Sigma_{e-ph}(T_e^5 - T_{ph}^5) + \nabla\kappa_e(T_e)\nabla T_e + \vec{j}\cdot\vec{E} \tag{2.1}$$

and

$$C_{ph}(T_{ph})\frac{\partial T_{ph}}{\partial t} = \Sigma_{e-ph}(T_e^5 - T_{ph}^5) + \nabla\kappa_{ph}(T_{ph})\nabla T_{ph} \tag{2.2}$$

where C_e and C_{ph} are the electron and phonon heat capacities, Σ_{e-ph} is the electron-phonon coupling factor, κ_e and κ_{ph} are the thermal conductivities, and $\vec{j}\cdot\vec{E}$ describes the Joule heating effect ($\vec{j}$ is the current density in the metal and $\vec{E}$ the electric field). The heat fluxes exchanged with the other materials is included in the equations by the term $\nabla\kappa(T)\nabla T$. Two analogous equations can be used to describe the superconducting channel, but in that case the Joule heating term is present only when the nanowire is normal. The dielectric can be modeled with just the phonon equation 2.3 with the electron-phonon interaction neglected. In all the phonon equations the acoustic mismatch model (AMM) [37] was used for the heat transfer between material 1 and 2, obtaining:

$$\kappa_{ph,1}(T_{ph,1})\nabla_\perp T_{ph,1} = -\frac{G_{12}}{4}(T_{ph,1}^4 - T_{ph,2}^4) \tag{2.3}$$

where $\kappa_{ph,1}$ is the local phonon thermal conductivity, $\nabla_\perp$ indicates the outward gradient in the direction normal to the interface, and G_{12} is the boundary conductance. Considering the equation associated with the electron and phonon systems of all the involved films, it is possible to define the complete 3D electrothermal model that describes the heat transfer in the device. Figure 2.5 show the block diagram of the model.

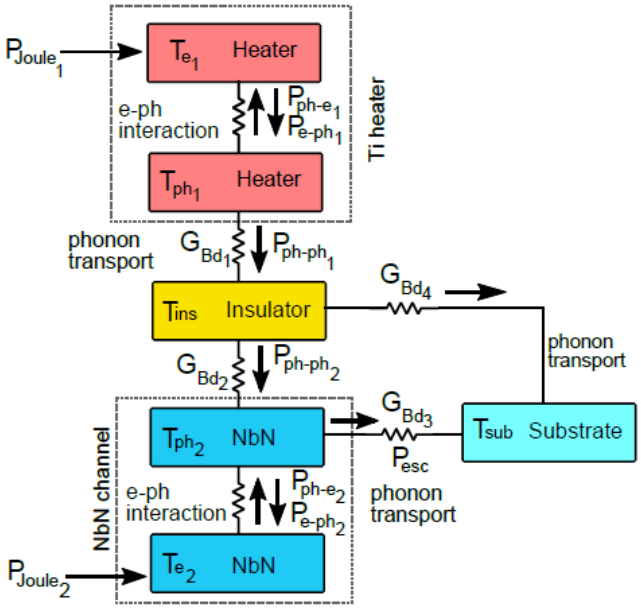

Figure 2.5: Block diagram of the electrothermal model described by the equations previously introduced. Figure from [4].

In this representation the heat can be modeled as a current ($P_{ph-e,1}$ etc. in figure), and the boundary conductances can be associated to thermal resistances ($G_{Bd,1}$, $G_{Bd,2}$, etc. in the figure). When a current I_H passes though the heater (H), the electron temperature $T_{e,H}$ ($T_{e,1}$ in the figure) is increased due to Joule effect ($P_{joule,1}$ in the figure), and the phonon system is heated thanks to the coupling. The phonons of the metal transfer heat to the dielectric (D) that in turn heats the phonons system of the superconducting channel (S), and the electron temperature $T_{e,S}$ ($T_{e,2}$ in the figure) increases. If $T_{e,S}$ exceeds the current-dependent T_c of the current biased channel, it switches generating Joule heat ($P_{joule,2}$ in the figure) and the current is provided to the load, like for the nTron. The channel cools down dissipating heat both to the dielectric and the substrate. If a meandered nanowire is added in series with the channel, the reset time is dependent on its kinetic inductance. It was observed experimentally that the hTron fabricated in [4], has a critical current that follows the expression:

$$I_c = 126.6 \left[1 - \left(\frac{T_{e,S}}{T_c} \right)^3 \right]^{2.1} \quad \mu A \tag{2.4}$$

Using the described complete electrothermal model, the hTron ability to amplify a current pulse was studied by [4], simulating the circuit shown in figure 2.6a with a load of 50 Ω, 500-nm wide heater and 600-nm wide channel.

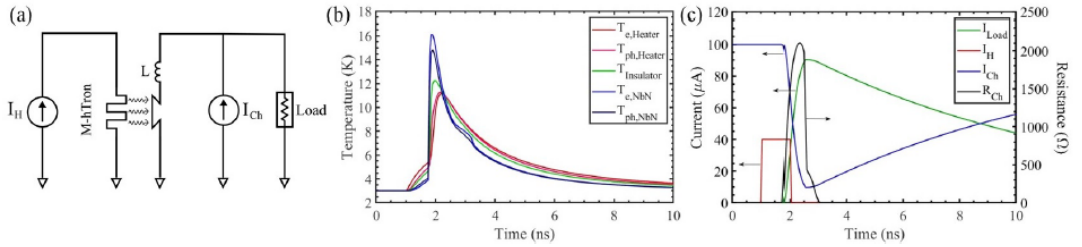

Figure 2.6: **(a)** Schematic of the circuit simulated using the complete elctrothermal model. A 1ns-long pulse of $I_H = 40\ \mu A$ is provided to the hTron. $I_{Ch} = 100\ \mu A$, $T_{sub} = 3\ K$. **(b)** Time-domain evolution of all the temperatures considered in the model **(c)** electrical response of the hTron. Figure from [4].

Figure 2.6b shows the time-domain evolution of the temperatures of all the materials after a 40 μA 1 ns-long current pulse is provided to the heater. Figure 2.6c shows that the hTron is able to sense the pulse generating a output current of 90 μA.

Simplified model

A simplified model that reduces the computation complexity for the M-hTron was introduced in [4]. In the limit of wide and thin heater, wide and thin channel and thin dielectric film, the entire system can be described by a 0D set of differential equations. This model is useful to estimate the thermal dissipation needed to make the channel switch. For the heater the equations are:

$$C_{e,H}(T_{e,H})\frac{\partial T_{e,H}}{\partial t} = -\Sigma_{e-ph,H}(T_{e,H}^5 - T_{ph,H}^5) + \frac{I_H^2 \rho_H}{(w_H d_H)^2} \tag{2.5}$$

and

$$C_{ph,H}(T_{ph,H})\frac{\partial T_{ph,H}}{\partial t} = \Sigma_{e-ph,H}(T_{e,H}^5 - T_{ph,H}^5) - \frac{G_{H-D}}{4d_H}(T_{ph,H}^4 - T_D^4) \tag{2.6}$$

where ρ_H is the heater resistivity, d_H the heater thickness and w_H the heater width. G_{H-D} is the heater-dielectric boundary conductance and T_D the phonon dielectric temperature. The electron heat capacity is linear dependent on the temperature: $C_{e,H} = \gamma_H T_{e,H}$. For the phonon heat capacity the Debye model is used: $C_{ph,H} = \alpha_H T_{ph,H}$.
For the dielectric the equation is:

$$C_D(T_D)\frac{\partial T_D}{\partial t} = \frac{G_{H-D}}{4d_D}(T_{ph,H}^4 - T_D^4) - \frac{G_{D-S}}{4d_D}(T_{ph,D}^4 - T_{ph,S}^4) \tag{2.7}$$

where d_D the dielectric thickness and G_{D-S} is the dielectric-superconductor boundary conductance. Also for the dieletrcic the Debye model can be used: $C_D = \alpha_D T_D$.
For the superconductor the equations are:

$$C_{e,S}(T_{e,S})\frac{\partial T_{e,S}}{\partial t} = -\Sigma_{e-ph}(T_{e,S}^5 - T_{ph,S}^5) \tag{2.8}$$

and

$$C_{ph,S}(T_{ph,S})\frac{\partial T_{ph,S}}{\partial t} = \Sigma_{e-ph,S}(T_{e,S}^5 - T_{ph,S}^5) - \frac{G_{D-S}}{4d_S}(T_{ph,S}^4 - T_D^4) - \frac{G_{S-sub}}{4d_S}(T_{ph,S}^4 - T_{sub}^4) \tag{2.9}$$

where d_S the superconductor thickness, G_{S-sub} is the superconductor-substrate boundary conductance and T_{sub} is the substrate temperature. The phonon heat capacity follows the Debye model ($C_{ph,S} = \alpha_S T_{ph,S}$), while the electron capacity is state-dependent: it increases exponentially with $T_{e,S}$ if $T_{e,S} > T_c$, otherwise it is linear dependent on the temperature.
Figure 2.7 shows the time-domain evolution of the temperatures obtained both with the simplified 0D model and the complete 3D model, with the heater current of 30 μA which turns on after 1 ns. The two results match almost perfectly.

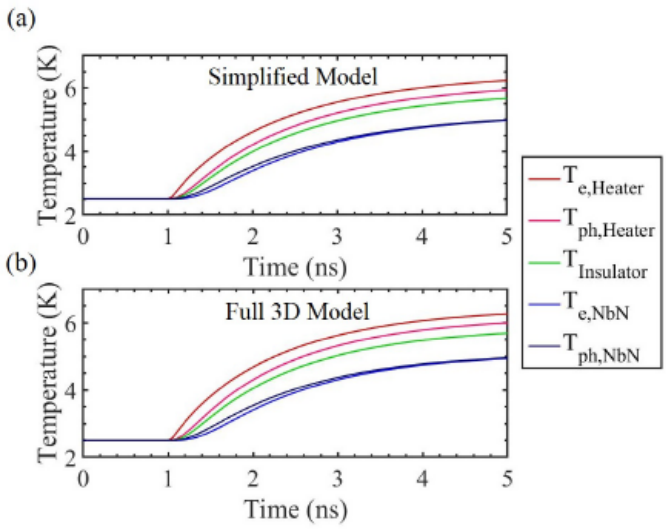

Figure 2.7: Time-domain evolution of the temperatures obtained with 500-nm wide heater and 600-nm wide channel and $I_H = 30\ \mu A$ after 1 ns. **a** Simplified 0D model. **(b)** Complete 3D model. Figure from [4].

In the steady state condition it is possible to compute the electron temperature of the channel as function of I_H:

$$T_{e,S}(I_H) = \left[\frac{4\rho_{\square,H}}{G_{S-sub}} \left(\frac{I_H}{w_H} \right)^2 + T_{sub}^4 \right]^{1/4} \tag{2.10}$$

where $\rho_{\square,H}$ is the sheet resistance of the heater. Once the electron temperature of the channel is known, the critical current can be computed with equation 2.4. The simplified thermal model here explained was used to electrically model the hTron, as described in section 3.4.

Chapter 3

Modeling

The results of simulations shown in this work were obtained through the software LTspice. For each exploited superconducting device, a SPICE model was created. This chapter describes the structure and behavior of all the models used to simulate the circuits of JJ neurons, NW neurons and the innovative synapses that will be introduced in the next chapter: (1) Josephson junctions; (2) Superconducting nanowires; (3) nTrons; (4) hTrons.

3.1 Josephson junction

In order to realize a model of the JJ neuron and perform the analysis of section 6, it is necessary to model the Josepjson junctions. The SPICE component of the Josephson junction was created using an enhanced version of a voltage-based model based on the resistively shunted model (RCSJ) explained in section 1.2.1. The structure, shown in figure 3.1, was inspired by [38].

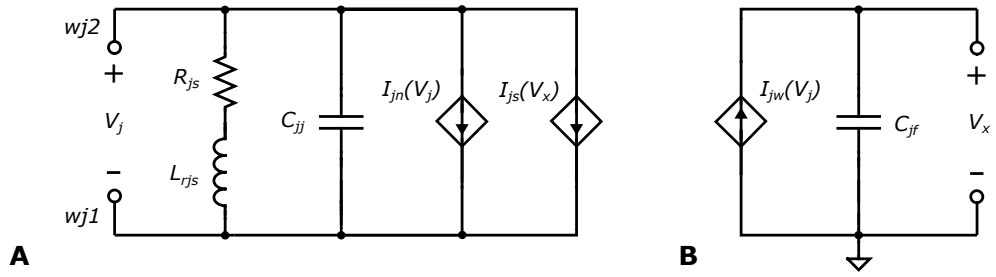

Figure 3.1: Circuit schematic of the SPICE model of a shunted Josephson junction. **(A)** Main circuit. **(B)** Subcircuit used to compute the phase of the junction integrating the voltage V_j.

The circuit can be divided in two subcircuits: the main circuit (A) is composed

by all the elements of the RCSJ model, while the subcircuit B is used to compute the phase difference φ of the junction. The two terminal of the junction are $wj1$ and $wj2$ and the voltage across it is V_j. $I_{js}(V_x)$ is the superconducting current, which is controlled by V_x:

$$I_{js}(V_x) = I_c \sin\varphi = I_c \sin\left(1000 \cdot V_x\right) \tag{3.1}$$

V_x is obtained by the integration of the the current $I_{jw}(V_j)$ in the subcircuit B. Thanks to the equation 1.2, we know that:

$$\varphi(t) = \frac{2\pi}{\Phi_0} \int V_j(t)dt \tag{3.2}$$

so setting $I_{jw}(V_j) = V_j$ and $C_{jf} = 1000 \cdot \Phi_0/2\pi$, the current provided by I_{jw} charges C_{jf} generating a voltage equal to:

$$V_x(t) = \frac{1}{C_{jf}} \int V_j(t)dt = \frac{\varphi(t)}{1000} \tag{3.3}$$

The factor 1000 was introduced just to obtain a reasonable value of capacitance C_{jf} in the order of picoseconds, so that LTspice can easily handle it. The model allows to set the initial value of φ, in case it would be necessary to do a simulation in which the superconducting current is not zero at $t = 0$. $I_{jn}(V_j)$ is the normal current of the junction and it is controlled by V_j:

$$I_{jn}(V_j) = \begin{cases} \frac{V_j}{R_{sg}}, & \text{if } |V_j| < V_g \\ \\ \frac{V_j}{R_n}, & \text{if } |V_j| > V_g \end{cases} \tag{3.4}$$

The values of R_{sg} and R_n depend on the structure and the fabrication process of the junction. Experimentally it was seen that I_cR_n and I_cR_{sg} are almost constant for a given process, so it is possible to compute the two resistance if I_c is known:

$$R_n(I_c) = \frac{IcRn}{I_c} \qquad\qquad R_{sg}(I_c) = \frac{V_m}{I_c} \tag{3.5}$$

In this model $IcRn = 1.65\ mV$ and $V_m = 16.5\ mV$. The values are obtained from experimental data of the MIT Lincoln Laboratory SFQ5EE process [39], as done in the software WRspice (the most used for simulations of Josephson junctions) [40].

R_{js} is the shunt resistor of the RCSJ model that typically has a parasitic series inductance L_{rjs}, whose value can be estimated with the tool InductEx, if the geometry of the device is known [38]. A too high inductance could alter the high-frequency dynamic of the junction, if it is not correctly designed [41]. The

parasitic inductance was introduced in the model, but it was neglected in all the simulations for sake of simplicity.

The model allows to set the following parameters: C_{jj}, L_{rjs}, I_c, $\varphi(t=0)$ and R_{tot}. The latter is the total resistance resulting from the parallel between R_{js} and R_{sg}, so R_{js} is set automatically in order to obtain the selected R_{tot}. Normally $R_{js} \ll R_{sg}$, so $R_{tot} \simeq R_{js}$.

The I-V characteristic of an overdamped Josephson junction model ($\Gamma > 1$) is shown in figure 3.2. The insets display the spiking behavior at two different values of bias current. It is the ability of the junction to generate pulses like the ones shown here that allowed to design the JJ neuron. The parameters of [1] necessary to obtain a Class I JJ neuron were used for this simulation ($\Gamma = 1.5$).

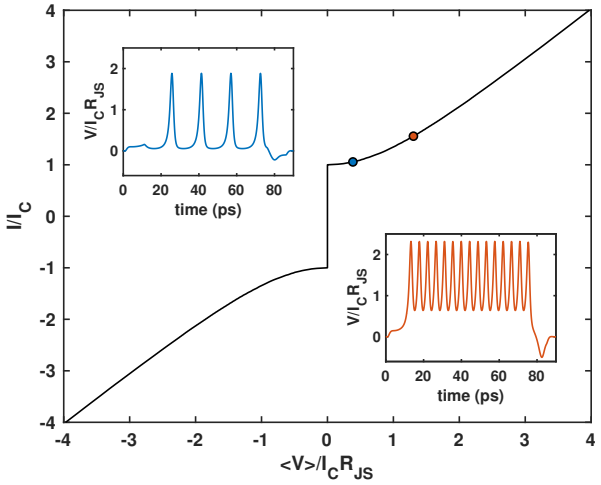

Figure 3.2: I-V characteristic with normalized axis of a current biased overdamped Josephson junction ($\Gamma = 1.5$) obtained in LTspice. For each point of the black curve the voltage is averaged over 300 ps. Parameters: $R_{tot} = 4\ \Omega$, $L_{rjs} = 0$ H, $C_{jj} = 102$ fF, $I_C = 90\ \mu A$. The blue curve in the inset is the normalized voltage in time-domain with a short pulse (65 ps) of 95 μA as bias current. The red curve with a pulse of 140 μA.

The model was realized to match the behavior of the standard model for Josephson junctions of WRspice. Figure 3.3 shows the I-V curves of a current biased underdamped Josephson simulated with LTspice and WRspice. The curve presents hysteresis because the junction is underdamped ($\Gamma = 0.54$), and at $< V > /I_c R_{tot} \approx 2.5$ the current drastically increases because the superconductivity is broken and R_n is lower than R_{tot}. The two curves matches almost perfectly. There are only small differences in the hysteresis and during the breaking of the

superconductivity, but considering that in JJ neurons the junctions do not work in these regimes the model can be trusted.

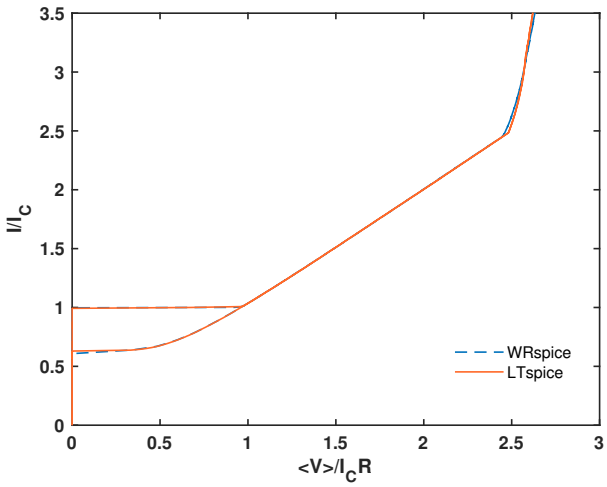

Figure 3.3: I-V characteristic with normalized axis of a current biased underdamped Josephson junction with hysteresis, obtained with WRspice (blue) and LTspice (orange). For each point of the curves the voltage is averaged over 300 ps. Parameters: $R_{tot} = 5\ \Omega$, $L_{rjs} = 0$ H, $C_{jj} = 230$ fF, $I_C = 200\ \mu A$ ($\Gamma = 0.54$).

3.2 Superconducting nanowire

It is important to model the superconducting nanowire, in order to perform electrical simulations of the NW neurons described in chapter 4 and 6. It was modeled through the dynamic SPICE model introduced by [23], which is based on the simplified 0D model explained in section 1.2.2. It was already shown that it can reproduce experimental results of SNSPD. Here It was slightly modified in order to adapt it to simulations of shunted nanowires, maintaining the consistency with experimental works. The circuit schematic of the modified model is shown in figure 3.4. It is composed by three distinct subcircuits: The main circuit (A) describes the nanowire as a kinetic inductance $L_k(I_d, V_{hs})$ in series with a variable normal resistance, which is modeled by the controlled voltage source $V_{res}(V_{hs}, I_{res})$ and the switch $S_{restore,1}(V_n)$.

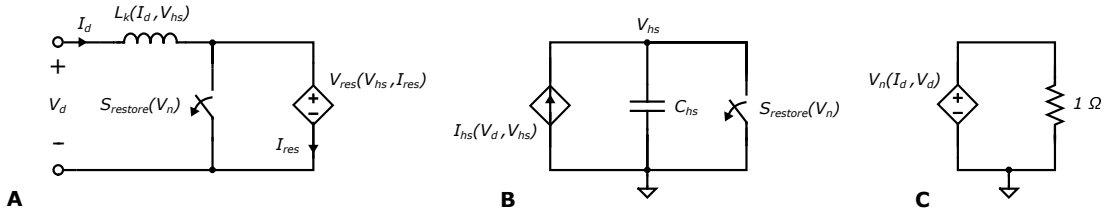

Figure 3.4: Circuit schematic of the SPICE model of a superconducting nanowire. **(A)** Main circuit with kinetic inductance and variable resistance. **(B)** Integration subcircuit circuit used to simulate the hotspot growth. **(C)** Subcircuit that stores the phase of the nanowire.

The state of the nanowire is stored in the subcircuit C: if $V_n(I_d, V_d) = 0\ V$ it is superconductive; if the current through $L_k(I_d, V_{hs})$ exceeds I_{sw}, it becomes normal ($V_n(I_d, V_d) = 1\ V$). The normal state is maintained until the voltage across the nanowire drops below a threshold. When the nanowire is superconductive ($V_n(I_d, V_d) = 0\ V$), $S_{restore,1}(V_n)$ is closed so that $V_d = 0$, otherwise it is open.

$V_{res}(V_{hs}, I_{res})$ represents the voltage across the hotspot resistance $R_{hs}(t)$, and it is controlled by V_{hs}, the voltage on the subcircuit B, which is equal to $R_{hs}(t)$:

$$V_{res}(V_{hs}, I_{res}) = \frac{|V_{hs}| + V_{hs}}{2} I_{res} \tag{3.6}$$

The expression is not simply $V_{hs}I_{res}$, in order to avoid a negative value of resistance. The subcircuit B is an integration circuit that models the hotpot growth, similar to the one exploited for the Josephson junction: the controlled current generator $I_{hs}(V_n, I_d, V_{hs})$ charges $C_{hs} = w/(2v_0R_\square)$ with a current proportional to the propagation velocity of the hotspot dl_{hs}/dt along the nanowire, when it is in the normal

phase. The resulting expression for V_{hs} in time is the following:

$$V_{hs}(t) = R_{\square}\frac{l_{hs}(t)}{w} = R_{\square}\frac{v_0}{w}\int \frac{\Psi(I_d/I_{sw})^2 - 2}{\sqrt{\Psi(I_d/I_{sw})^2 - 1 + |\Psi(I_d/I_{sw})^2 - 1|/2 + \delta}}dt \quad (3.7)$$

where w is the width of the nanowire and δ is just a small constant introduced to avoid zero in the denominator. The integration is stopped by $I_{hs}(V_n, I_d, V_{hs})$, when the whole nanowire has become normal ($R_{hs} = R_{norm}$). $S_{restore,2}(V_n)$ restores the superconducting phase discharging the capacitor when $V_n(I_d, V_d) = 0$. Figure 3.5 shows how the subcircuit B models the hotspot growth.

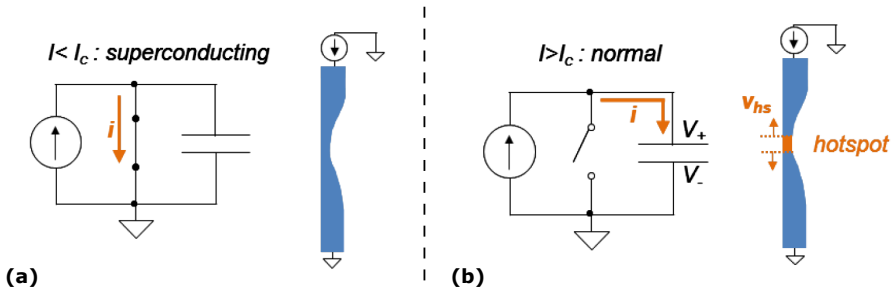

Figure 3.5: Subcircuit B used to simulate the hotspot growth. **(a)** When the nanowire is superconductive the switch is close so the hotspot resistance is zero.**(b)** When the nanowire is in normal state, the switch is open and the behavioral current source charges the capacitor, so that the resistance of the hotspot can grow. Modified Figure from [18].

The kinetic inductance $L_k(I_d, V_{hs})$ is computed with equation 1.2.2 in the superconducting phase. In the normal state its value is multiplied by the factor $1 - R_{hs}(t)/R_{norm}$ to consider that the normal region of the nanowire does not contribute to the kinetic inductance.

3.2.1 Shunted nanowire

In order to obtain a relaxation oscillator the nanowire is shunted with a resistor. As starting point for the simulations of the next sections, the model with shunt resistor was tested with the parameters used by [2] for NW neurons. Figure 3.6 shows that the I-V curve and the spiking behavior are similar to the ones of a Josephson junction in figure 3.2, but have fundamental differences: (1) for a shunted nanowire the spiking frequency is two orders of magnitudes lower; (2) the electrothermal oscillations are not generated by a coherent effect like in the case of the Josephson junction; and (3) the shunted nanowire presents the latching behavior when the

steady state current I_{ss} is reached, so that the nanowire remains constantly in the normal state without generating oscillations. The similarities between the two devices allowed to design the NW neuron with the same structure of the JJ neuron.

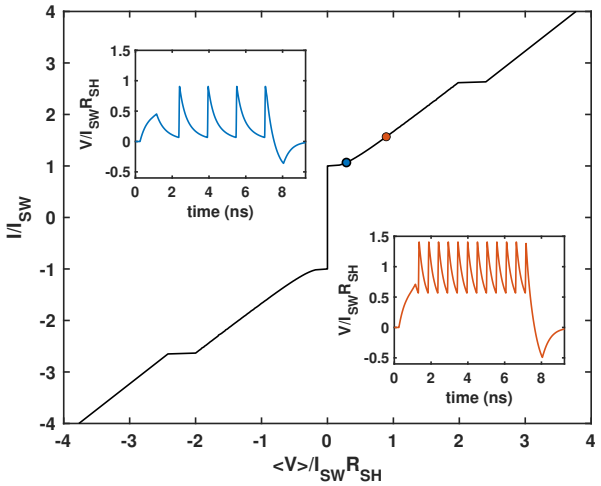

Figure 3.6: I-V characteristic with normalized axis of a current biased overdamped shunted nanowire obtained in LTspice. For each point of the black curve the voltage is averaged over 200 ns. When $|I| \approx 2.5 I_{SW}$, the nanowire latches in the normal state, so the voltage does not oscillates. Parameters: $R_{sh} = 10\ \Omega$, L = 4 nH, $I_{sw} = 30\ \mu A$. The blue curve in the inset is the normalized voltage in time-domain with a short pulse (6 ns) of 32 μA as bias current. The red curve with a pulse of 47 μA.

3.3 Nano-cryotron

It is necessary to model the nTron, in order to perform all the electrical simulations of chapter 4, and some of chapter 6. An electrical model of the nTron was already realized by Andrew Wagner (Raytheon BBN Technologies) using the verilog-A language. Here a SPICE model, which uses the same experimental parameters of the aforementioned model, is introduced. The behavior of the nTron is reproduced exploiting the previously explained model for superoconducting nanowires. The structure of the device can be divided in four elements: the gate nanowire, the drain nanowire, the source nanowire and the channel. Each element can be modeled as a nanowire, with its own width, length and critical current. With this simplification, it is possible to use the same circuital elements of section 3.2 with some changes applied, to form the structure of the nTron, shown in figure 3.7.

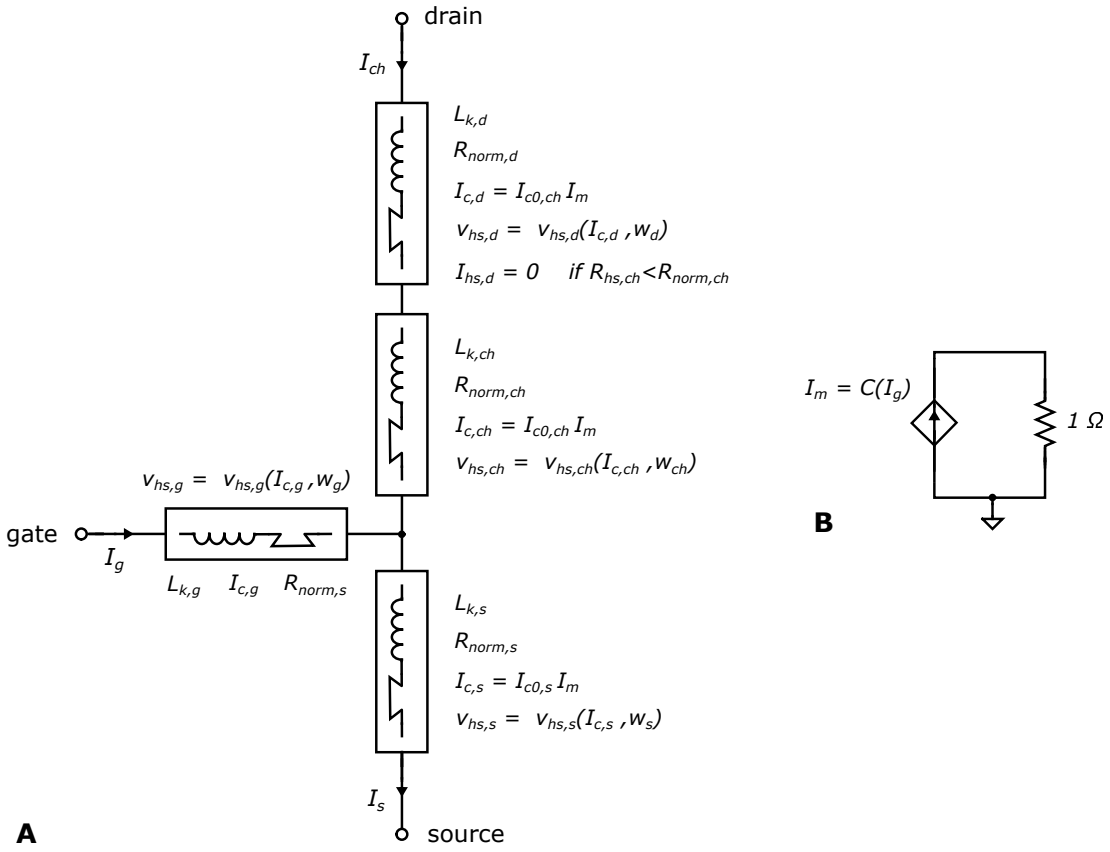

Figure 3.7: Circuit schematic of the SPICE model of the nTron **(A)** Main circuit composed by four superconducting nanowires. The image shows the important parameters and their dependencies for each nanowire element. **(b)** Subcircuit used to store the value of the coefficient that modulates the channel critical current.

The model allows to set the length and width of all the elements. The width of the entire gate nanowire w_g is approximated to the width of the choke. In order to reproduce the geometry in figure 2.1A, the channel is narrower than the drain and the source nanowires, and the gate is connected to the channel and the source nanowires. The critical current of the entire vertical nanowire (drain, channel, source) is set to $I_{c,ch}$, so that the growth of the hotspot can be simulated in the whole nanowire, as soon as it is formed. For the computation of v_{hs} each part of the vertical nanowire uses the value of its own critical current which can be different by $I_{c,ch}$.

$I_{c,ch}$ must be modified when the hotspot is generated. DC measurements on samples fabricated by Raytheon BBN Technologies showed that when the choke is normal the critical current of the channel is modulated by the gate current as follows:

$$I_{c,ch} = I_{c0,ch} C(I_g, I_g > I_{c,g}) = A_1 e^{-\frac{|I_g| - I_{c,g}}{\beta}} \tag{3.8}$$

where $I_{c0,ch}$ is the critical current without hotspot, $A_1 = 0.4$ and $\beta = 12.82\ \mu A^{-1}$. The subcircuit B in figure 3.7 is used to store the value of the coefficient $C(I_g)$, which is equal to 1 when $I_g < I_{c,g}$, and follows the expression 3.3 when $I_g > I_{c,g}$. This factor is inserted in the elements of the drain, channel and source. After the hotspot is formed, it initially expands only through the channel and the source. The integration subcircuit of the drain was modified to ensure that only if the channel becomes entirely normal ($R_{ch} = R_{norm,ch}$), it starts integrating to make the hotpsot propagate through the drain nanowire. The expansion velocity of the hotspot v_{hs} for a single element of the nTron is half of the one used in a single nanowire, because it expands only in one direction.

The model here introduced was tested in LTspice with different circuits configurations to demonstrate that it is able to amplify pulses or digital signals, and that its behavior is consistent with the one of the verilog-A model (simulated in WRspice). After checking the consistency, the model was also tested as amplifier of current pulses, setting the same parameters of the device fabricated by [3]. Figure 3.8a shows the simulated circuit schematic, while 3.8b shows the time-domain behavior. The magnitude of the output pulses depends only on $I_{b,nT}$, that can also be set to negative values. The signal can be amplified by a factor of 18. Changing the geometry of the nTron it is possible to have a different critical current for both the gate and the channel, obtaining different input threshold and maximum amplification.

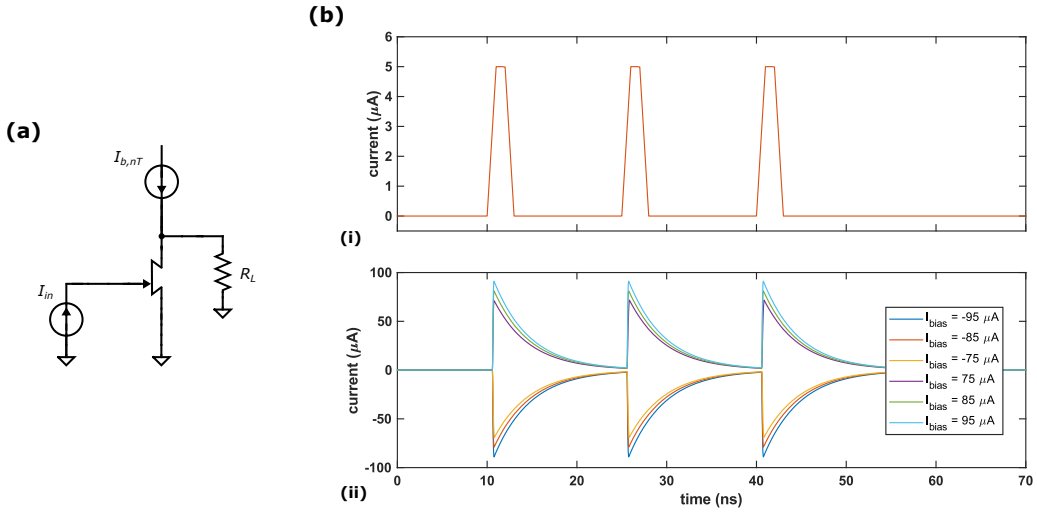

Figure 3.8: **(a)** Circuit schematic of the nTron used as amplifier of current pulses. **(b)** behavior of the circuit in time-domain. Parameters: $I_{c0,ch} = 106\ \mu A$, $I_{c,g} = 2.9\ \mu A$, $A_1 = 0.7$, $L_{k,ch} = 14\ nH$, $R_L = 5\ \Omega$. The panel **(i)** shows the current through the gate; the panel **(ii)** shows the current through R_L.

The behavior of the nTron is also defined by its load, which can be seen as shunt resistor of the channel. As explained in section 1.2.2, if the load impedance is too high the nanowire latches, and it is necessary to reset the bias current after each pulse to restore the superconductivity. This is an undesired behavior if there is not a clock signal in the system, so the load impedance must be kept low. The latching depends also on the bias current: a too high channel critical current, so consequently high bias current, would cause the nanowire to latch with lower load impedance, so there is a limit in the gain we can obtain with a certain load.

3.4 Heater nano-cryotron

The M-hTron introduced in section 2.2, was electrically modeled in LTspice, exploiting the 0D simplified model to describe the evolution in time of the heat transfer between the heater the insulator, the channel and the substrate outside the steady state condition. As consequence its validity in theory could be demonstrated only in the limit of the assumption made by the simplified thermal model. The channel was modeled as a superconducting nanowire with the method explained in section 3.2. Figure 3.9 shows the circuit schematic of the SPICE element, formed by three subcircuits.

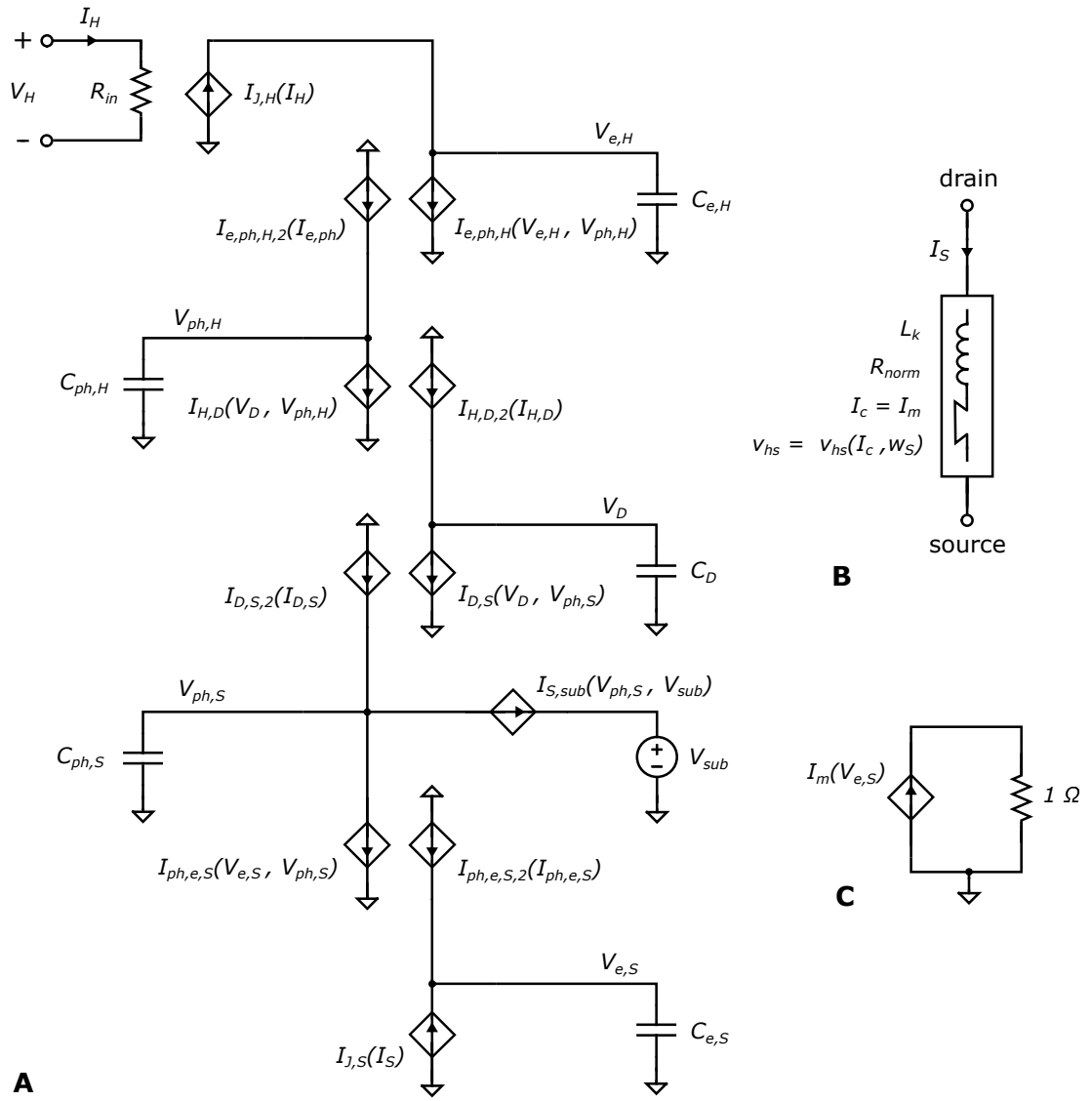

Figure 3.9: Circuit schematic of the M-hTron model. **(A)** Subcircuit that contains the heater and computes the electron temperature of the superconductor as function of the current through the heater I_H. **(B)** Nanowire model that reproduce the behavior of the channel. The critical current is modulated by the subcircuit B. **(C)** Subcircuit that stores the value of critical current as function of the electron temperature ($V_{e,S}$).

The subcircuit A, used to compute the electron temperature of the superconductor knowing I_H, is in turn composed by 5 circuits. Each of them associated with one of the differential equations in the simplified model. The subcircuit B is the nanowire model of the channel, whose critical current value is stored in

the subcircuit C. The latter obtains I_c using equation 2.4, which depends on $T_{e,S}$, computed by the subcircuit A.

In the subcircuit A, the heater is represented by the normal resistance R_{in}, on which the voltage drop is V_H. In order to compose the remaining part of the subcircuit, the equations of the simplified thermal model were rearranged in order to be correctly solved by LTspice. For example, the equation describing the electron temperature of the heater was rewritten as:

$$\gamma_H \frac{\partial T_{e,H}}{\partial t} = -\frac{\Sigma_{e-ph,H}}{T_{e,H}}(T_{e,H}^5 - T_{ph,H}^5) + \frac{I_H^2 \rho_H}{T_{e,H}(w_H d_H)^2} \tag{3.9}$$

If the temperature is associated to a voltage $V_{e,H}$, γ_H to a capacitor $C_{e,H}$ and the other terms to behavioral current sources, the equation is converted in a circuit obtaining the following current equation:

$$C_{e,H} \frac{\partial V_{e,H}}{\partial t} = -I_{e,ph,H}(V_{e,H}, V_{ph,H}) + I_{J,H}(I_H) \tag{3.10}$$

where $I_{e,ph,H}$ is associated with the flow of heat from the electron system to the phonon system, and $I_{J,H}$ with the Joule heat introduced by I_H. The same conversion method can be used for the phonon heat equation, in which the electron-phonon term is also present and can be rewritten as $I_{e,ph,H,2} = I_{e,ph,H}(T_{e,H}/T_{ph,H})$. After all the equations are converted with the same method, the subcircuit A is obtained. The current flowing from the top of it to the bottom represents the heat that flows from the heater to the channel. Since the substrate is considered at constant temperature, it is modeled by a simple voltage source (V_{sub}).
In the electron equation of the superconductor it was added a term $I_{J,S}(I_S)$ to consider the Joule heat generated by the current I_S through the channel just after the normal switching. It is zero if $I_S < Ic$ and $I_S^2 \rho_S / T_{e,S}(w_S d_S)^2$ if $I_S > Ic$. For sake of simplicity the electron heat capacity of the superconductor was modeled as linear dependent on the temperature. The value of γ_S was chosen to obtain the same results of [4]. All the other parameters was taken by [4].
The described circuit is able to correctly reproduce the behavior of the simplified electrothermal model. Figure 3.10 shows the time-domain evolution of the temperatures, which results to be consistent with figure 2.7a, obtained with the simplified model.

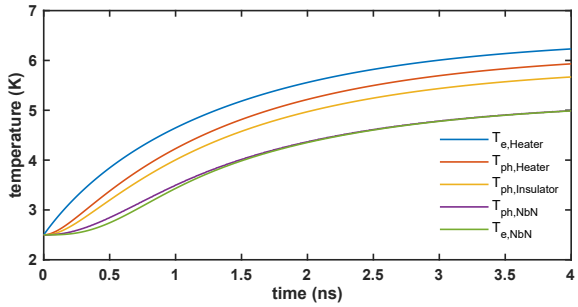

Figure 3.10: Simulated (with the SPICE model) time-domain evolution of the temperatures obtained with 500-nm wide heater and 600-nm wide channel and $I_H = 30\ \mu A$ after 1 ns.

The circuit simulated with the complete model, shown in figure 2.6a, was tested also with the SPICE model obtaining the following result

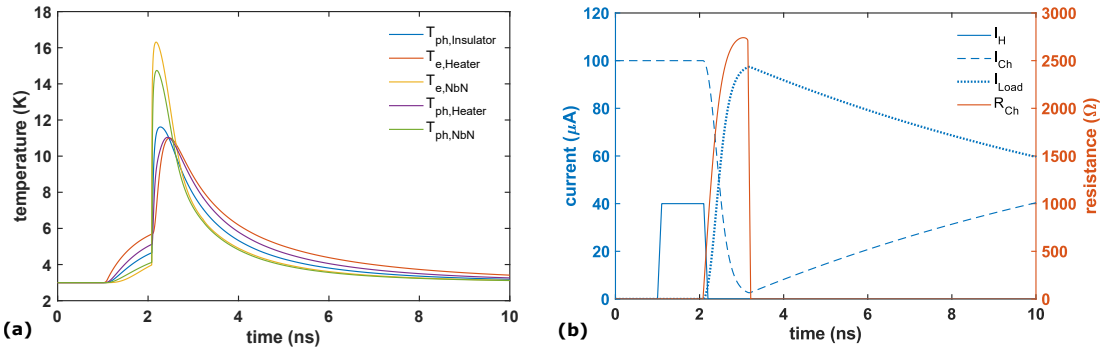

Figure 3.11: Time-domain behavior of the circuit shown in figure 2.6a, simulated using the SPICE model. A 1ns-long pulse of $I_H = 40\ \mu A$ is provided to the hTron. $I_{Ch} = 100\ \mu A$, $T_{sub} = 3\ K$, $R_L = 37\ \Omega$. **(a)** Time-domain evolution of all the temperatures considered in the model **(b)** electrical response of the hTron.

The general behavior is consistent with the one of the 3D model, but some differences are present. First of all the load impedance was set to 37 Ω instead of 50 Ω otherwise the nanowire would have latched. Since the latching phenomenon depends on the capability of the nanowire to dissipate heat after the hotspot formation, its thermal parameters used in the model need to be changed. In fact the Stekly parameter, ratio between the Joule heat and the dissipated heat, should be modified to consider that the heat can be transferred both to the substrate and the dielectric. Applying this modification the cooling down of the hotspot probably would be faster and the latching would appear at higher load resistances. Moreover some other parameters, like thermal conductivity of the superconductor,

are different from the electrothermal to the nanowire models. Here they were not modified in order to keep the behavior of the two 0D models coherent with the respective experimental data. An additional secondary improvement would be to include the state-dependent heat capacity of the superconductor.
Even if all the listed improvements were applied, probably the result would still differ from the one of the 3D model, for the intrinsic simplification: the heat transfer is modeled only in a single point of the nanowire. However the created SPICE model can be useful to perform electrical simulations of innovative circuit structures and demonstrate their working principle, as done in section 4.3.

Chapter 4

Nanocryotron-based synapse: Design and electrical analysis

The inductive synapse can reproduce some characteristics of its biological counterpart, but it still needs to be improved. The main problems of the inductive synapse for the nanowire-based system are:

- It is a passive component like a biological electrical synapse, so the input action potential can not be amplified as in the case of chemical synapses. Moreover the output current depends on the structure of the network, and can be very low. All previous simulations of the inductive synapse with coupled neurons are performed introducing a current source in front of the target neuron, because it can not provide enough current to activate it.

- It can reproduce the inhibitory behavior only if the main neuron is biased negatively. In biological systems the control depends only on the synapse.

- The tunability of the output current is low if compared with some of other synapses introduced in section 1.3.4. The ratio between the maximum and minimum current it can provide with different bias currents is $I_{out,max}/I_{out,min} = 1.20$. In the synapse based on nanotextured Josephson junctions the typical ratio is $I_{out,max}/I_{out,min} = 7$ [32].

The simplest solution to introduce the inhibitory control that depends only on the synapse is using the inductive coupling as done by [7].

4.1 Inductive synapse with inductive coupling

Introducing a transformer into the structure of the inductive synapse allows to choose between excitatory and inhibitory control, just changing the type of coupling from positive to negative. Figure 4.1 shows the circuit schematic of a neuron connected to an inductive synapse modified to implement the inductive coupling.

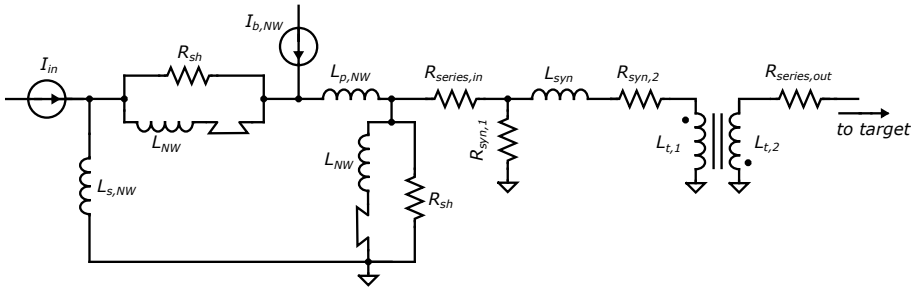

Figure 4.1: Circuit schematic of a NW neuron connected to a synapse with inductive coupling.

The characteristic currents equation of a transformer composed by two coupled inductors $L_{t,1}$ and $L_{t,2}$ is:

$$I_{t,2} = \frac{j\omega M}{j\omega L_{t,2} + Z_L} I_{t,1} = \frac{j\omega k\sqrt{L_{t,1}L_{t,2}}}{j\omega L_{t,2} + Z_L} I_{t,1} \tag{4.1}$$

where $I_{t,1}$ and $I_{t,2}$ are respectively the input and output current of the transformer, Z_L is the load impedance of the transformer, $M = k\sqrt{L_{t,1}L_{t,2}}$ is the mutual inductance and k is the coupling coefficient. If the latter is positive the coupling is positive, otherwise it is negative. If $L_{t,1} = L_{t,2}$ and $j\omega L_{t,2} \gg Z_L$, the output current of the transformer can be approximated to $I_{t,2} = kI_{t,1}$. In the ideal case k can be 1 or -1, so that the input current can be completely transferred to the output without loss. If $k = -1$ the control of the synapse is inhibitory, because the current is subtracted by the loop of the target neuron.

In the circuit shown in figure 4.1, Z_L is equal to $R_{series,out} + Z_{in,target}$, the input impedance of the target neuron. Considering that $Z_{in,target}$ is on the order of $j\omega L_{s,NW}/2 \approx 5\ nH$, $L_{t,1}$ and $L_{t,2}$ were set to 300 nH to approximately fulfill the previous conditions, having $|I_{t,1}| \approx |I_{t,2}|$. $R_{series,out}$ was set to 8 Ω to ensure isolation from the target neuron, even if it slightly lowers the output current of the synapse. Figure 4.2 shows the dynamic in time-domain of two NW neurons coupled through the modified inductive synapse with negative coupling. The same parameters of the neurons in simulations performed by [2] were used.

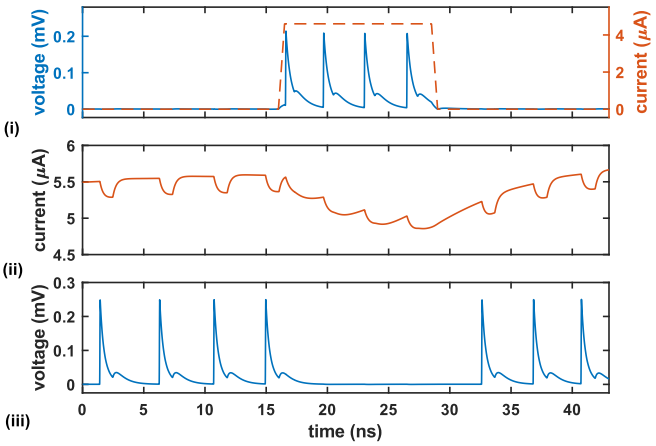

Figure 4.2: Time-domain behavior of two NW neurons coupled through inductive synapse with inductive coupling (inhibitory control). Parameters: $I_{b,NW,main} = 59\ \mu A$, $R_{series,in} = 5\ \Omega$, $R_{syn,1} = 40\ \Omega$, $L_{syn} = 265\ nH$, $R_{syn,2} = 20\ \Omega$, $I_{b,NW,target} = 57.17\ \mu A$, $I_{in,main} = 4.6\ \mu A$, $L_{t,1} = L_{t,2} = 300\ nH$, $R_{series,out} = 8\ \Omega$. The target neuron is activated with a current source connected to its input: $I_{in,target} = 5.6\ \mu A$. Panel **(i)** displays the output voltage and the input current of the main neuron; Panel **(ii)** displays the input current of the target neuron (sum of the current through $R_{series,out}$ and $I_{in,target}$); Panel **(iii)** displays the output voltage of the target neuron.

The inductive coupling solves the problem of the inhibition, but the synapse is still not versatile because the type of control can not be changed during the normal operations of the synapse, but only during the fabrication. Moreover the kinetic inductance can not be exploited to create the coupled inductors, since it does not depend on magnetic phenomena, so different materials and structure should be used. This means that integrating a transformer in a nanowire-based system would not be efficient in terms of fabrication processes and occupied area.

4.2 nTron synapse

The use of the inductive coupling was discarded, and the nano-cryotron (nTron) was added to the structure of the synapse in order to made the inductive synapse more similar to a biological chemical synapse and solve all the other problems listed above. The nTron is able to sense and amplify current pulses on the order of few μA, thanks to the presence of the narrow choke in the gate. It was already demonstrated the possibility of coupling it with an SNSPD [3], so it is reasonable to think we can do the same with shunted nanowires. In this way the nTron would become the perfect candidate for sensing action potentials coming from neurons, and modulating their amplification to introduce variability in the synapse. As shown in section 2.1, the SPICE model can reproduce the ability of the nTron to amplify pulses with different strengths, so it was used for all the following simulations performed to design the nTron synapse.

In the new structure proposed to improve the inductive synapse, the nTron gate is directly connected to the output of the main neuron, and the drain to a standard inductive synapse, that acts as load, as shown in figure 4.3. The nTron is designed in order to allow a single pulse of the upstream neuron to exceed the critical current of the gate, and generate a spike. The spikes charge L_{syn}, which slowly discharges through $R_{series,out}$, making the target neuron fire, just like in the simple inductive synapse.

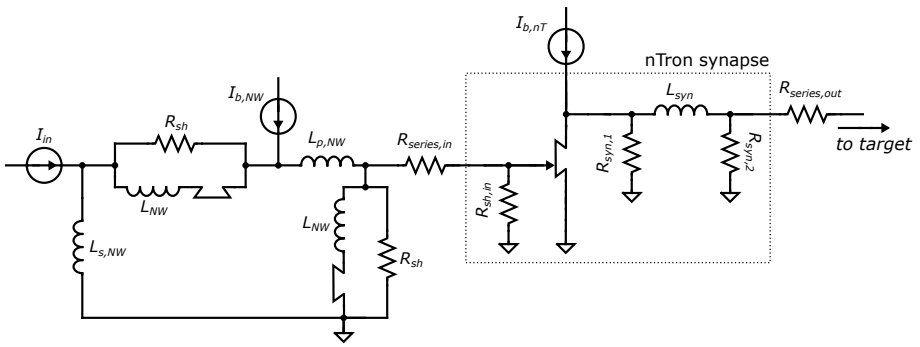

Figure 4.3: Circuit schematic of a NW neuron connected to a nTron synapse. The output voltage of the neuron make the nTron switch. The spikes generated by the nTron charge L_{syn}, that slowly discharges providing current to the target NW neuron.

In order to design the synapse optimising its characteristics some considerations were made:

- **Load impedance and latching:** as explained in section 2.1 the load of the nTron characterizes its behavior in the time-domain. The load impedance of the nTron is:

$$Z_L = \left(\frac{1}{R_{syn,1}} + \frac{1}{j\omega L_{syn} + (R_{syn,2} // Z_{in,target})} \right)^{-1} \tag{4.2}$$

  In the inductive synapse designed by [2], $R_{syn,1}$ is 40 Ω and $L_{syn} = 265\ nH$. With the resulting load resistance the nTron fabricated by [3] would latch according to simulations. Therefore it was chosen to lower $R_{syn,1}$ and L_{syn} respectively to 15 Ω and 100 nH. Considering that the latching depends also on the bias current, $I_{c,ch}$ were reduced from 106 μA to 40 μA. $R_{syn,2}$ was risen to increase the output current of the synapse, and improve the isolation from a target neuron.

- **Falling edge:** the nTron must be able to follow a pulses train without loosing information in order to charge the L_{syn} with more spikes, so the constant time $\sim L_{ch}/Z_L$ that defines the falling edge of the pulses, has to be lower than the typical spiking period of a neuron, which as minimum can be approximately 3 ns. The value of kinetic inductance of the channel L_{ch} was chosen to have a constant time of $\approx 700\ ps$.

- **Shunt resistor for the gate:** when the gate switches, the input impedance rises drastically, so the gate current is diverted to R_{sh} of the main oscillator and the profile of the action potential is clearly modified. To avoid this behavior, the gate was shunted with a resistor $R_{sh,in}$ which is on the order of few Ω, so that the change of input impedance becomes negligible. Figure 4.4 shows how the action potential and the input current of the nTron behave in the time-domain, when the gate switches, with and without $R_{sh,in}$.

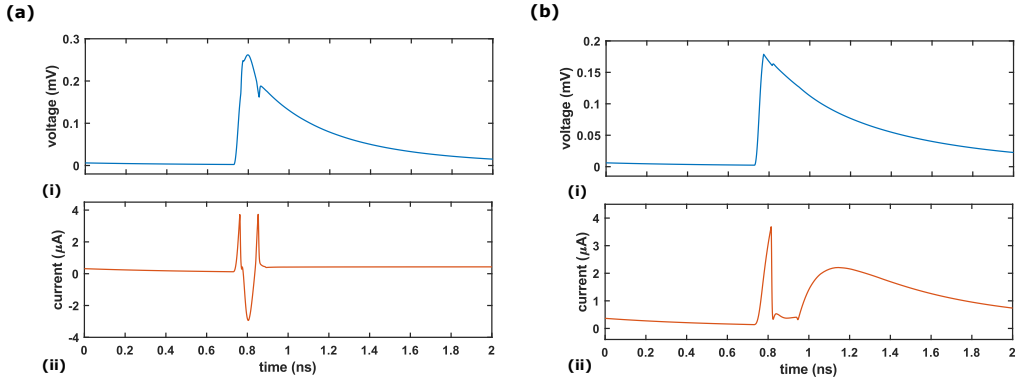

Figure 4.4: Time-domain behavior of the gate current in a nTron synapse that back-propagates toward the main neuron, and its effect on the action potential. Parameters: $I_{c,g} = 3.8\ \mu A$, $L_{k,g} = L_{k,s} = 260\ pH$. **(a)** Without shunt resistor $R_{sh,in} = 5\ \Omega$ on the gate of the nTron. **(b)** With shunt resistor $R_{sh,in}$ on the gate of the nTron.. For both the images: panel (i) displays the action potential of the NW neuron; panel (ii) displays the current through the gate of the nTron.

With the shunt resistor the gate current does not becomes negative after the switch, so the back-propagation is limited. The input impedance of the gate is $Z_{in} = j\omega(L_{k,g} + L_{k,s})$. If $R_{sh,in} < Z_{in}$ it could happen that the gate current is not able to make switch the gate. $R_{sh,in}$ has also to be lower than R_{sh} of the main oscillator. Therefore it is chosen in the range $Z_{in} < R_{sh,in} < R$. For certain values of Z_{in} and $R_{sh,in}$ it might be possible to generate relaxation oscillations in the gate nanowire, that could slightly alters the response of the nTron, so it is preferable to avoid this phenomenon.

- **Current stored in the inductor:** the synaptic inductance L_{syn} stores the pulses generated by the nTron, similarly to the superconducting loop of the synapse of [7] introduced in section 1.3.4 that is charged by a Josephson junction. In both cases the loop has a maximum value of current that can be stored. For the nTron synapse once a new pulse charge the inductor, the discharging time slightly decreases until it becomes equal to the charging time and a steady state is reached, so that the current saturates to its maximum value. Therefore the maximum current is higher when the period of the input pulses is much lower than the discharge time constant of L_{syn}, which approximately is $\tau_{syn} \approx L_{syn}/(R_{series,out} + Z_{in,target})$. The number of intermediate values (or number of pulses needed to reach the maximum current) depends on τ_{syn}: higher is τ_{syn}, fewer values can be obtained. $R_{series,out}$ was set to obtain about 10 possible levels and also ensure enough isolation from the target neuron. It is important to notice that the time during which the

current remains stored after the main neuron stops firing, depends also on $R_{series,out} + Z_{in,target}$.

The main advantage of using the nTron is that the pulses are amplified and their magnitude depends only on the bias current. As consequence, even if the upstream neuron was driving more synapses, the output current of a single one would be the same. Moreover with a negative bias current the synapse has an inhibitory control on the downstream neuron. This solves the problem of using inductive coupling, making possible to change type of control even after the fabrication of the devices. Therefore a network based on nTron synapse would be much more versatile. Figure 4.5 shows how the synapse behaves in the time-domain with excitatory and inhibitory control.

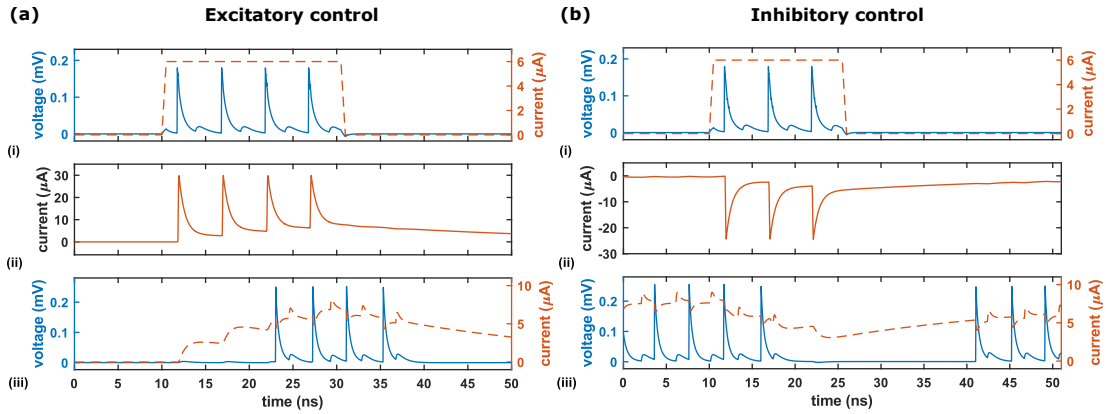

Figure 4.5: Time-domain behavior of two NW neurons coupled through an nTron synapse. Neurons parameters: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 10\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{bias} = 57\ \mu A$, $R_{series,in} = 20\ \Omega$, $R_{series,out} = 5\ \Omega$. nTron synapse parameters: $I_{c,g} = 3.7\ \mu A$, $I_{c,ch} = 40\ \mu A$, $R_{sh,in} = 5\ \Omega$, $R_{syn,1} = 15\ \Omega$, $L_{syn} = 100\ nH$, $R_{syn,2} = 50\ \Omega$. **(a)** Excitatory control. Parameters: $I_{b,nT} = 36\ \mu A$. **(b)** Inhibitory control. Parameters: $I_{b,nT} = -30\ \mu A$, the target neuron is activated with a current source on its input that continuously provides 8 μA. For both cases: Panel **(i)** displays the input current and the output voltage of the upstream neuron; Panel **(ii)** displays the output current of the nTron, that enter into the inductive loop; Panel **(iii)** displays the input current and the output voltage of the downstream neuron.

4.2.1 Variable synaptic strength

A key characteristic of the nTron synapse is the possibility to tune the bias current $I_{b,nT}$ to obtain a variable synaptic strength, without modulating the kinetic inductance using an additional current source. The range of possible values that

can be set for the bias current depends on how the creation of the hotspot in the gate modulates the critical current of the channel. When the gate current reaches the threshold, the critical current of the channel becomes $I_{c,ch} = A_1 I_{c0,ch}$. Then most of the gate current is diverted into $R_{sh,in}$ because the gate becomes normal and the neuron is not an ideal current source, so $I_{c,ch}$ can not be further decreased below $A_1 I_{c0,ch}$. As consequence, it is necessary to set a bias current higher than $A_1 I_{c0,ch}$ to ensure the channel becomes normal after the creation of the hotspot. Moreover if $|I_{b,nT}| > 0.9 I_{c0,ch}$, an undesired behavior such as photon- and noise-induced hotspot generation was observed experimentally [3], so the bias current of the nTron has to be set such that:

$$A_1 I_{c0,ch} < |I_{b,nT}| < 0.9 I_{c0,ch} \tag{4.3}$$

Lower is A_1, larger is the range of operation. The value of A_1 depends on the geometry of the channel in the region where the hotspot is generated. It was observed that the hotspot is able to suppress J_c only within one diffusion length L_D of its perimeter [3]. For thin films of NbN the typical diffusion length is around 100 nm, so the width of the channel has to be on the same scale to generate a sharp dropoff in $I_{c,ch}$ and minimize A_1. By simulations and measurements of [3], it results that $A_1 \approx 0.7$, with a channel width of approximately 150 nm and a thickness of 10 nm, while by measurements performed to realize the verilog-A model it seems possible to obtain $A_1 = 0.4$ with a channel width of 108 nm and a thickness of 4 nm. For the nTron synapse it was chosen to keep almost the same parameters of the verilog-A model, to obtain $A_1 = 0.4$. Considering that A_1 depends on the channel width, it was maintained the value 108 nm, but the thickness of the film was risen to 8 nm to obtain $I_{c0,ch} = 40\mu A$.

Response to a single pulse

The response of the nTron synapse to a single action potential from the upstream neuron was studied, because hypothetically it could be possible to realize a neural network that exploits the magnitude of single pulses to encode information, without charging the synaptic inductance with pulses trains. It was already demonstrated by [29], that NW neurons with inductive synapses can form a neural network for image recognition, if they use this kind of encoding. Figure 4.6a shows how the profile of the output current through $R_{series,out}$ varies as function of the bias current of the nTron, when a single action potential activates it (the load of the synapse is a NW neuron with zero bias current). The nTron can not be biased very close to $A_1 I_{c0,ch}$, so it was chosen to set as minimum possible $I_{b,nT}$ a current 12.5% higher. Therefore we obtain a ratio between maximum and minimum output currents the

nTron can provide equal to:

$$\frac{I_{out,nT}^{max}}{I_{out,nT}^{min}} = \frac{0.9 I_{c0,ch}}{1.125\ A_1 I_{c0,ch}} = \frac{0.8}{A_1} \tag{4.4}$$

that is equal to 2 with $A_1 = 0.4$. Figure 4.6b shows how the peak current of a single pulse on the output varies as function of $I_{b,nT}$. The range of bias currents and associated maximum output current in which the nTron works properly is displayed by the orange squares. The peak current with $I_{b,nT}/I_{c0,ch} = \pm 0.9$ is $\pm 3.2\ \mu A$, while it is $\pm 1.45\ \mu A$ for $I_{b,nT}/I_{c0,ch} = \pm 0.45$, so the resulting tunability is $I_{out}^{max}/I_{out}^{min} = 2.2$ (120 %). This value is higher than the ratio computed above for the output current of the nTron, and it is 1.83 times larger than the tunability of an inductive synapse. Designing a nTron synapse with higher channel critical current would allow to make fire a target neuron with a single spike. This can not be done with inductive synapses, without increasing the critical current of the main neuron.

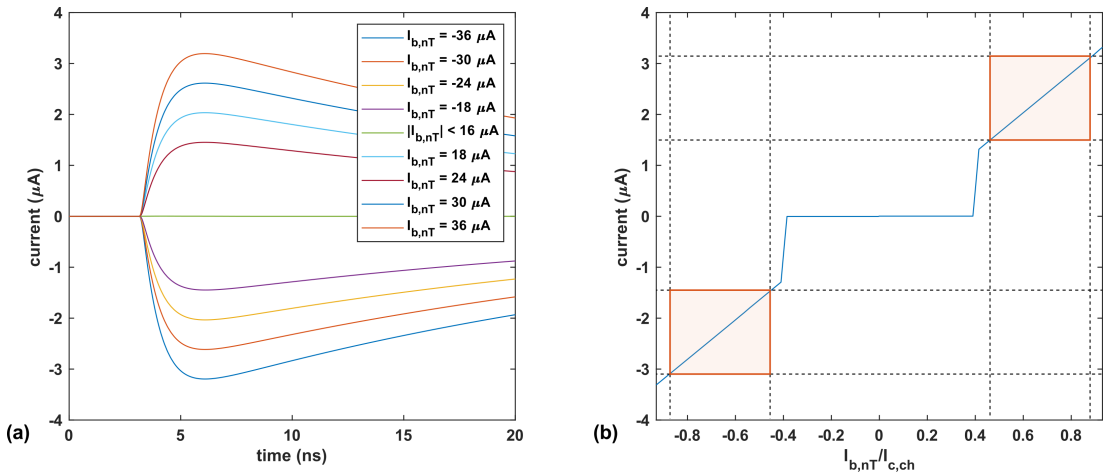

Figure 4.6: nTron synapse activated by a single action potential from a NW neuron. The load of the synapse is a NW neuron with bias current set to zero. Neurons parameters: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 10\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{bias} = 57\ \mu A$, $R_{series,in} = 20\ \Omega$, $R_{series,out} = 5\ \Omega$. nTron synapse parameters: $I_{c,g} = 3.7\ \mu A$, $I_{c,ch} = 40\ \mu A$, $R_{sh,in} = 3\ \Omega$, $R_{syn,1} = 15\ \Omega$, $L_{syn} = 100\ nH$, $R_{syn,2} = 50\ \Omega$. **(a)** Profile of the output current of the synapse (current through $R_{series,out}$) in the time domain, with different values of bias current $I_{b,nT}$. **(b)** Peak current of the output pulses generated by the nTron synapse versus the ratio between $I_{b,nT}$ and $I_{c0,ch}$. For each value of $I_{b,nT}/I_{c0,ch}$ the peak current is obtained measuring the maximum of the pulses that are like the ones shown in figure (a). The orange areas represent the intervals of current in which the nTron works properly with single spikes

Response to multiple pulses

The nTron synapse, as shwon in figure 4.5, can work with trains of pulses like the inductive synapse, making possible to provide higher current to the target neuron. As explained previously, with a fixed $I_{b,nT}$ the synaptic inductor can be charged at different levels of current controlling the number of incoming action potentials, and there is a maximum current the inductor can store (with a fixed spiking frequency of the main neuron). It was seen that when the current through L_{syn} saturates to its maximum value, and action potentials continue to activate the nTron, a bias current too close to $A_1 I_{c,ch}$ does not allow to sense every new incoming pulses. Therefore we set as minimum possible $I_{b,nT}$ a current 50% higher than $A_1 I_{c,ch}$ to ensure any information is lost. We obtain a ratio between maximum and minimum output currents the nTron can provide equal to:

$$\frac{I^{max}_{out,nT}}{I^{min}_{out,nT}} = \frac{0.9 I_{c0,ch}}{1.5\ A_1 I_{c0,ch}} = \frac{0.6}{A_1} \tag{4.5}$$

that is equal to 1.5 with $A_1 = 0.4$. Figure 4.7a and 4.7b show how the profile of the output current through $R_{series,out}$ varies as function of the bias current of the nTron, when a train of action potential activate it and make the inductor saturate (the load of the synapse is a NW neuron with zero bias current).

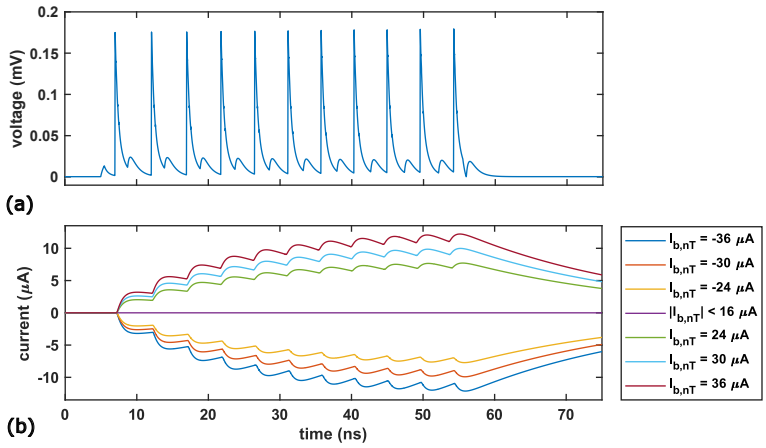

Figure 4.7: nTron synapse activated by a 11 consecutive action potentials from a NW neuron. The load of the synapse is a NW neuron with bias current set to zero. Neurons parameters: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 10\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{bias} = 57\ \mu A$, $R_{series,in} = 20\ \Omega$, $R_{series,out} = 5\ \Omega$. nTron synapse parameters: $I_{c,g} = 3.7\ \mu A$, $I_{c,ch} = 40\ \mu A$, $R_{sh,in} = 3\ \Omega$, $R_{syn,1} = 15\ \Omega$, $L_{syn} = 100\ nH$, $R_{syn,2} = 50\ \Omega$. **(a)** Output voltage of the main neuron in the time domain. **(b)** Profile of the output current of the synapse (current through $R_{series,out}$) in the time domain, with different values of bias current $I_{b,nT}$.

Figure 4.8 shows how the maximum output peak current between 0 and 70 ns (with 11 incoming pulses at a spiking frequency of 100 MHz) varies as function of $I_{b,nT}$. The orange squares display the ranges of bias currents and associated maximum output current in which the nTron works properly. It is possible to notice that for $0.4 < |I_{b,nT}/I_{c0,ch}| < 0.6$ the output current starts decreasing because some pulses are not amplified by the nTron.

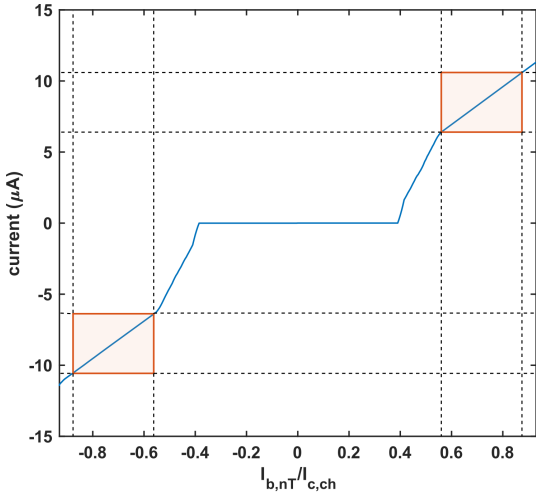

Figure 4.8: Peak current after 11 action potentials generated by the nTron have charged L_{syn}, versus the ratio between $I_{b,nT}$ and $I_{c0,ch}$. For each value of $I_{b,nT}/I_{c0,ch}$ the peak current is obtained measuring the maximum of the current profiles, that are like the ones shown in figure 4.7a. The orange areas represent the intervals of current in which the nTron works properly with multiple spikes. Neurons parameters of figure 4.7

The maximum peak current with $I_{b,nT}/I_{c0,ch} = \pm 0.9$ is $\pm 10.92\ \mu A$, while it is $\pm 6.94\ \mu A$ for $I_{b,nT}/I_{c0,ch} = \pm 0.6$, so the resulting tunability is $I_{out}^{max}/I_{out}^{min} = 1.57$ (57 %). Even in this case the value is higher than the ratio computed for the output current of the nTron. It is 1.3 times higher than the tunability of the inductive synapse, but much lower than the result obtained with single pulses. Considering that 120% of tunability can be obtained with a single spike, it might be possible to obtain the same result. Here the problem is not analyzed in details. Further analysis would be necessary to understand how to change the parameters of the nTron or the structure of the synapse, in order to reach the desired value. There could be also a problem in the model that might be solved with additional experimental data. Another way to increase the tunability would be to design a new geometry for the channel that can drastically decrease A_1.
The energy consumption of the nTron synapse, considering the variable strength, is in the range 5-12 aJ per spike. If $|I_{b,nT}/I_{c0,ch}| = 0.6$ it has the same energy

consumption of a inductive synapse, otherwise it is slightly higher, but still lower than the one of the NW neuron.

4.2.2 Fan-out/fan-in limitations

The goal of this section is to study the fan-out/fan-in limitations of the nTron synapse, to understand if it is possible to realize a large network, that might reach the level of connectivity of biological system, in which every neuron is connected to thousands of others. Here only the simplest complete network (shown in section 1.1.1) is chosen as starting point for the following analysis.

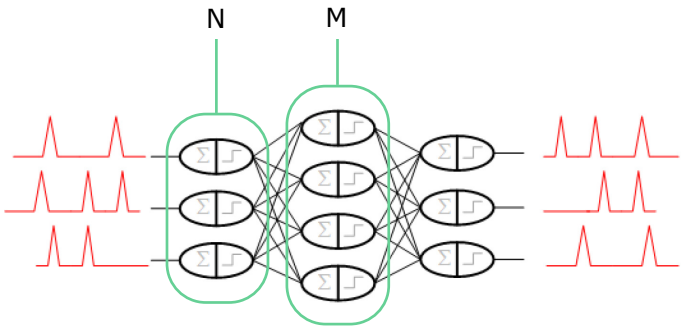

Figure 4.9: Simplified schematic of a SNN. The neurons are indicated as ovals and the connecting lines can be associated with the synapses. The neurons can be divided into two parts like in the integrate-and -fire model: the input half, which sums the inputs (integration), and the output half, which generates a spike if the threshold is exceeded (firing). Both the input and output signals of the network are patterns of spikes in time. N and M are respectively the numbers of neurons in the first and second layer. Modified figure from [11].

Let's consider just two layers of the entire simplified system, so that the first one is composed by N neurons and the second one by M neurons (see figure 4.9), in order to understand how the input current of a neuron in the second layer depends on the parameters of the first one. This is important because an higher current that can be possibly provided to a neuron in the second layer is a consequence of larger maximum achievable fan-out/fan-in. Each firing neuron of the first layer can be very roughly considered as a generator of current $I_{out,n1}$, with a certain output impedance $Z_{out,n1}$, that drives M synapses with input impedance $Z_{in,syn}$. If we assume for sake of simplicity that all the synapses have the same weight $\overline{W}$, average of all the real weights, the output current of the neurons in the first layer ($I_{out,n1}$)

equally splits between the synapses, and the output current of each synapse is:

$$I_{out,syn} \approx I_{out,n1}\overline{W}\frac{Z_{out,n1}}{MZ_{out,n1} + Z_{in,syn}} \tag{4.6}$$

Usually $Z_{in,syn}$ is lower than $Z_{out,n1}$, so if M is large enough, $MZ_{out,n1} \gg Z_{in,syn}$, and the current becomes:

$$I_{out,syn} \approx \frac{I_{out,n1}\overline{W}}{M} \tag{4.7}$$

Considering that there are N_{on} neurons in the first layer that are firing and supposing that the synapses can be seen as current generators with output current $I_{out,syn}$ and impedance $Z_{out,syn}$, the input current of each neuron in the second layer is:

$$I_{in,n2} \approx I_{out,n1}\overline{W}\left(\frac{N_{on}}{M}\right)\frac{1}{1 + N\left(\frac{Z_{in,n2}}{Z_{out,syn}}\right)} \tag{4.8}$$

If we have $N = N_{on} = 1$ we are in the simplest case for which the output current of each synapse is equal to $I_{in,n2}$ and its expression is the equation 4.7. This configuration is studied in section 4.2.2(Fan-out), to find the maximum fan-out (maximum number of neurons M_{on} that can be driven by a single neuron) for NW neurons using nTron synapses. The neurons in the second layer have a minimum input current $I^{min}_{in,n2}$ that makes them fire, so the maximum fan-out is:

$$FO_{max} = M_{on,max} \approx \overline{W}\left(\frac{I_{out,n1}}{I^{min}_{in,n2}}\right) \tag{4.9}$$

In the worst case of $N_{on} = M = 1$ and $N \gg 1$, we are in the simple situation for which all the neuron in the first layer fan into a neuron in the second one, and only one of them fires. This configuration is analyzed in section 4.2.2(Fan-in) for NW neurons with nTron synapses to find the maximum fan-in (maximum number of neurons N that allows to activate one neuron in the second layer with only one neuron firing in the first). From equation 4.7 it is possible to obtain the maximum fan-in:

$$FI_{max} = N_{max} \approx \left(\frac{Z_{out,syn}}{Z_{in,n2}}\right)\left(\frac{\overline{W}I_{out,n1}}{I^{min}_{in,n2}} - 1\right) \tag{4.10}$$

In the most general configuration with $M > 1$, $N > 1$, and $N_{on} > 1$, we can not define a maximum fan-in or fan-out, but a useful factor would be the maximum ratio between the number of activated neuron in the second layer M_{on} and the number of firing neuron N_{on} in the first one:

$$GFO_{max} = \left(\frac{M_{on}}{N_{on}}\right)_{max} \approx \overline{W}\left(\frac{I_{out,n1}}{I^{min}_{in,n2}}\right)\frac{1}{1 + N\left(\frac{Z_{in,n2}}{Z_{out,syn}}\right)} \tag{4.11}$$

This expression can be seen as a generalized definition of maximum fan-out ("generalized fan-out"), that intrinsically depends on the maximum fan-in, so higher it is, better are both the fan-in and fan-out characteristics. It is analyzed in section 6.3.3 for a hybrid neural network that integrate NW neurons with JJ neurons and uses inductive synapses.

Fan-out

In order to analyze the fan-out of a system composed by NW neurons and nTron synapses, we can consider the case in which the main neuron fires multiple times activating all the nTron synapses, that provide enough current to the target neurons to make them fire. For sake of simplicity all the weights of the synapses ($I_{b,nT}$) are set to the maximum value. The structure of the analyzed circuit is shown in figure 4.10.

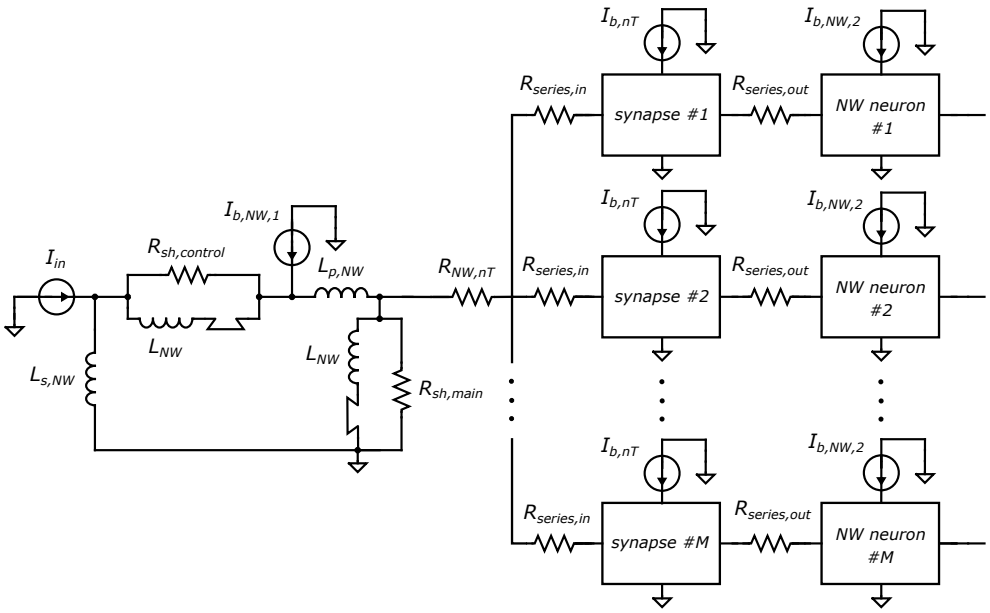

Figure 4.10: Schematic of the simulated circuit: the main neuron drives M NW neurons through nTron synapses

The expression of the fan-out is slightly different from equation 4.9, because here each synapse can certainly activates the target neuron, so we have to consider $I^{min}_{in,syn}$, the threshold current of the synapse, instead of $I^{min}_{in,n2}$, obtaining:

$$FO_{max} = \frac{I_{out,n1}}{I^{min}_{in,syn}} \tag{4.12}$$

When the main neuron fires, the bias current of the main oscillator is diverted from the nanowire to $L_{p,NW}$, $R_{sh,main}$ and the load impedance. Higher is the bias current, more current flows through the load, so $I_{out,n1}$ is proportional to the bias current, which in turn depends on the switching current $I_{sw,n1}$. Therefore, increasing $I_{sw,n1}$, it is possible to obtain an higher maximum fan-out. As already explained, $I_{sw,n1} = wdJ_c$, where d is the thickness of the film and w the width. Values of critical current higher than 100 μA have already been obtained by [3] with $d = 10\ nm$, and [4] with $d = 20\ nm$. Starting from the parameters of the fabricated NbN film, extracted experimentally in chapter 5, here it was supposed that it would be possible to reach 200 μA of switching current with a thickness of 16.8 nm, just increasing the width to 297.6 nm. This is only an assumption that could be not true, since there is not experimental evidence that it can be realized. In order to increase I_{sw} starting from the parameters obtained experimentally here and in [9], it is necessary to do some considerations:

- The value of shunt resistor R_{sh} for which the nanowires latch depends on I_{sw}: higher is the bias current (proportional to I_{sw}), slower is the timescale to reset the superconductivity after the switch, because more Joule heat has to be dissipated. Therefore lower is the electrical time constant $\tau_e = L_{nw}/R_{sh}$ for which latching occurs, as explained in section 1.2.2. This means that in order to increase $I_{sw,n1}$, τ_e has to be increased to avoid latching. Here it was chosen to maintain always the same R_{sh} and just rise L_{nw} proportionally to $I_{sw,n1}$, so that the output impedance of the neuron is not altered much.

- To ensure that the neuron fires also $L_{s,NW}$ and $L_{p,NW}$ have to be proportionally increased.

- The bias current $I_{b,NW}$ was set at $1.9I_{sw}$ to maintain enough margin from external noise that could make the neuron fire also when the input current is zero. Even if the nanowire can be activated with lower input current I_{in}, it was set to $0.2I_{sw}$ to ensure the activation.

The most relevant parameters used for the simulations are shown in table 4.1.

d	J_c	w	R_{sh} (Ω)	L_{nw}	$L_{s,NW}, L_{p,NW}$	$I_{b,NW}$	I_{in}
16.8 nm	40 $mA/\mu m^2$	$I_{sw}/(dJ_c)$	10	(130 $pH/\mu A$)I_{sw}	(325 $pH/\mu A$)I_{sw}	$1.9I_{sw}$	$0.2I_{sw}$

Table 4.1: Parameters used to ensure that the NW neuron does not latch during simulations. Some are functions of I_{sw}.

Increasing the time constant τ_e that defines the falling edge of an action potential, the spiking frequency decreases approximately like $1/L(I_{sw})$, following equation 1.10. Figure 4.11 shows the spiking frequency as function of $I_{sw,n1}$.

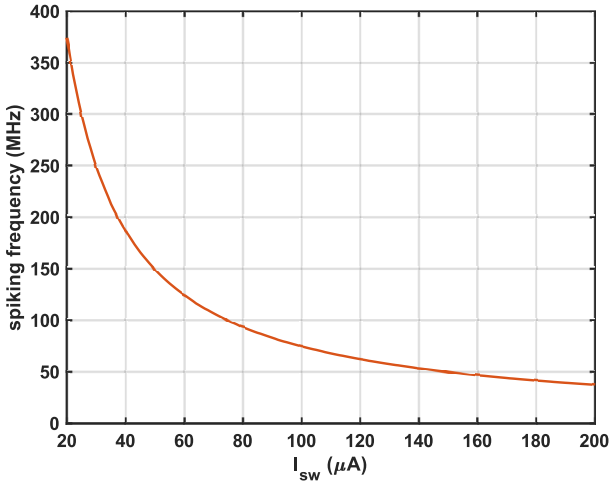

Figure 4.11: Spiking frequency of a NW neuron as function of the switching current, obtained through simulations with the parameters shown in table 4.1

Trying to increase the fan-out, it is necessary to consider that the spiking frequency is lowered, so the the nTron synapse will behave differently. As explained in section 4.2, the ratio between the spiking period and the discharging time of the synaptic inductor defines the maximum level of current that can be stored in the synapse. If we want to maintain approximately the same maximum current even at higher switching currents to ensure the target neurons fire, we have to increase the discharging time. This can be done decreasing $R_{series,out}$. In the following simulations, it was set to (100 $/I_{sw}$) Ω, so that it is 5 Ω when $I_{sw} = 20\ \mu A$ as in previous simulations, and 10 times smaller (0.5 Ω) when $I_{sw} = 200\ \mu A$, to counterbalance the decrease of frequency. An undesired consequence of lowering $R_{series,out}$, is that the isolation of the synapse from the downstream neuron is less efficient, so further work would be necessary to understand how to solve this issue. As explained above, the output current of the neuron depends on the current divider between $R_{sh,main}$, $L_{p,NW}$ and the load impedance of the neuron. Higher is $R_{sh,main}/Z_{load}$, more current flows to the load and higher is the maximum fan-out. As shown in figure 4.10, the load is $R_{NW,nT} + R_{series,in}/M$, which can be approximated to $R_{NW,nT}$ if M is large. Setting $R_{sh,main}$ to a value 5 times larger than $R_{NW,nT}$ and ensuring that $R_{sh,main}//R_{NW,nT} \approx R_{sh,control}$, it is possible to have almost 40% of $I_{b,NW}$ flowing to the load. Increasing $R_{sh,main}$ more, the behavior of the main oscillator would become too dependent on the load, and the shape of the action potential would be modified. For this reason $R_{sh,main}$ was set to 50 Ω and $R_{NW,nT}$ to 10 Ω. Figure 4.12 shows the simulated dynamic in time-domain of a NW neuron with $I_{sw,n1} = 200\ \mu A$ and the increase of $R_{sh,main}$, that drives 35 NW neurons with $I_{sw,n2} = 30\ \mu A$.

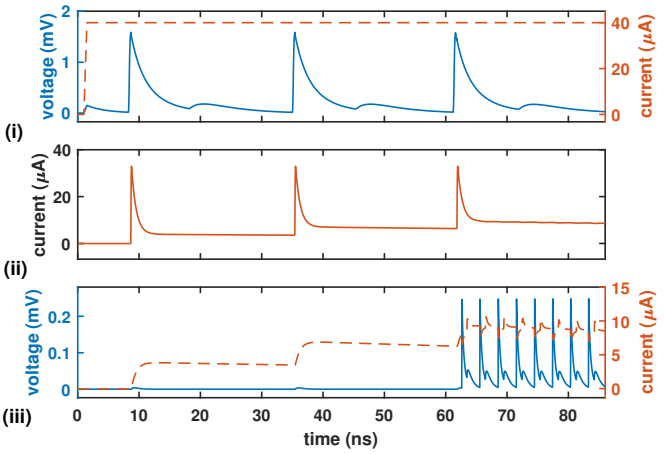

Figure 4.12: Time-domain simulation of a NW neuron that drives 35 NW neurons, through nTron synapses. The parameters of the neurons follow expressions shown in table 4.1. Main neuron parameters: $I_{b,n1} = 200\ \mu A$, $R_{sh,main} = 50\ \Omega$, $R_{NW,nT} = 10\ \Omega$. Synapses parameters: $R_{series,in} = 1\ \Omega$, $R_{sh,in} = 5\ \Omega$, $I_{c,g} = 3.7\ \mu A$, $I_{b,nT} = 36\ \mu A$, $R_{series,out} = 0.5\ \Omega$. Target neurons parameters: $I_{sw,n2} = 30\ \mu A$. Panel **(i)** displays the output voltage and the input current of the main neuron; Panel **(ii)** displays the output current of the nTron; Panel **(iii)** displays the input current and output voltage of one of the target neurons.

In order to perform a more complete analysis on the fan-out limitation and confirm if equation 4.12 is correct, it was created a Matlab script, integrated with LTspice, that can simulate a system with different number of target neurons and check if they fire. In this way it is possible to find the maximum fan-out for each possible configuration.

Figure 4.13 shows the maximum fan-out as function of $I_{sw,n1}$ (when the weights of the synapses are maximized), for different $I_{c,g}$, obtained with the aforementioned method. The maximum fan-out is obviously boosted by the amplification of the nTron synapse. Using standard inductive synapses, such numbers could not be reached. The image shows that FO_{max} depends on $I_{out,n1}/I^{min}_{in,nT}$, indeed the slope of the curve is higher with lower $I^{min}_{in,nT}$ as expected, but $I^{min}_{in,nT}$ is different from the critical current of the gate. For the case of $I_{c,g} = 3.7\ \mu A$, $I^{min}_{in,nT}$ is equal to $4.4\ \mu A$. This is due to the presence of $R_{sh,in}$ and $R_{series,in}$ in each synapse, that lower the current flowing into the gate.

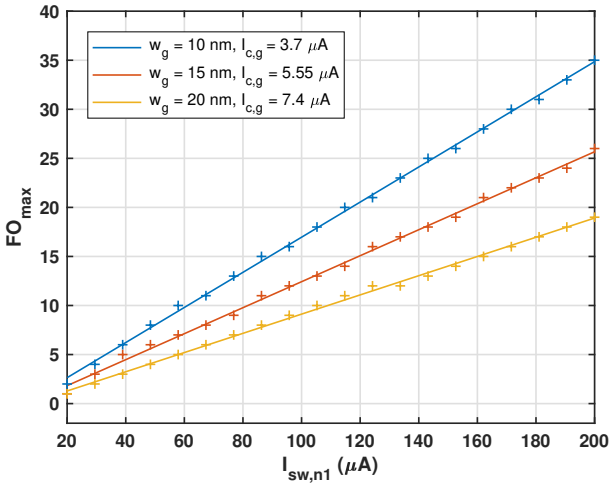

Figure 4.13: Maximum fan-out as function of the switching current $I_{sw,n1}$ of the main neuron, at different values of gate critical current $I_{c,g}$. The marked points were obtained by simulations, the curves fit the points. The parameters of the neurons follow expressions shown in table 4.1. Main neuron parameters: $R_{sh,main} = 50\ \Omega$, $R_{NW,nT} = 10\ \Omega$. Synapses parameters: $R_{series,in} = 1\ \Omega$, $R_{sh,in} = 5\ \Omega$, $I_{b,nT} = 36\ \mu A$, $R_{series,out} = (100\ \mu A / I_{sw,n1})\ \Omega$. Target neurons parameters: $I_{sw,n2} = 30\ \mu A$.

A different configuration was tested, in which all $R_{sh,in}$ and $R_{series,in}$ were removed, and a single shunt resistor was put in parallel with the gates of all the nTrons. Doing so, $I^{min}_{in,nT} = I_{c,g}$ and the maximum fan-out can be increased by 17%, but the the system loses in terms of stability. Indeed the synapses are not more isolated from each other. When they have different weights, the nTrons are biased at different currents, so the propagation velocities of the hotspot are also different. This means that the gates does not switch at the same time, and back-propagation of the gate currents alters the behavior of some synapses. As consequence this configuration was discarded, but the same effect was observed to a lesser extent also in the configuration shown in figure 4.10. Therefore Further work is necessary to improve the isolation between synapses (also between main and target neurons). In section 4.3 the hTron, a different type of cryotron, is proposed as possible solution.

Fan-in

The circuit schematic in figure 4.14 was analyzed and simulated to find the maximum fan-in. The expression 4.10 for the maximum fan-in is only a rough estimation, that is not really valid exploiting nTron synapses, due to the charging and discharging dynamic of the synaptic inductance with multiple spikes, not considered in the equation. However it is still useful to understand the dependencies and what can

be done to improve the fan-in characteristics.

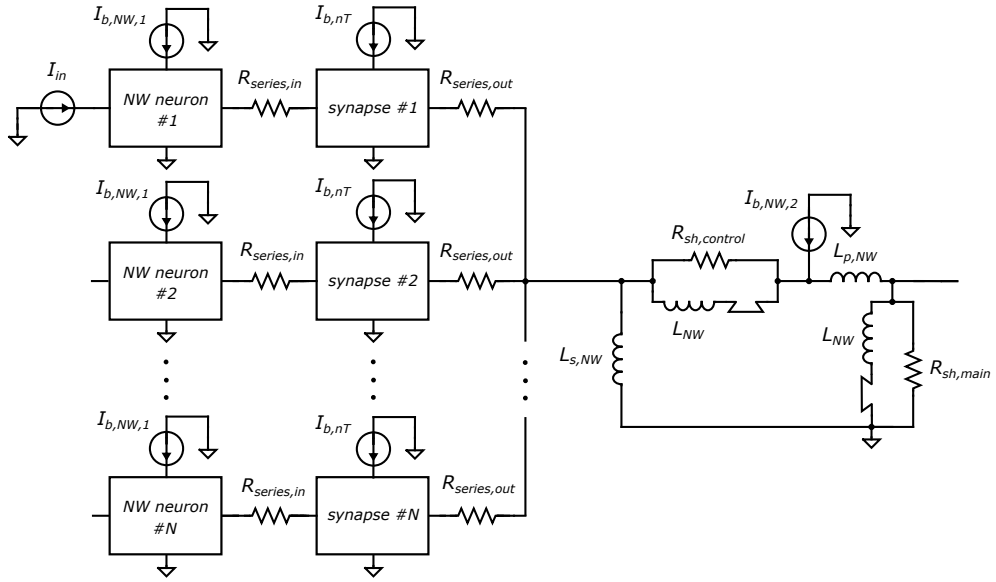

Figure 4.14: Schematic of the circuit simulated to study fan-in performances. N NW neurons fan into one target neuron through nTron synapses, and only one of the N neurons is activated through I_{in}.

In this configuration, the current provided to the downstream neuron is:

$$I_{in,n2} \approx WI_{out,n1} \frac{1}{1 + N\left(\frac{Z_{in,n2}}{Z_{out,syn}}\right)} \tag{4.13}$$

where $WI_{out,n1}$ depends on the bias current of the nTron. W is function of the number of incoming pulses to the synapse. Using the synaptic parameters previously introduced, with 9 pulses it is possible to reach approximately the maximum value $WI_{out,n1} = 12\ \mu A$. $I_{in,n2}$ is also function of $Z_{out,syn}/Z_{in,n2}$: higher it is, more current can be provided to the neuron. In fact, part of the current generated by a single synapse can back-propagate to the other synapses (cross-talk current), and its value is lower if the output impedance of the synapses is much higher than the input impedance of the target neuron. The input impedance of the target neuron is:

$$Z_{in,n2} = \left(\frac{1}{j\omega L_{s,NW}} + \frac{1}{2(R_{sh}//j\omega L_{NW}) + j\omega L_{p,NW}}\right)^{-1} \tag{4.14}$$

while the output impedance of the synapse is:

$$Z_{out,syn} = \left(\frac{1}{R_{syn,2}} + \frac{1}{(R_{syn,1}//j\omega L_{nT}) + j\omega L_{syn}} \right)^{-1} \tag{4.15}$$

The ratio $Z_{out,syn}/Z_{in,n2}$ depends on the FFT of the output signal of the synapse, so it varies with the temporal pattern of spikes that charges the synaptic inductance. By a first rough estimation, if we consider that the output signal is almost a DC signal, $Z_{in,n2} = j\omega L_{s,NW}//j\omega(2L_{NW} + L_{p,NW})$ and $Z_{out,syn} = j\omega(L_{nT} + L_{syn})$. Using the parameters of the previous section, the ratio is approximately $Z_{out,syn}/Z_{in,n2} = 20$. This means that the fan-in structure can work properly with a low cross-talk current up to N close to 20, but a more detailed analysis is needed to find the correct ratio through simulations.
The circuit in figure 4.14 was simulated to obtain the peak input current $I_{in,n2}$ as function of N and the the number of incoming spikes ($I_{b,nT} = 36\ \mu A$). The result obtained using all the parameters of table 4.1 and $I_{sw,n1} = I_{sw,n1} = 30\ \mu A$ is shown in figure 4.15.

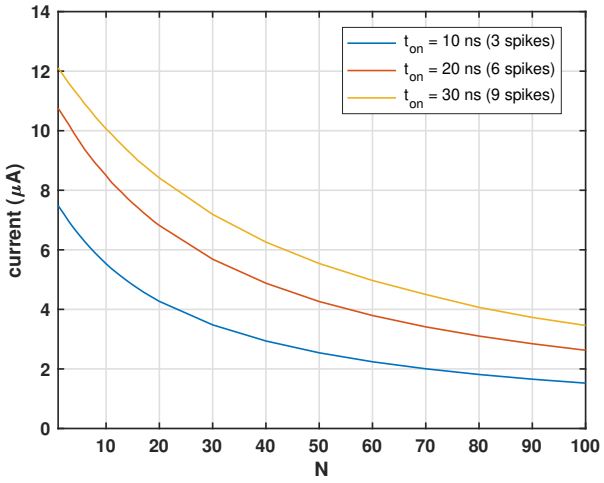

Figure 4.15: Input current $I_{in,n2}$ of the target neuron of figure 4.14, as function of the number of neurons in the first layer N, for different number of pulses generated by the NW neuron 1 (or different t_{on} of the pulse generated by I_{in}). The neurons parameters are the ones of table 4.1 with $I_{sw} = 30\ \mu A$. The target neuron has $I_{b,NW,2} = 0\ \mu A$. Synapses parameters: $R_{series,in} = 20\ \Omega$, $R_{sh,in} = 5\ \Omega$, $I_{b,nT} = 36\ \mu A$, $R_{series,out} = 5\ \Omega$.

The three curves were fitted with the function of equation 4.13, obtaining three different ratios $Z_{out,syn}/Z_{in,n2}$: ~ 25 for 3 spikes, ~ 35 for 6 spikes and ~ 42 for 9 spikes. The minimum current to activate a NW neuron with parameters of table

4.1, is $I_{in,n2}^{min} = (2.07I_{sw,n2} - I_{b,NW,2})$. So the expression of the maximum fan-in becomes:

$$FI_{max} \approx \left(\frac{Z_{out,syn}}{Z_{in,n2}}\right)\left(\frac{WI_{out,n1}}{2.07I_{sw,n2}} \cdot \frac{1}{1-\left(\frac{I_{b,NW,2}}{2.07I_{sw,n2}}\right)} - 1\right) \tag{4.16}$$

This equation shows that the fan-in can be increased also decreasing $I_{sw,n2}$, or biasing the target neuron closer to $2I_{sw,n2}$. This last option would decrease the noise margins. Setting $I_{b,NW,2} = 1.9I_{sw,n2}$, in the best case of 9 spikes that charge the synapse, the resulting maximum fan-in is:

$$FI_{max} \approx 1.35 \cdot \left(\frac{Z_{out,syn}}{Z_{in,n2}}\right) \approx 56 \tag{4.17}$$

This is a good result if compared with the one of the JJ-based network that exploits the current method for the fan-in, introduced in [11], for which $FI_{max} \approx 20$. it is not an exciting result if compared with the one of the flux method [11], that can reach a fan-in of ~ 100. Further work is needed to improve the performances, but it should be reasonable to reach values higher than 100. The first things to do are decrease $Z_{in,n2}$ using lower $L_{s,NW}$, and decrease $I_{sw,n2}$. Afterwards it could be possible to increase $Z_{out,syn}$, just increasing L_{syn} and $R_{series,out}$, or increase $WI_{out,n1}$ using an higher I_c for the nTron. It is important to notice that the first method would lower the output current, and the second would require to lower $R_{syn,1}$ to avoid latching, consequently decreasing the output current too. A right balance between parameters needs to be found.

4.2.3 nTron splitter for fan-out boosting

The maximum fan-out achievable by the NW system with nTron synapses is higher than the one obtained with inductive synapses, but it is still far from values of biological system in which a neuron can drives thousands of others. Moreover the structure introduced in the previous section is limited by the gate threshold current, and the need to increase the switching current of the main neuron with the consequent decrease of spiking frequency. For these reason, it was designed a new structure inspired by a recent work [11], in which the fan-out of a system based on JJ neurons was boosted exploiting JJ splitters, typically used in SFQ circuits [25]. The splitter is able to generate a pulse on two output lines receiving one pulse on the input. Creating a fan-out circuit composed by nested splitters organized on different layers allows to drastically increase the maximum fan-out, ideally without limit. Obviously the drawback is the larger occupied area by the network, and the increased power consumption.
Here it is proposed a similar structure based on nTron splitters, that can be nested

to boost the fan-out of NW neurons, reaching performances of biological systems and avoiding the increase of switching current.
The circuit schematic of a single nTron splitter is shown in figure 4.16A. It is composed by n nTrons that are activated by a single current pulse through $R_{series,in}$ on the input line. All the nTrons fire contemporary so that n pulses are produced on the output lines. It is possible to use just one $R_{sh,in}$ shared between nTrons, because they are all biased with the same current $I_{b,nT}$, so there are not problems of cross-talk. The maximum number of nTrons n that can be used depends on how much current can be provided on the input. The nTron splitter has to be designed in order to allow a single output line to drive another splitter, so that a nested structure organized on more layers can be created. Figure 4.16B shows a 1-to-4 2-layers structure formed by 1-to-2 nTron splitters.

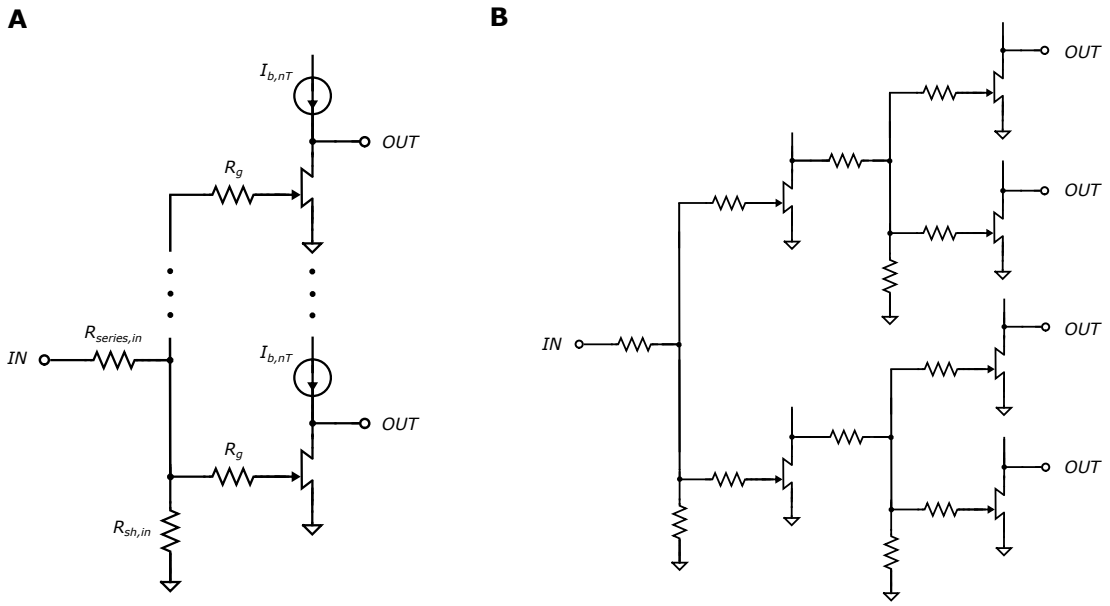

Figure 4.16: **(A)** Circuit schematic of a 1-to-n nTron splitter. **(B)** Circuit schematic of 1-to-4 tree composed by 1-to-2 ntron splitters (current sources are omitted for simplicity).

More layers can be added to produce more spikes. If a NW neuron drives the whole structure and the nTrons of the last layer are replaced with nTron synapses, the action potentials of the main neuron can activate a very high number of synapses. The fan-out grows exponentially with the number of layers:

$$FO = n^{N_L} \tag{4.18}$$

Higher is n, less layers are needed to reach a certain fan-out. The number of

nTron used in the structure is:

$$N_{nT} = \sum_{k=1}^{N_L} n^k \tag{4.19}$$

In order to demonstrate the utility of the structure, it was simulated a 3-layers nested tree formed by 1-to-3 nTron splitters, used to obtain a fan-out of 27, with a NW neuron that has $I_{sw} = 30\ \mu A$. Figure 4.17 shows the time-domain behavior of the system. Each action potential of the main neuron activate the nTrons in the first layer that in turn activate the ones in the second layer, which make the nTron synapses switch, so that the target neurons can fire.

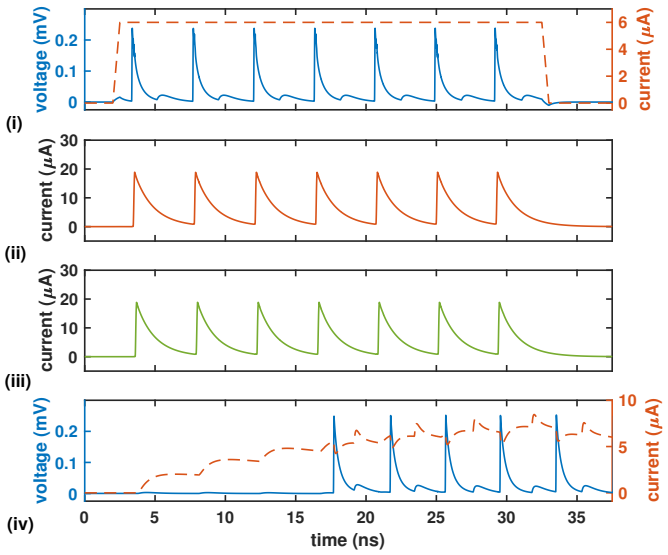

Figure 4.17: Time-domain behavior of a NW neuron that drives a 3-layers 1-to-27 fan-out tree composed by 1-to-3 nTron splitters. In the last layer the nTrons are replaced by nTron synapses, that activate NW neurons. The NW neurons parameters are the ones of table 4.1 ($I_{sw} = 30\ \mu A$). nTron synapses parameters: $I_{b,nT} = 36\ \mu A$. nTron splitters parameters: the nTrons have the same parameters of the nTron synapses, but $I_{b,nT} = 25\ \mu A$, $R_g = 2\ \Omega$, $R_{series,in} = 8\ \Omega$. **(i)** Input current of and output voltage of the main neuron. **(ii)** Output current of a nTron in the first layer. **(iii)** Output current of a nTron in the second layer. **(iv)** Input current and output voltage of one of the target neurons.

Using the same parameters of the nTron synapse for all the nTrons and a bias current of 25 μA, the energy consumption per spike is approximately 5 aJ, ten

times smaller than the energy consumed by a NW neuron $E_{n,NW}$.
In the JJ-based system introduced in [11] n is always 2 three junctions are used for each splitter. If we compare the total number of elements with a structure based on 1-to-2 nTron splitters, the number of used junctions is

$$N_{JJ} = 3 \cdot \sum_{k=0}^{N_L} 2^{k-1} = \frac{3}{2} \cdot \sum_{k=1}^{N_L} 2^k = \frac{3}{2} \cdot N_{nT} \tag{4.20}$$

so less elements are used in a NW-based structure, than in a JJ-based one. Moreover the nTron splitter is more flexible in terms of number of output line and layers, considering that the JJ splitter only generates two pulses. For the JJ-based system it was also proposed a different structure more similar to nTron splitters named "current-based fan-out" that also allows to set the number n, but it is less efficient for its higher number of junctions, so it is not considered here. The energy consumed by a junction in the JJ splitters is approximately 0.5 aJ per spike, 2 times smaller than the energy consumption of a JJ neuron $E_{n,JJ}$. The Josephson junction will always be more energy efficient than nanowires, so in order to compare the two systems is more fear to normalize the total energy of the splitters trees with the one of the associated neuron: the normalized energy of the NW-based splitters system is $E_{NW}/E_{n,JJ} = N_{nT}/10$, while for JJs it is $E_{JJ}/E_{n,JJ} = 1.5 \cdot N_{nT}/2$, so in these terms the nTron splitter seems to be 7.5 times more efficient.
The nTron splitter is firstly proposed here and there are not experimental data to demonstrate it can work properly. Furthermore a more accurate analysis would be needed to understand how much area would be occupied by such structure.

4.3 hTron synapse for isolation improvement

The nTron synapses presented problems of isolation between input and output due to the intrinsic leakage current of the nTron. It was noticed in simulations, that for certain set of synaptic parameters the spiking of the target neuron altered the activation of the nTron. Moreover, as already explained, in the fan-out configuration it was observed that setting different values of bias current for nTron synapses driven by a single neuron, some nTrons do not spike or latch if their gates are not correctly isolated from the others. These problems could be solved replacing the nTron with a M-hTron, because the input impedance of the synapse would be constant since the heater is normal, and the output would be almost completely isolated by the input. Moreover the driving capability of the synapse would be increased. The SPICE model of the hTron, introduced in section 3.4, was used to study a possible synapse based on the hTron. The hTron was placed in the same configuration of the nTron, but the shunt resistance of the gate $R_{sh,in}$ was removed, since the heater is normal. The critical current of the hTron was lowered to 40 μA

to have the same strength of the nTron synapse (the channel and heater widths were lowered to 200 nm). The other parameters of the synapse were changed in order to adapt to the hTron characteristics. The input resistance of the heater was set to 10 Ω and for sake of simplicity the thicknesses of the NbN, Ti and SiO_2 films were kept equal to the ones realized by [4]. The time-domain behavior of two neurons coupled with excitatory control through hTron synapse is shown in figure 4.18.

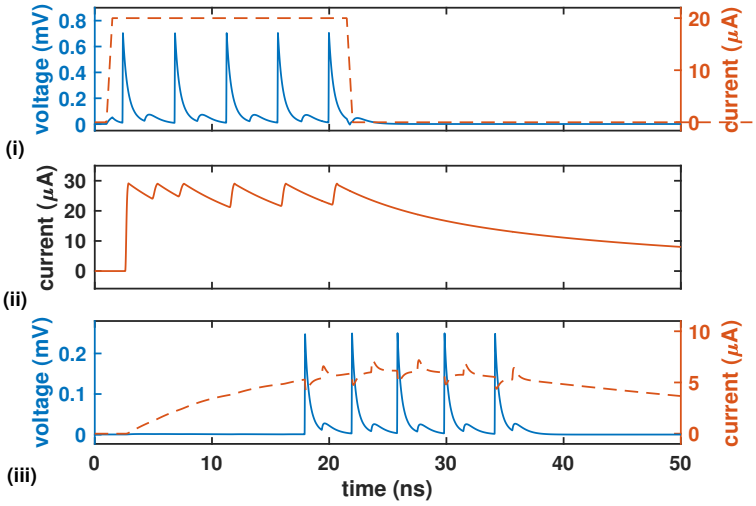

Figure 4.18: Time-domain behavior of two NW neurons coupled with excitatory control through an hTron synapse. Parameters shared by the two neurons: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh,control} = 10\ \Omega$. Main neuron parameters: $I_{sw} = 100\ \mu A$, $I_{bias} = 190\ \mu A$, $I_{in} = 20\ \mu A$, $R_{sh,main} = 50\ \Omega$ Target neuron parameters: $I_{sw} = 30\ \mu A$, $I_{bias} = 57\ \mu A$, $R_{sh,main} = 10\ \Omega$. hTron synapse parameters: $L_{drain} = 100\ nH$, $R_{series,in} = 0\ \Omega$, $R_{in} = 10\ \Omega$, $I_{c,ch} = 114.5\ \mu A$, $R_{sh,in} = 0\ \Omega$, $R_{syn,1} = 10\ \Omega$, $L_{syn} = 300\ nH$, $R_{syn,2} = 50\ \Omega$, $R_{series,out} = 20\ \Omega$. **(i)** Parameters: $I_{b,nT} = 36\ \mu A$. Panel **(i)** displays the input current and the output voltage of the main neuron; Panel **(ii)** displays the output current of the hTron, that enters into the inductive loop; Panel **(iii)** displays the input current and the output voltage of the target neuron.

Since the thicknesses of the films in the hTron were not modified, the threshold current through the heater that makes the channel switch is much higher of $I_{c,g}$ in the nTron. Therefore the switching current of the NW neuron was set to 100 μA, and $R_{sh,main}$ to 50 Ω (The parameters of table 4.1 were not used). Changing the structure of the hTron ideally could be possible to decrease the threshold current. The parameters used in this simulations are not optimized and further experimental

work would be needed to understand if it is possible to realize hTron with these characteristics, however the result is useful to show that in theory it would be possible to create a hTron synapse that would improve the isolation, and would increase the output current. The possibility of inhibitory control and variable strength were not analyzed.
Considering that it was demonstrated the hTron has a higher driving capability, it could be possible to exploit it also for splitters in order to further boost the fan-out.

Chapter 5

Nanowire-based Neurons and synapses: fabrication and characterization

In chapter 4 the nTron synapse was introduced, and a detailed analysis of its characteristics was done, thanks to simulations. However a synapse used to connect two NW neurons was never concretely realized, so the coupling between two neurons was not demonstrated. In this work the first steps for a long-term plan, that aims to demonstrate experimentally the possibility of coupling two neurons with a nTron synapse, were set. First of all the basic elements of neurons and synapses were designed and fabricated: (1) relaxation oscillators; (2) nTrons; (3) inductive synapses. Afterwards they were characterized, and the first attempts to test the neurons and the synapses were made. In this chapter the fabrication process, the layout, the experimental setups and the results of the characterization are shown. Due to lack of time, the characterization of all the fabricated devices was not completed, and the operation of neurons and nTron synapses was not well demonstrated. However interesting results that can help a future design of the devices are here presented. Definitely further work will be needed to reach the final goal of the plan.

5.1 Neuron

The NW neuron has already been fabricated and tested [29], but an appropriate analysis of the fabricated shunt resistors has not been yet realized. Considering that the process to obtain L_{NW} is well standardized, it is important to have a good control of the shunt resistance in order to set the time constant L_{NW}/R_{sh}

and consequently the spiking frequency. The dependence of the spiking frequency on the resistance has not been studied experimentally. Furthermore the value of the resistance used in previous works [29] is very low (1.2 Ω), so the output peak voltage of the neuron approximately equal to $I_{sw}R_{sh}$ results to be on the order of 50 μV. With this low value the effect of the noise is high and the driving capabilities are limited, since the output impedance is low. Realizing oscillators with resistance in the order of 10 Ω like the one used in simulations would solve the problems. For these reasons isolated oscillators and neurons with different shunt resistances were fabricated and characterized. The results of the measurements are shown in this section.

5.1.1 Fabrication

For both the resonators and the somas, the same structures designed in [29] were reproduced. Figure 5.1 shows SEM images of a fabricated oscillator and soma. The nanowire is composed by a 60-nm wide switching element, and two meandered structures 150-nm wide. The critical current depends on the width of the constriction in the switching element, where the hotspot is generated.

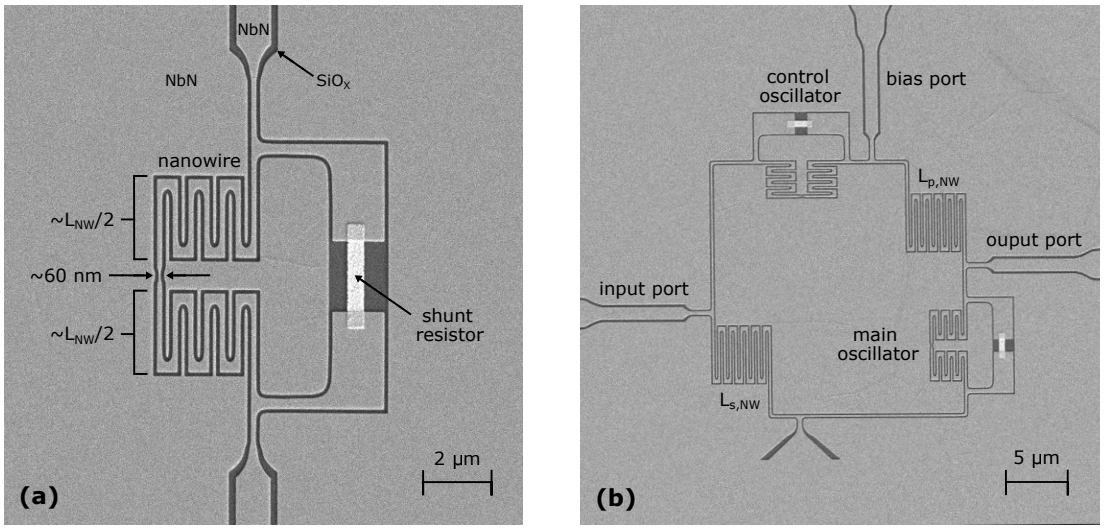

Figure 5.1: **(a)** SEM image of a relaxation oscillator with 4-squares shunt resistor. **(b)** SEM image of a NW neuron composed by two coupled oscillators

The kinetic inductance L_{NW} depends on the number of squares (~ 200) of the entire nanowire, and the inductivity of the film, which is directly proportional to the sheet resistance of the NbN film. The nanowire is designed to have a time constant in the order of nanoseconds, depositing a film with a sheet resistance of $\sim 150\ \Omega/sq$, and as consequence an inductivity of $\sim 20 - 50\ nH/sq$. The shunt resistor was fabricated as close as possible to the nanowire.
The neuron is formed by a superconducting loop composed by two coupled oscillators and two kinetic inductances ($L_{s,NW}$ and $L_{p,NW}$). The latter were realized with meandered nanowires. The two branches of the loop were designed to have the same number of squares (~ 634), so that the bias current can be equally split.

On the same 1x1 cm^2 die cut from a 4 in. Si wafer single resistors, oscillators and somas were fabricated all together. Each element were fabricated with 5 different structures in which the resistors have same film thickness but different numbers of squares (0.5, 1, 2, 3, 4). Three attempts were done before reaching the desired result. The fabrication process used for the last attempt is described in the next section, while the results of the two first attempts are explained afterwards.

Fabrication process

The isolated resistors, oscillators and the somas were fabricated with the multi-steps lithographic process shown in figure 5.2, on a silicon substrate (covered by 300 nm-thick thermal oxide).

As done by other works on NbN nanowires [42], the resistors were placed beneath the NbN film, obtaining a two-layers structure. The marks necessary to align the two layers were patterned with electron-beam lithography (Elionix E125) through a bilayer resist process. They were realized with two separate steps, due to their different thicknesses. For both the steps, firstly the positive-tone resist ZEP520 was spun on the substrate at 5 krpm, and baked at $180°C$. Then the resist was developed in O-Xylene at $5°C$ for 90 s, and rinsed in IPA at room temperature.

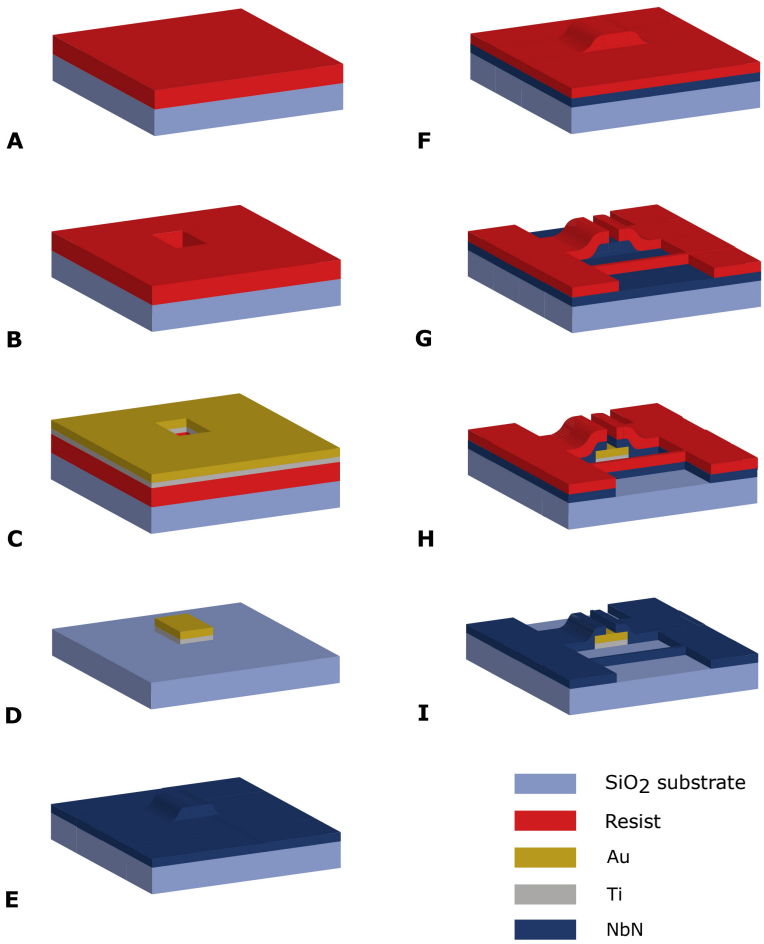

Figure 5.2: Fabrication process for shunted NbN nanowires (the shape of the structure is simplified in this figure). **(A)** Resist spinning (ZEP520). **(B)** E-beam lithography and development of the shunt resistor pattern. **(C)** Deposition of Ti-Au film. **(D)** Lift-off. The shunt resistor is patterned. **(E)** Deposition of NbN. **(F)** Resist spinning (ZEP530). **(G)** E-beam lithography and development for the nanowire and the contacts with the resistor. **(H)** Reactive Ion Etching (RIE). **(I)** Resist removal. The shunted nanowire is patterned.

For the markers a 10-nm layer of Ti and a 50-nm layer of Au were evaporated on the resist, and the lift-off was performed in 1-methyl-2-pyrrolidone (NMP) at $60°C$ for one hour. For the shunt resistors a 10-nm Ti + 25-nm Au layer was realized with the same technique. Before starting creating the NbN layer, the chip was ashed in O_2He plasma with a power of 100 W for 1 min., in order to remove impurities from the metallic film and ensure a good adhesion with the NbN film. Better is the adhesion lower is the contact resistance between the two films, and more controllable is the value of the shunt resistance. The NbN film was deposited

on the entire chip in AJA sputtering system for 390 s with a peak power of 160 W, obtaining a thickness of 16.8 nm. Then the nanowires structures were patterned with a second electron-beam lithography step, exploiting the previously fabricated marks to align them with the shunt resistors. In this lithographic process a different resist was used (ZEP530), with the same recipe explained above. In this case the final structure was obtained removing the regions of NbN uncovered by resist, using reactive ion etching (RIE) with CF_4 at 50 W of power. The residual resist was then removed again with NMP.

The three different attempts for the fabrication of the resistances are:

- **5-nm Ti + 15-nm Au and 13-nm NbN:** The thickness of the metallic film was reduced respect to previous works to increase the sheet resistance, while the NbN film was made thinner to obtain a faster cooling down, so a higher latching current. In this way it should have been easier to increase the shunt resistor. The resulting resistances were all on the order of $k\Omega$ and were not linear proportional to the number of squares, so the oscillators were not shunted. Probably the metallic film was too thin, or the contact resistances too high because it was not performed the ashing step in the process to clean the metallic surface.

- **10-nm Ti + 15-nm Au and 16.8-nm NbN:** After the first attempt the ashing step was introduced, the thicknesses of the Ti film and the NbN film were increased. Two chips were fabricated with same NbN and Ti thiknesses, but different Au films (15 nm and 25 nm). The resistances of the 15-nm film were in the range $90 - 250\ \Omega$, and it was seen that the values were not almost dependent on the number of squares.

- **10-nm Ti + 25-nm Au and 16.8-nm NbN:** With this Au thickness the resistances resulted to be on the order of $10\ \Omega$, with lower dependence on the contacts.

All the following results of the characterization were obtained with the 10-nm Ti + 25-nm Au 16.8-nm NbN chip. Firstly isolated oscillators were tested, secondly the somas.

5.1.2 Relaxation oscillators: characterization

First of all a I-V curves of a unshunted nanowire and shunted nanowires with different resistances were realized in helium Dewar at 4.2 K. A bias current was applied to them through a battery source in series with a 100 $k\Omega$ resistor and a 1.9-MHz low-pass filter used to filter any kind of noise that could make the nanowire

switch. The voltage across the nanowire was measured by a multi-meter with a 1 $M\Omega$ resistor in series. Figure 5.3a shows the result for the unshunted nanowire. The switching current is 38.2 μA, and the curve is hysteretic with $I_r = 5.2\ \mu A$. Figure 5.3b shows the I-V curves of the shunted nanowires with different shunt resistances. The switching currents are in the range $(30.3 - 38.2)\ \mu A$, with average value of $I_{sw} = 34.4\ \mu A$.

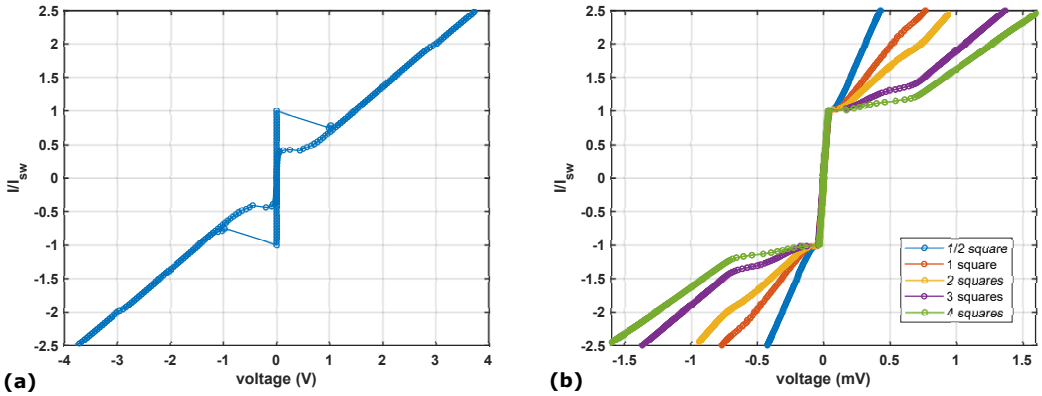

Figure 5.3: **(a)** I-V curve of an unshunted nanowire. **(b)** I-V curves of shunted nanowires with different shunt resistances. There is a 1 Ω parasitic resistance due to the cables used to measure.

The shunted nanowires are not hysteretic, and the voltage increases with the number of squares. At a certain latching current, the nanowires stop spiking, so above that point the slope of the curve corresponds to the value of the shunt resistance, therefore it is possible to extract it. It has to be considered that there is a 1 Ω parasitic resistance that comes from the cables used for the measurements, that can be seen in the superconducting state. The graph clearly shows that the latching current decreases with increasing resistance, as expected.

Resistances

The values of the resistances were extracted from the latching state in the previous graph (for the 0.5-squares oscillator it was extracted from its spiking behavior shown in the next section, because the latching current was higher than the switching current of the entire structure 100 μA). For each oscillator, it was also measured the resistance of the associated isolated resistor, that has the same structure. The resulting values as function of the number of squares are shown in figure 5.4.

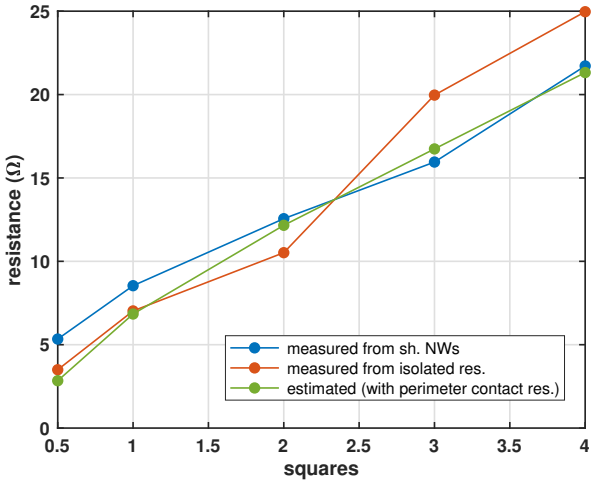

Figure 5.4: Resistance as function of the number of squares. (blue) extracted from the I-V curves of the shunted nanowires. (red) measured from the isolated resistors. (green) obtained with the fitting values of sheet and contact resistance.

The resistances were patterned with different widths, the overlap between the NbN and the Au films has a 0.5x0.5 um^2 area for 2,3,4-squares resistors, 0.5x1 μm^2 for 1-square and 0.5x2 μm^2 for 0.5-squares. The narrower cross section the current has to pass through is the one on the border of the contact region with an area of $\sim d \cdot P$, where d is the thickness of the NbN film and P is the perimeter of the contact region. Assuming this, the contact resistance has to scale with the perimeter of the contact area in the measured resistances. Considering that the 2,3,4-squares resistors have the same contact area, it was fitted the curve formed by that three points (blue curve) obtaining a sheet resistance $R_\square = 4.58\ \Omega/sq$ and a contact resistance $R_c = P \cdot 1\ \Omega \cdot \mu m$ (or $R_c = d \cdot P \cdot 0.017\ \Omega \cdot \mu m^2$). The green curve shows the estimated resistances considering the scaling of the perimeter contact resistance. It is consistent with the 0.5,1-square values obtained by isolated resistors (red curve).

Spiking behavior

After the extraction of the resistances, the spiking behavior of the oscillators was tested, measuring the spiking frequency as function of the bias current. The oscillator was connected to a bias-tee. The bias current was applied to the DC port by a battery source in series with a $100k\Omega$ resistor, while the RF port was connected to a 1 GHz-bandwidth 50dB amplifier (MITEQ AM-1309). The output of the amplifier was sent to the oscilloscope, used to apply the Fast Fourier Transform (FFT) to the signal and find the dominant frequency peak for each value of bias

current. Figure 5.5 shows the time-domain behavior of the 1-square oscillator with $I_{bias}/I_{sw} = 1.047$.

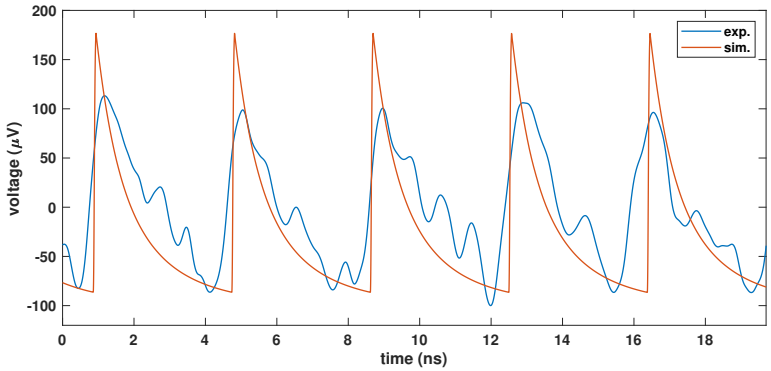

Figure 5.5: Comparison between experimental and simulated time-domain behavior of the 1-square oscillator. For simulation: $L_{NW} = 6.4\ nH$, $R_{sh} = 8.53\ \Omega$, $I_{sw} = 38.19\ \mu A$, $I_{bias}/I_{sw} = 1.047$

The orange curve was obtained with simulation of the LTspice model, in which the inductance was set to $L_{NW} = 6.4\ nH$ in order to obtain the spiking frequency of the experimental result. The same thing was done for all the oscillators and the average value of kinetic inductance resulted to be $L_{NW} = 6.5\ nH$. The experimental curve presents many reflection artefacts, because the nanowire is not matched with the 50 Ω line. Probably for these reason the measured voltage peak is slightly lower than the simulated one.

Figure 5.6 shows the spiking frequency as function of I_{bias} in the range between I_{sw} and I_L (the latching current) for all the oscillators. Outside of this range the spiking frequency is zero.
This graph clearly demonstartes that increasing the shunt resistance, it is possible to have the same spiking frequency with lower values of bias current, but the nanowire latches at lower current. For frequencies above 1 GHz the data are more noisy because we are in the cut-off region of the amplifier. However the maximum value reached is $\sim 1.25\ GHz$, which is very high if compared with the ones of previous works [29].

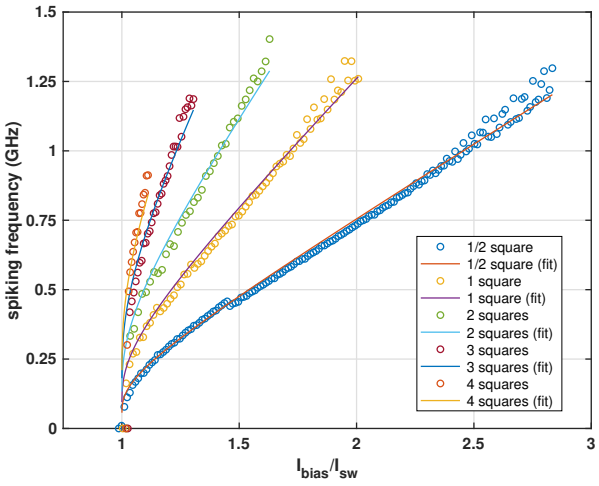

Figure 5.6: Spiking frequency of the oscillators as function of the bias current, at different shunt resistance. The graph shows the range between I_{sw} and I_L (the latching current). Outside of this range the spiking frequency is zero. The curves are fitted with equation 5.1.

The curves were fitted with the expression obtained by simplification of equation 1.10:

$$\frac{1}{f} \approx -\frac{L_{NW}}{R_{sh}} \ln\left(\frac{I_{bias} - I_{sw}}{I_{bias}}\right) \tag{5.1}$$

Using the previously extracted values of resistances for 1,2,3,4 squares, it was obtained $L_{NW} \approx 10\ nH$. From this value it was extracted the resistance of the 0.5-square oscillator. The inductance is higher than the one obtained matching the time behavior of simulation and experiment, but considering that equation 1.10 is just an approximation, the value 6.5 nH was taken as correct. This means that the expression of the spiking period should be divided by a 1.5 correction factor.

5.1.3 Soma: characterization

Once the oscillators had been characterized, the somas with 0.5 and 2-squares shunt resistors were tested. Firstly they were overbiased above their critical current with zero input current, so that the two shunted nanowires oscillate in phase, in order to check if the neuron can fire. The same frequency analysis of the previous section was performed, obtaining the graph in figure 5.7A. Here the bias current was provided by a battery source in series with a 100 $k\Omega$-resistor and the filter. The input signal was sent to a bias-tee with 50 Ω-termination on the DC port in series with the amplifier.

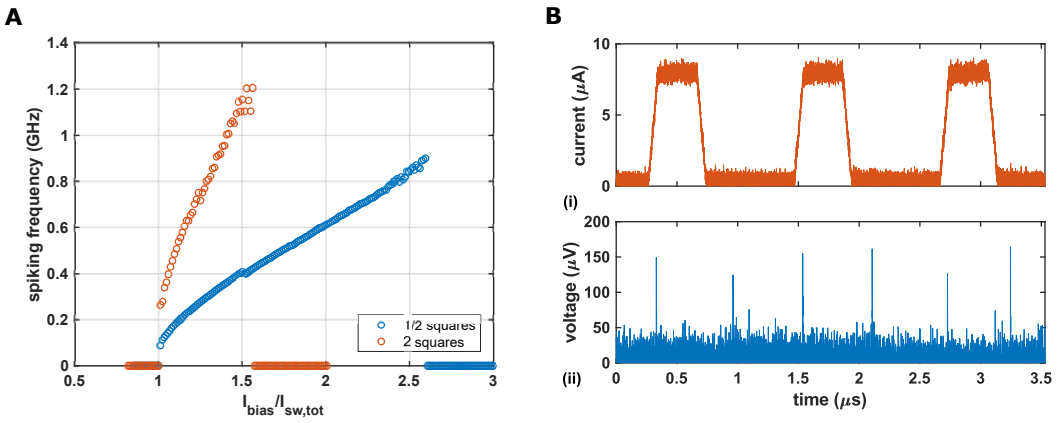

Figure 5.7: **(A)**Spiking frequency of the neuron as function of the bias current, at different shunt resistance and zero input current. The voltage is measured on the output of the neuron. $I_{sw,tot}$ is the switching current of the neuron, which can be different from $2I_{sw}$ if the two oscillators have different switching current. **(B)** Experimental time-domain behavior of 0.5-square neuron. $I_{sw,tot} = 66\ \mu A$, $I_{bias} = 63\ \mu A$. **(i)** Input current of the neuron. **(ii)** Output voltage of the neuron.

The neurons can properly spike when the bias current is above the threshold $I_{sw,tot}$, and the frequency follows the same behavior of a single oscillator, but with a slightly lower maximum frequency.

Afterwards the two neurons were tested with an applied input current. Multiple combinations of I_{in} and I_{bias} were used, but both the neurons did not work properly. The input current pulse was provided by a waveform generator with a 10 $k\Omega$-resistor in series. The voltage on the output of the generator was directly sense by the oscilloscope (1 $M\Omega$-mode). As shown in figure 5.7B, the 0.5-square neuron with $I_{sw,tot} = 66\ \mu A$, $I_{bias} = 63\ \mu A$ and $I_{in} = 9\ \mu A$, fires only during the rising and falling edges of the input current pulse. The same thing happens for the 2-squares neuron. This behavior typically is caused by uneven oscillators that have different inductances, switching current or shunt resistances. It was observed in previous works [29], when the input current is higher than a certain threshold. Here it was tried to decrease the input current and increase the bias current to avoid this phenomenon, but the neuron started spiking randomly due to noise. Further work would be needed to understand where the problem comes from, performing electrical simulations, and testing more devices.

5.2 nTron synapse

The nTron synapse was not integrally fabricated. Firstly the nTrons and synaptic inductors were separately realized on the same chip of the oscillators and neurons, in order to characterize them before testing the performance of the nTron synapse. The work of characterization was not completed, so in this section only the preliminary results are shown.

5.2.1 Fabrication

Both for the nTrons and the synaptic inductors (or standard inductive synapses) were fabricated at the first fabrication attempt (5-nm Ti + 15-nm Au and 13-nm NbN), following the process explained in section 5.1.1. The sheet resistance of the obtained NbN film is $\sim 180\ \Omega/sq$. Figure 5.8a shows the SEM image of a nTron, while figure 5.8b shows the SEM image of the inductor with associated parallel resistances.

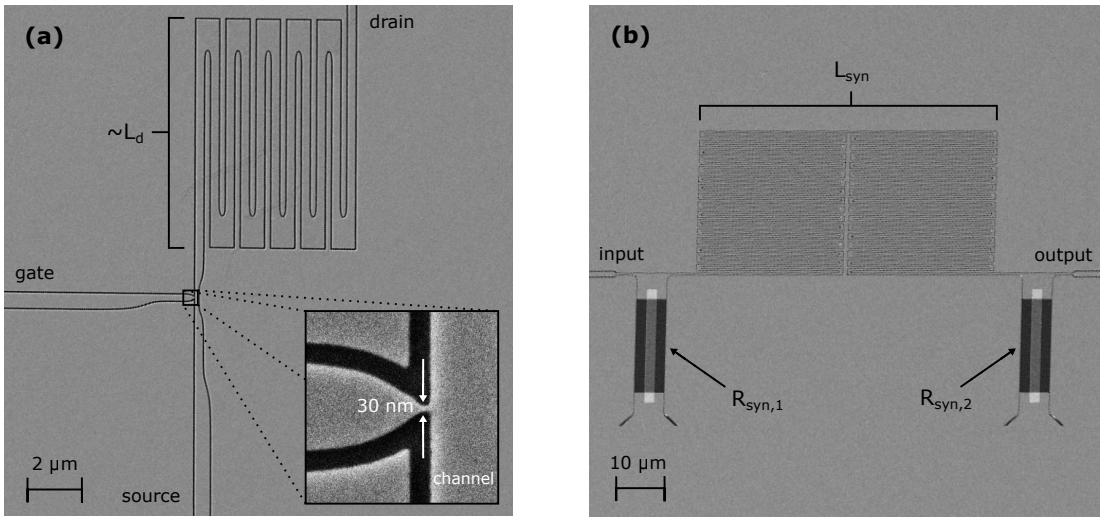

Figure 5.8: (a) SEM image of a fabricated nTron (30-nm wide choke and 300-nm wide channel). L_d is the drain kinetic inductance ($\sim$ 300 squares), the total drain nanowire is $\sim$ 400-squares long. The gate is 175-squares long, while the source is 120 squares (There were realized also nTrons with $\sim$ 20squares long source and gate). The small inset shows the 30-nm wide choke (there were realized also nTrons with the choke in proximity of the bottom of the channel). **(b)** SEM image of the inductive synapse. Both $R_{syn,1}$ and $R_{syn,1}$ are 9-squares long.

The nTron does not have a resistances in its structure, so it is formed only by a 13-nm thick NbN film. The kinetic inductor of the drain was realized with a meandered nanowire ($\sim$ 400 squares). The gate nanowire is $\sim$ 170-squares long, while the source is $\sim$ 120-squares long. Multiple nTrons were fabricated on the same chip with slightly different geometries, to study how their dimensions control the property of the device. There were fabricated nTrons with different choke width (10 nm, 20 nm, 30 nm, 40 nm) and channel width (100 nm, 200 nm, 300 nm ,400 nm). Some nTrons were also fabricated differently respect to the one shown in figure 5.8a: the choke was realized in proximity of the bottom part of the channel, and the gate and source inductances were lowered, reducing the number of squares to $\sim$ 20. Only some devices were tested among them.

The inductive synapses were designed to have a inductance L_{syn} at least higher than 200 nH, realizing a 8500-squares meandered nanowire. With a film inductivity of $\sim 30-40$ pH/sq, the resulting inductance is in the range $250-340$ nH. Multiple inductors were realized with same number of squares, but different resistances $R_{syn,1}$ and $R_{syn,2}$ (from 3 to 12 squares), in order to test the circuit with different time constants. Since the fabricated resistances resulted to be on the order of $k\Omega$, they were added externally with resistors soldered on the PCB, as it will be explained afterwards.

5.2.2 nTron: characterization

Firstly I-V curves of some nTrons with zero gate current were obtained in helium Dewar at 4.2 K. A bias current was applied to the drain through a battery source in series with a 100 $k\Omega$ resistor (at room temperature). The drain voltage was measured by a multi-meter with a 1 $M\Omega$ resistor in series. Figure 5.9 shows the result for a nTron with 20-nm wide choke and 200-nm wide channel.

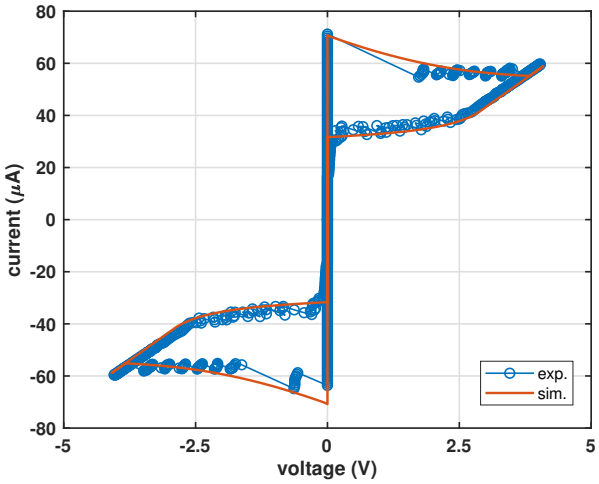

Figure 5.9: I-V curve of the nTron with 20-nm wide choke and 200-nm wide channel (long source and gate), at zero gate current. The gate is terminated at 50 Ω. The orange curve shows the I-V curve obtained with simulation of the SPICE model. Its parameter were set according to the geometry of the nTron.

The curve is hysteretic because the load of the nTron is 1 $M\Omega$, and the critical current is $I_{c,ch} = 71\ \mu A$. The orange curve was obtained by simulation of the LTspice model, in which all the geometrical parameters were set. The model approximately fits the experimental data. During the transaction from the superconducting to the normal state the curves are slightly different probably because the model does not consider the real meandered geometry of the drain, moreover the negative critical current in the experimental curve is a bit lower respect to the positive one.

After the realization of the I-V curves of other nTrons, here not shown, a DC current was applied to the gate with the same experimental setup. The gate critical current $I_{c,g}$ was found applying a current ramp to the gate and measuring for which value of current the gate switches. Table 5.1 shows the obtained critical currents for three different nTrons (with long gate and source nanowires):

	10-nm choke, 100-nm ch.	20-nm choke, 200-nm ch.	30-nm choke, 300-nm ch.
$I_{c,g}$ (μA)	3.5	4.8	11.5
$I_{c,ch}$ (μA)	66	71	181

Table 5.1: Gate critical current $I_{c,g}$ and channel critical current $I_{c,ch}$ for three different nTrons, measured at 4.2 K.

The gate critical current is proportional to the choke width as expected, but the proportionality factor is slightly different for each nTron. This means that there is not enough control on the gate dimensions. The same consideration is valid for the channel critical current.
The behavior of the nTron with a load on the order of 10 Ω was studied to analyze the configuration that could be used in the nTron synapse, and estimate the value of the parameter A_1, introduced in section 3.3. An external resistor (at room temperature) of 22 Ω was connected to the drain of the nTrons and V-I curves with DC gate current higher than zero were realized. Figure 5.10 shows the result for 10-nm and 20-nm nTrons.

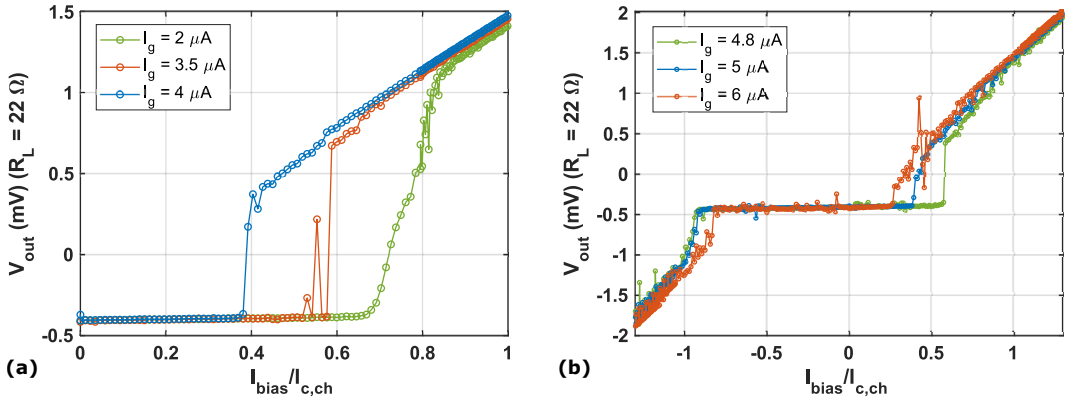

Figure 5.10: Voltage across the load resistance ($R_L = 22\ \Omega$) of the nTron as function of the bias current, at different values of gate current (DC). **(a)** nTron with 10-nm wide choke and 100-nm wide channel ($I_{c,ch} = 66\ \mu A$, $I_{c,g} = 3.5\ \mu A$). Only the result with positive bias currents is shown. **(b)** nTron with 20-nm wide choke and 200-nm wide channel ($I_{c,ch} = 71\ \mu A$, $I_{c,g} = 4.8\ \mu A$).

Figure 5.10a shows the positive part of the V-I curve obtained with the 10-nm nTron. The signal is very noisy because using external resistors at room temperature reflection signals are generated and the curve is slightly shifted along the y axis due to parasitic resistances introduced by the setup, but it is possible to notice that the nTron is correctly working since the critical current decreases with increasing gate current. At $I_g = I_{c,g} = 3.5\ \mu A$ the critical current is suppressed by ~ 0.55, which is the factor A_1 of the nTron.
Figure 5.10b shows the full V-I curve obtained with the 20-nm nTron. Even in this case the nTron is working but the switching seems to be asymmetric. With negative bias currents the suppression of $I_{c,ch}$ is much lower. Further work would be needed to understand how to solve this problem, that would not allow to bias the nTron synapse with negative currents to obtain the inhibitory control. Also for this nTron $A_1 \approx 0.55$ (from green curve). This result is consistent with the fact

that the two nTrons have almost same channel critical currents, so in theory also same channel width, even if designed to have it different.
With the obtained A_1, a synapse formed by these nTrons, with the considerations made in chapter 4, would have a resulting tunability of $\sim 60\%$ for single spikes and $\sim 14\%$ for multiple spikes. These values are not optimal, so further work is needed to increase A_1, changing the geometry of the nTron. First of all a nTron with 100-nm wide channel should be correctly fabricated, since the channel of the tested device is probably different from 100 nm.

Elimination of reflections

The experimental setup used for the previous results presents the problem of reflections, that alter the output signal. The reflections were even more evident in setups used to demonstrate the ability of the nTron to amplify current pulses, with time-domain characterizations. The reflected signals are generated by the coaxial cables that act as transmission lines, and are not matched with the input impedances of the devices. These signals can be almost removed placing the 100 $k\Omega$ resistors and the load resistors at cryogenic temperature in close proximity of the nTrons, avoiding to use coaxial cables between the resistors and the devices ports. For each signal that it is not constant in time, like the current provided to the gate, a 50 Ω off-shelf resistor was placed on the PCB, in parallel with the battery source in order to transfer the voltage to the chip being matched. After this resistor the 100 $k\Omega$ resistor was placed to set the current. For bias lines the 50 Ω resistor is not necessary, so in this case only 100 $k\Omega$ were used. Placing the resistors at cryogenic temperature also lowers the thermal noise. Figure 5.11 shows the realized PCB with the 1x1 cm^2 die glued in the center and the resistors soldered on the lateral pads. A 20-nm wide choke 200-nm wide channel nTron (with short gate and source) was wire bonded to the associated pads, to test the amplification ability with a load of 10 Ω.
Considering that R_{syn} of the inductive synapses were too high, a 8500-squares inductor was wire bonded to off-shelf resistances ($R_{syn,1} = 10$, $R_{syn,2} = 5$), in order to demonstrate that the kinetic inductor can be charged by a train of spikes.
Moreover a nTron synapse was realized, connecting the drain of a 20-nm nTron to an inductor with $R_{syn,1} = 10$ and $R_{syn,2} = 5$, to test the behavior of the ntron synapse. The obtained results are shown in the following sections

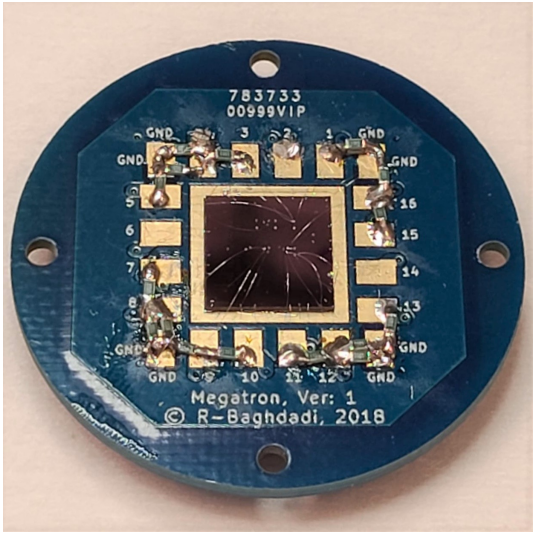

Figure 5.11: PCB with the die glued in the center and off-shelf resistors soldered on the pads. A nTron, a meandered inductor and a nTron synapse are wire bonded to the associated pins and resistors.

The I-V curve with variable gate current of the 20-nm nTron with resistors at cryogenic temperature ($R_L = 10\ \Omega$) is shown in figure 5.12.

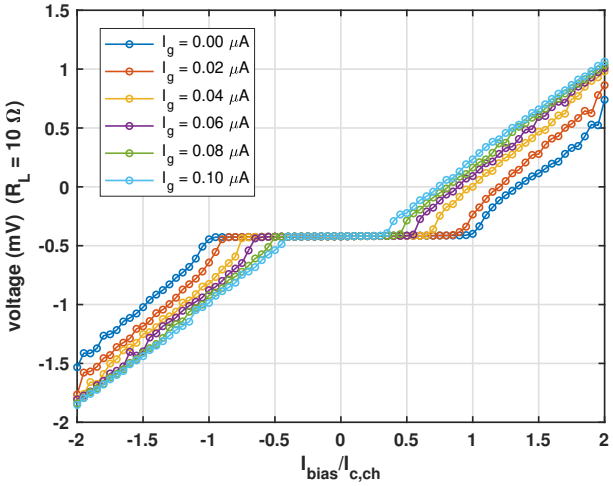

Figure 5.12: Voltage across the load resistance ($R_L = 10\ \Omega$) of the nTron as function of the bias current, at different values of gate current (DC). 20-nm wide choke and 200-nm wide channel, short gate and source (~ 20 squares), resistances at cryogenic temperature, $I_{c,ch} = 80\ \mu A$, $I_{c,g} = 20 - 50\ nA$.

Surprisingly the tested nTron seemed to have a very low gate critical current in the range 20-50 nA. Such a low value could be obtained with a real choke width of $\sim 1\ nm$. It is possible that the fabricated choke is on this order of magnitude, but is unlikely that the nTron would not be activated by noise with this $I_{c,g}$. Since the activation seems to be stable probably the 100 $k\Omega$ resistor changed is value at low temperature or it was made a mistake during the soldering and in reality a higher current was provided to the gate. Even if the value of critical current probably has to be scaled, figure 5.12 shows that this nTron with short gate and source (~ 20 squares) has a symmetric I-V curve, and the suppression of the critical current is dependent on the gate current. The transition from superconducting to normal state is more smooth than the one of figure 5.10, so the value of A_1 was extracted from the time-domain results.

Pulse amplification

Firstly the ability of the nTron to amplify a current pulses, crucial for the realization of the nTron synapse, was demonstrated using a 20-nm nTron (long source and gate) with resistors at room temperature. The setup of the measurement is shown in figure 5.13A ($R_L = 50\ \Omega$), while 5.13B shows the time-domain behavior. The train of pulses was generated by a waveform generator with a 100 $k\Omega$ resistor in series.

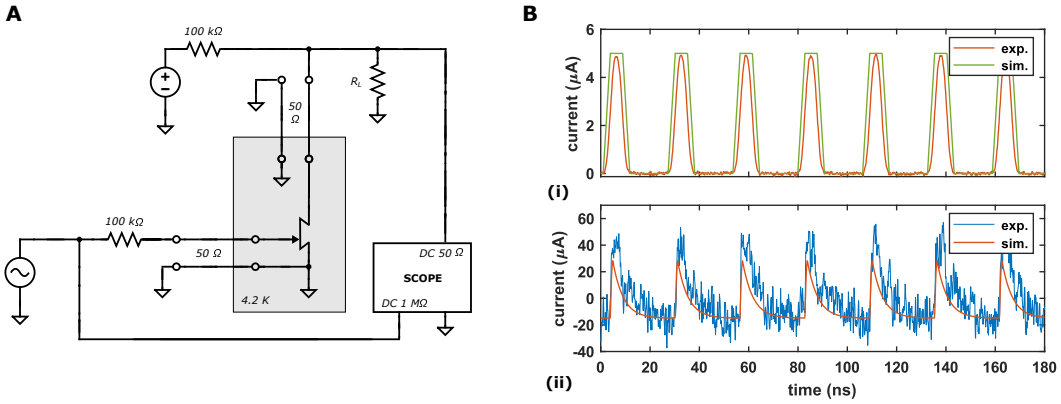

Figure 5.13: **(A)** Circuit schematic of the experimental setup used to amplify current pulses with a nTron. The resistors are not at cryogenic temperature. $R_L = 50\ \Omega$ but the input impedance of the oscilloscope is set to 50 Ω, so the total load resistance is 50 Ω. The input voltage is measured by the scope from a secondary channel of the waveform generator. **(B)** Time-domain behavior of the nTron with 20-nm wide choke and 200-nm wide channel (long source and gate). The nTron was simulated with the same setup (without coaxial cables), using the SPICE model with the same parameter of figure 5.9. **(i)** Current provided to the gate **(ii)** Current flowing through the load resistor, obtained dividing the output voltage by the load resistance.

The input signal does not present high noise because it was measured on a secondary port of the waveform generator. The SPICE model with the parameter used to match the I-V curve in figure 5.9, was used to simulate the same configuration of figure A (without including the coaxial cables). The simulated response seems to be coherent with the experimental result, even if reflections are present, so the model works properly. Due to reflections the maximum frequency of the current pulses provided to the gate, that can be followed by the nTron is only 38 MHz.

The same test was performed on the 20-nm nTron of figure 5.12 with the resistors soldered on the PCB. In this case the output signal was amplified by a

1GHz-bandwidth 50db amplifier. The obtained time-domain behavior is shown in figure 5.14

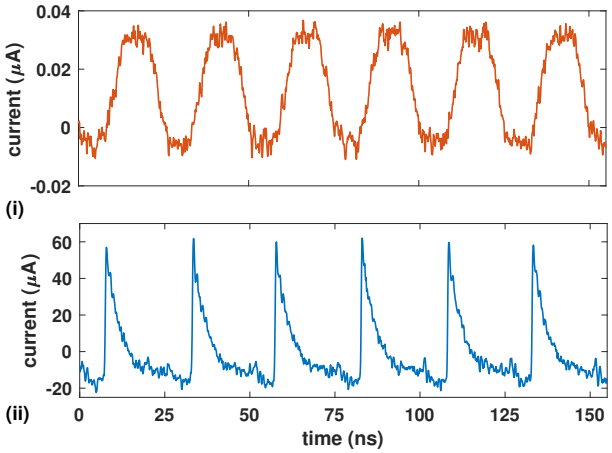

Figure 5.14: Experimental time-domain behavior of the nTron with 20-nm wide choke and 200-nm wide channel (short gate and source). The resistors are at cryogenic temperature, and the lines are matched ($I_{c,ch} = 80\ \mu A$, $I_{c,g} = 20 - 50\ nA$). **(i)** Current provided to the gate **(ii)** Current flowing through the load resistor ($R_L = 10\ \Omega$), obtained dividing the output voltage by the load resistance.

As expected the signal is much less noisy and similar to the one obtained by simulations. The exponential decay after the switching of the channel can be clearly recognized. The reflections are very limited and the nTron can work even at 80 MHz, the limit of the waveform generator. This nTron was tested also with negative values of bias current obtaining negative pulses on the output. This means that it can be used in a nTron synapse to set an inhibitory control. Since the signal is very clear, the result of figure 5.14 should be compared with the behavior of the SPICE model to eventually find corrections to it.

In order to find the value of A_1 for this last nTron, and realize a curve similar to the one obtained by simulations of the nTron synapse (without the inductor L_{syn}) of figure 4.6b, a sweep on the bias current of the nTron was performed obtaining the following output peak current as function of I_{bias} (the gate current is a train of pulses with high level of $I_{c,g} \approx 50\ nA$ at 80 MHz)

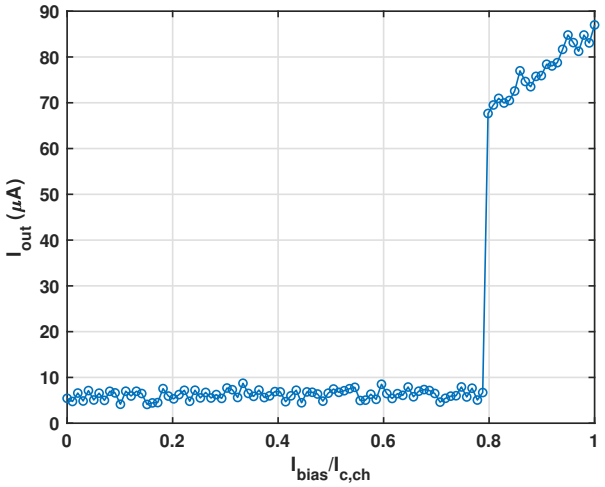

Figure 5.15: Output peak current through the 10 Ω load resistor of the 20-nm nTron ($I_{c,ch} = 80\ \mu A$, $I_{c,g} = 20 - 50\ nA$, resistors at cryogenic temperature) as function of the bias current. The peak current on the input is 50 nA, and the pulses frequency is 40 MHz. For each value of bias current the nTron is generating spikes, the peak current was extracted computing for each value of I_{bias} the maximum in the interval 0-180 ns.

Only positive bias currents were considered for this analysis. The obtained A_1 is ~ 0.8, which is too high to realize a nTron synapse with enough tunability.
The next step of this study would be to characterize all the fabricated nTrons with the method that exploits resistors soldered on the PCB. With more data it will be possible to understand better how to improve the geometry of the devices.

nTron + inductive synapse

The inductive synapse wire bonded on the chip shown in figure 5.11, was tested. A train of current pulses was provided to the input through the waveform generator as in the previous setup, and the voltage drop on the parallel resistance $R_{syn,2}$ was amplified and sensed by the oscilloscope. The goal was to show that a certain number of pulses could charge the inductor. The discharging current, that would have been the output current of the synapse if a NW neuron had been placed as load, could be extracted by the voltage drop on $R_{syn,2}$. Ideally increasing the frequency of the pulses, the DC signal on the output should have increased, but the tested device probably was shorted and the inductor did not charge. Previously the inductor was also tested with external resistances at room temperature, but the reflections were too high to allow a correct characterization. In the next steps the inductive synapses will need to be fabricated again using the correct fabrication

process, and more of them need to be characterized.

The final step necessary to demonstrate the nTron synapse behaves as designed, is to connect a nTron to an inductive synapse. This was done wire bonding the drain of a 20-nm nTron to the input of a inductive synapse. The structure was tested with the same method used for the inductive synapse, but the connected nTron did not work properly. The input pulses could pass through it as leakage current without being amplified. The same structure were previously tested with resistances at room temperature, but even in this case the reflections altered the response, and the inductor was not able to charge because the maximum frequency the nTron could tolerate was too low (8 MHz). A new fabrication has to be performed and more devices need to be characterized. The goal of this last measurement was to demonstrate the variable synaptic strength, the inhibitory control and find the real tunability of the nTron synapse.

Chapter 6

Integration of JJ and NW neurons: electrical analysis

The NW neurons are better than the JJ neurons in terms of fan-out because able to generate more fluxons. It is also simpler to test them, and integrate them with standard CMOS technologies, but they still can not reach the same level of parallelism of biological neurons. In a biological systems usually each neuron can connect to 1000s of neighbors [2]. In order to find a possible solution to this lack, the integration between nanowires and Josephson junctions-based technologies was analyzed by the electrical point of view. The JJ neurons have much lower characteristic impedance than NW neurons, and the NW neurons in general are able to provide more current to the target. The union of these two characteristics could be decisive to obtain an higher fan-out.

It was demonstrated by [43] that RSFQ circuits based on Josephson junctions can be interfaced with nanowire-based devices like nanocryotrons, and [7] proposed a complete design of a photonic spiking neural network in which JJ neurons are integrated with SNSPDs. Therefore NW neurons should be able to interface with Josephson junctions, which are the standard technologies for superconducting electronics. This section shows the results of the performed simulations to highlight pros and cons of the integration. Even if this integration did not result to be advantageous in term of fan-in/fan-out properties, it could be still useful exploiting NW neurons to allow the interface between JJs and CMOS.

6.1 Methods

The integration of the two technologies was tested only through electrical simulations, using the SPICE models explained in details in chapter 3. The complete LTspice model of the NW neurons was already used in [2] to demonstrate the basic

characteristics, while the model of the JJ neuron was only present in WRspice. Therefore the LTspice model of the JJ neuron was firstly tested, verifying that its time-domain behavior matches the results of [1].

The parameters of the two models were initially set respectively to the values used in both [1], and [2], and then slightly modified to find the optimal setting for the system. The NW neuron is a new technology. Few operating NW neurons were realized until now, and their characteristic and fabrication process are not well controlled and optimized yet, considering that the value of the shunt resistance results to be not well controlled. Here we supposed that the shunt resistance, critical current, inductances and other parameters are perfectly controllable for simplicity. A more in-depth analysis would be necessary to take into account any limitations introduced by the fabrication.

Parameters of JJ neurons

In this analysis only JJ neurons which belongs to Class I ($\Gamma = 1.5$) were taken in consideration. Their junctions are more damped than the ones of Class II, so it is easier to control their firing behavior, avoiding undesired unstable oscillations. The typical value of I_c for fabricated JJ neurons is 90 μA, and the junctions capacitance is linearly dependent on the critical current with a factor around 1.13 nF/A. It brings to a capacitance of 102 fF. The parallel of the shunt resistor and the subgap resistance (R_{tot}) was set to 4 Ω in order to obtain $\Gamma = 1.5$. The loop inductance $L_{tot} = (L_p + L_s)$ was set to values in the range from 50 to 100 pH.

In order to limit an unwanted activation of the neurons due to noise injected into the system, the bias current of both the devices was kept close to $1.9I_{sw}$ (or $1.9I_c$). In this way a 5% margin is maintained on both the oscillators (or junctions).

The axons of both JJ and NW neurons are not considered in all the simulations, moreover the configuration of inductive synapses that allows a variable synaptic strength and the inhibitory control through inductive coupling were not introduced for sake of simplicity.

6.2 Connections

In order to evaluate a possible integration of JJ neurons and NW neurons, different kinds of connections were simulated, in order to find the best configurations.

6.2.1 Direct connection

The easiest way to connect two neurons is through a resistor, so that the strength of the synaptic connection corresponds to the value of the resistance. Both NW-to-JJ and JJ-to-NW connections were analyzed.

NW-to-JJ connection

Considering that a single current pulse of the NW neuron is at least 100 times longer than a pulse of a JJ neuron and the typical spiking frequency of the NW neuron is at least one order of magnitude lower then the one of the JJ neuron, the NW neuron with $I_{sw} = 30\ \mu A$ can activate the JJ neuron, which fires each time the upstream neuron generates a spike. The direct connection through a resistor does not isolate the upstream neuron from the pulses generated by the downstream neuron, in fact the shape of the action potential is modified as shown in figure 6.1b, so the configuration was discarded.

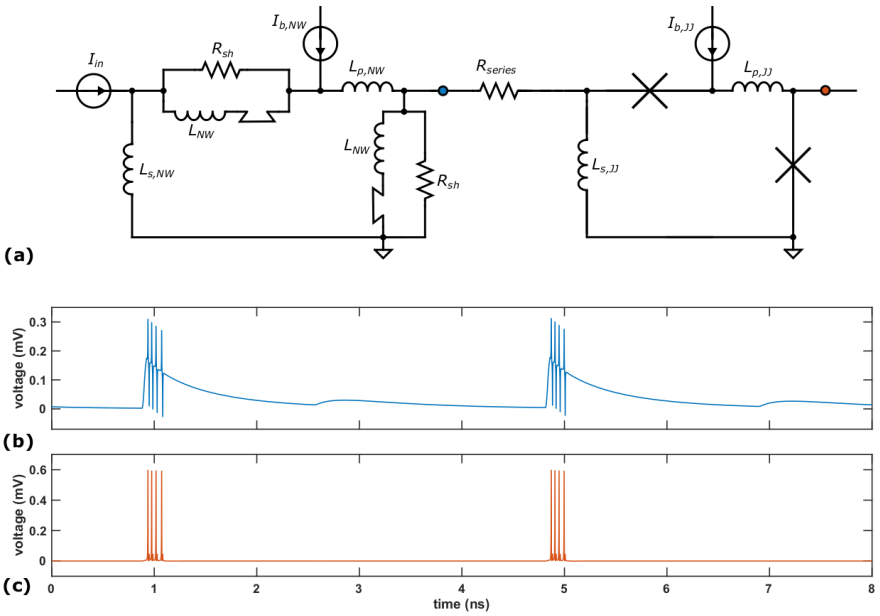

Figure 6.1: NW neuron connected to a JJ neuron through a resistor (direct connection). **(a)** Circuit schematic. Parameters: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 20\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{b,NW} = 57\ \mu A$, $R_{series} = 10\ \Omega$, $I_{b,JJ} = 171\ \mu A$, $I_{in} = 6\ \mu A$. **(b)** Time-domain of the output voltage of the NW neuron. **(c)** Time -domain of the output voltage of the JJ neuron.

JJ-to-NW connection

Obviously it is not possible to make a NW neuron fire with a single spike of current provided by a JJ neuron (JJ-to-NW connection) because a NW neuron needs more than one fluxon to be triggered, so the direct configuration is not suitable for this connection.

6.2.2 Connection through inductive synapse

The inductive synapse was chosen as preferable solution, because it can isolate the two neurons thanks to the presence of $R_{syn,1}$ and $R_{syn,2}$, and it ideally allows to exploit the non-linearity of the kinetic inductance to introduce the tunability of the synaptic strength, as explained in section 1.3.3.

NW-to-JJ connection

For the NW-to-JJ connection the kinetic inductance was set to values one or two orders of magnitude lower (1-10 nH) than ones used in [2], in order to maximize its output current and demonstrate that a NW neuron can make fire a JJ neuron with a single spike. With so low inductances the time constant of the synapse is not large enough to permit the input current pulses to charge the inductor with a train of spikes. This means that the synapse does not work as a standard inductive synapse, but more closely to a direct connection with isolation. The action potential directly propagates through it, its voltage and current levels are lowered, and its shape is broadened.
Figure 6.2 shows the dynamic in time-domain of the NW-to-JJ connection in both the excitatory and inhibitory behavior, for a single spike generated by the main neuron (in the inhibitory configuration a current source on the input of the JJ neuron is placed to activate it). In both cases the NW neuron is able to control the JJ neuron, which generates more spikes respect to the result of figure 6.1 since it is biased closer to $2I_c$. Using oscillators with higher I_{sw} would allow to decrease the bias current, obtaining the same level of firing. Moreover it is possible to notice that the spiking frequency of the JJ neuron is strongly dependent on the input current, clearly belonging to Class I.

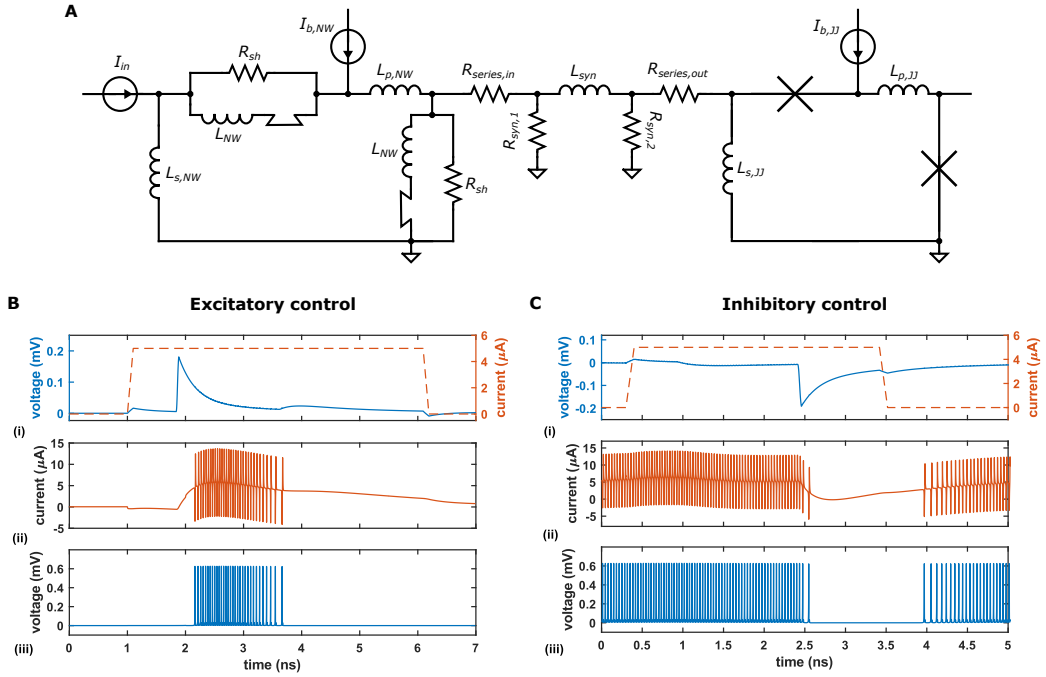

Figure 6.2: NW neuron driving a JJ neuron through inductive synapse. **(A)** Circuit schematic. Parameters of the NW neuron: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 10\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{in} = 6\ \mu A$ **(B)** Time-domain of the excitatory control. Parameters: $I_{b,NW} = 57\ \mu A$, $R_{series,in} = 5\ \Omega$, $L_{syn} = 5\ nH$, $R_{syn,1} = R_{syn,2} = 20\ \Omega$, $R_{series,out} = 1\ \Omega$. **(C)** Time-domain of the inhibitory control. Parameters: $I_{b,NW} = -57\ \mu A$, $R_{series,in} = 10\ \Omega$, $L_{syn} = 3\ nH$, $R_{syn,1} = R_{syn,2} = 20\ \Omega$, $R_{series,out} = 1\ \Omega$, $I_{in,JJ} = 6\ \mu A$. For both the types of control: inset (i) shows the input current and output voltage of the NW neuron; inset (ii) shows the output current of the synapse; inset (iii) shows the output voltage of the JJ neuron.

JJ-to-NW connection

Also for the JJ-to-NW connection, the reduction of the synaptic inductance is necessary to ensure enough output current, since the impedance of the JJ neuron is much lower than the one of the NW neuron and only a fluxon per spike is generated. In this case the pulses of the JJ neuron charges the small inductor of the synapse, that slowly discharges and activates the NW neuron.
Figure 6.3 shows the dynamic of this connection. The JJ neuron is able to drive a NW neuron both with excitatory and inhibitory control, firing for at least 5 ns. Increasing $I_{b,NW}$, less spikes would be necessary to activate the NW neuron. The inset (i) of figure 6.3B shows that the spiking frequency of the JJ neuron is strongly

dependent on the current passing though the synapse. It decreases from 25 GHz to 15 GHz, between 1 ns and 5 ns, and then rises again to 16 GHz when the NW neuron fires. This last change is due to the back-propagation of the signal and could be limited or increased, varying the synaptic inductance.

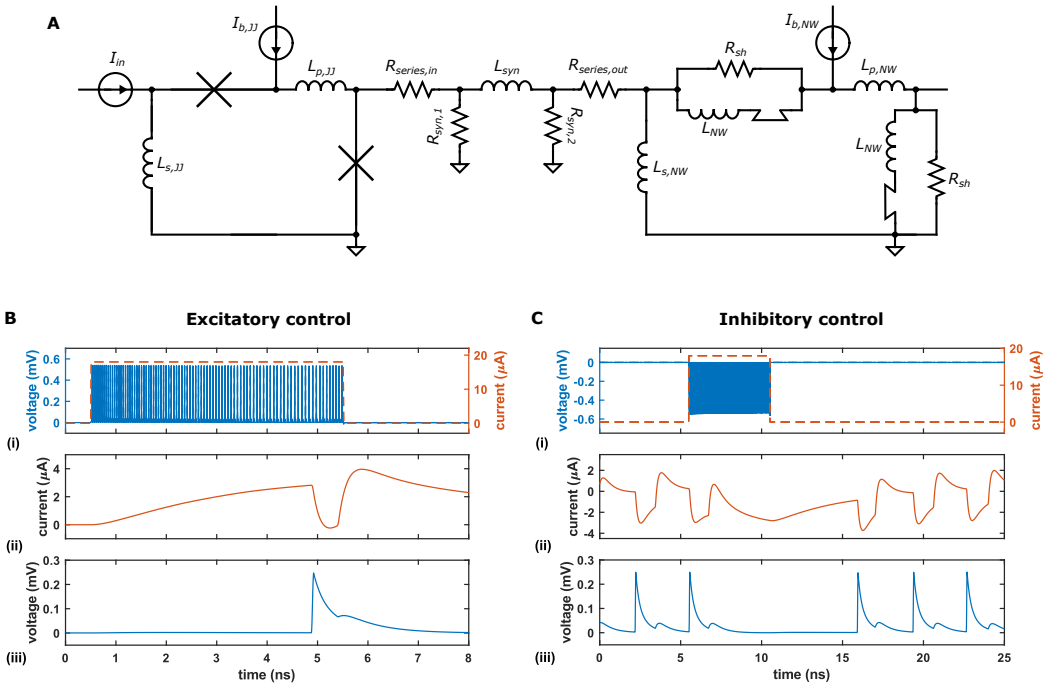

Figure 6.3: JJ neuron driving a NW neuron through inductive synapse. **(A)** Circuit schematic. Parameters of the NW neuron: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 10\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{b,NW} = 58.5\ \mu A$. Parameters of the JJ neuron: $L_{s,JJ,1} = L_{p,JJ,2} = 50\ pH$, $I_{in} = 20\ \mu A$ **(B)** Time-domain of the excitatory control. Parameters: $I_{b,JJ} = 171\ \mu A$, $R_{series,in,1} = 5\ \Omega$, $R_{syn1,1} = 20$, $L_{syn,1} = 20\ nH$, $R_{syn2,1} = 10$, $R_{series,out,1} = 1\ \Omega$. **(C)** Time-domain of the inhibitory control. Parameters: Parameters: $I_{b,JJ} = -171\ \mu A$, $R_{series,in,1} = 5\ \Omega$, $R_{syn1,1} = 20$, $L_{syn,1} = 20\ nH$, $R_{syn2,1} = 10$, $R_{series,out,1} = 3\ \Omega$, $I_{in,NW} = 6\ \mu A$. For both the types of control: panel (i) shows the input current and output voltage of the NW neuron; panel (ii) shows the output current of the synapse; panel (iii) shows the output voltage of the JJ neuron.

JJ-to-NW-to-JJ connection

In order to demonstrate that an hybrid system can be realized, it is necessary to show that the neuronal signal can propagate from JJ neurons to NW neurons and return again to the JJ neurons. Figure 6.4a and 6.4b show that it is possible, if

all the neurons are biased closer to $2I_c$ (or $2I_{sw}$). Despite this could be a problem in terms of noise margins, this simulation shows that the two technology can communicate. A single spike of the NW neuron can make the JJ neuron generate more spikes, that in turn can induce the NW neuron to fire once. Obviously some improvements have to be introduced to obtain a real integration and realize a large network.

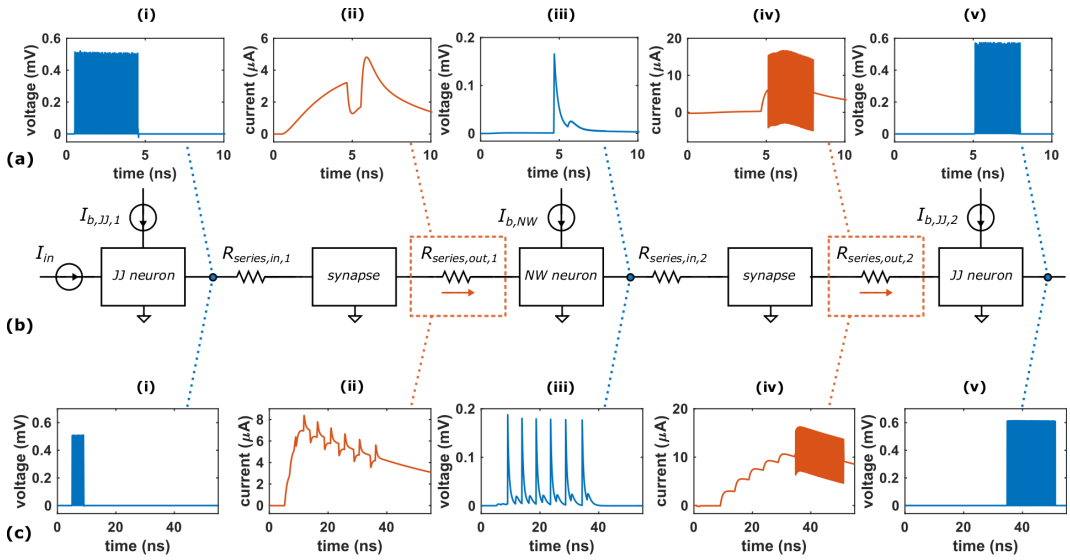

Figure 6.4: JJ-to-NW-to-JJ connection with inductive synapses and nTron synapses. **(a)** Time-domain with inductive synapses. First synapse parameters: $R_{series,in,1} = 5\ \Omega$, $R_{syn1,1} = 20$, $L_{syn,1} = 20\ nH$, $R_{syn2,1} = 10$, $R_{series,out,1} = 1\ \Omega$. Second synapse parameters: $R_{series,in,2} = 3\ \Omega$, $R_{syn1,1} = 20$, $L_{syn,1} = 5\ nH$, $R_{syn2,1} = 10$, $R_{series,out,1} = 1\ \Omega$. JJ parameters: $L_{s,JJ,2} = L_{p,JJ,2} = 100\ pH$. **(b)** Simplified circuit schematic. NW neuron parameters: $L_{s,NW} = L_{p,NW} = 10$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 10\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{b,NW} = 58.5\ \mu A$. JJ neuron parameters: $I_{b,JJ,1} = 171\ \mu A$, $L_{s,JJ,1} = L_{s,JJ,1} = 50\ pH$, $I_{b,JJ,2} = 175.5\ \mu A$, $I_{in} = 20\ \mu A$. **(c)** Time-domain with nTron synapses. Parameters for both synapses: $R_{sh,in} = 5\ \Omega$, $R_{syn1,1} = 15$, $L_{syn,1} = 100\ nH$, $R_{syn2,1} = 50\ \Omega$, $R_{series,in,1} = 10\ \Omega$, $R_{series,in,2} = 20\ \Omega$, $R_{series,out,1} = R_{series,out,2} = 2.5\ \Omega$, $I_{c,ch} = 40\ \mu A$, $I_{c,gate,1} = 2.2\ \mu A$, $I_{c,gate,2} = 3.7\ \mu A$, $I_{b,nT} = 36\ \mu A$. JJ neuron parameters: $L_{s,JJ,2} = L_{s,JJ,2} = 80\ pH$. For both inductive and nTron synapses the panels show the time-domain of: (i) output voltage of the first JJ neuron; (ii) current through $R_{series,out,1}$; (iii) output voltage of the NW neuron; (iv) current through $R_{series,out,2}$; (v) output voltage of the last JJ neuron.

The two technologies in the examples presented here, transfer information in two different way: the information encoded in a single spike of a NW neuron is translated into a train of spikes with variable frequency by the JJ neuron. It would

be preferable to have the same kind of codification for both the devices, but it is hard to obtain it, because of the enormous difference between their operating frequencies.
A more complete analysis would consist in the simulation of a closed loop formed by only one JJ neuron and one NW neuron, with the respective axons, that are connected through inductive synapses. The signal would travel along the loop, and the neurons would fire continuously with a certain phase difference from each other. This synchronization dynamic was studied on the picosecond time scale in coupled JJ neurons by [27].

6.2.3 Connection through nTron synapse

In chapter 4, the nTron synapse was introduced to connect NW neurons, but theoretically it could be used to connect also neurons based on different technologies. The NW-to-JJ connection is easily achievable, since the dependence of the nTron synapse on the load impedance is almost the same of a simple inductive synapse. For JJ-to-NW connections it is necessary to make some considerations. It has been demonstrated in [43] that the nTron can be triggered by a single pulse generated by a Josephson junction, but the time scale of the nTron pulse is much longer than the one of a JJ pulse. As consequence the nTron is not able to follow the fast firing of the Josephson junction, and a series of pulses could be seen as a single spike by the nTron.
Normally on the gate of the nTron there is a non-negligible kinetic inductance. If it is large enough (hundreds of pH), the spikes of the JJ neuron charge the input inductor with a time constant that depends on the shunt resistor of the gate ($R_{sh,in}$) and output impedance of the JJ neuron. When the current reaches the threshold the nTron switches and the inductor is quickly discharged. After the superconductivity is restored in the nTron, the gate is ready to switch again. The number of spikes needed to activate the nTron can be set by changing the value of the input inductance, so the number of squares of the gate and the source.

JJ-to-NW-to-JJ connection

The JJ-to-NW-to-JJ connection can be implemented also with nTron synapses. Figure 6.4c shows the time-domain behavior of the configuration, in order to compare it with the one based on inductive synapses. With the same number of spikes, the first JJ neuron through nTron synapse (with maximum $I_{b,nT}$) can make the NW neuron fire 6 times, charging the gate of the nTron multiple times. The NW neuron induces the last JJ neuron to fire for about 20 ns instead of 2.5 ns. All the neurons can be biased at current lower than $1.90I_c$, thanks to the amplification introduced by the nTron, they becomes more robust to external noise.

Moreover the nTron synapses always has the capability to store current in the inductive loop, while the simple inductive synapse can not for NW-to-JJ connections, due to the low value of L_{syn}. All the advantages of the nTron synapse could be exploited in a hybrid system, but additional power consumption would be added, and the real advantages of the integration deriving from the impedance mismatch between JJs and NWs would be lost (explained in following parts). For these reason next sections will try to analyze networks that does not exploit nTron synapses.

6.3 Fan-out/Fan-in limitations

The goal of this section is to study the fan-out/fan-in limitations of the hybrid system, as done in sections 4.2.2 for the NW-based system, to understand if it is possible to realize a large network with integrated JJ and NW neurons. The same notations and definitions of fan-out/fan-in are used here.

6.3.1 Fan-out

A system based on JJ neurons that does not exploits splitters (see section 4.2.3) or synapses able to amplify the signals, can only reach a fan-out of 2 or 3 elements [11], because the output flux of a single action potential is always equal to one fluxon Φ_0. In a system based on NW neurons that exploits inductive synapses the maximum fan-out is higher but still not comparable with the one of the brain. In order to show the advantages of the integration of nanowires and Josephson junctions for the fan-out, simulations on the following systems were performed: (1) one NW neuron driving M NW neurons; (2) one NW neuron driving M JJ neurons. For both these systems the inductive synapses are used instead of nTron synapses to remove every kind of signal amplification and make a fair comparison. The comparison with NW-based network that uses nTron synapses was included (without splitters). The fan-out limit of this system depends mainly only on the gate critical current of the nTron, so the results of section 4.2.2 should be valid also to describe one NW neuron that drives M JJ neurons though nTron synapses. The fan-out of a JJ neuron that drives M NW neurons was not analyzed because a JJ neuron is not able to provide enough current.

For systems (1) and (2) the synaptic inductances were kept as low as possible to maximize the output current of the synapses, but ensure a low cross-talk current, so an acceptable fan-in, in the hypothetical system with two layers introduced in section 4.2.2. L_{syn} of system (2) was chosen to be on the order of 1 nH. With the same value, in system (1) $Z_{out,syn}/Z_{in,n2}$ could be two orders of magnitude lower than the one of system (2), so it should be necessary to use a L_{syn} even two order of magnitude higher to obtain the same level of isolation (low cross-talk current). Considering that a so high inductance would lower the maximum fan-out to almost one, it was chosen to use 15 nH, a value that allows to have a two layer system with at least $N = 10$. These considerations already show the advantages of a possible integration of JJ and NW neurons: the fan-in connection from NWs to JJs is advantageous because the input impedance of a JJ neuron is much lower than the one of a NW neuron, so it is possible to drastically decrease the value of L_{syn} still keeping isolation from back-propagation, obtaining an higher fan-out because more current can be provided by the synapses. The second advantages of using NW neuron to drive JJ neurons is the possibility to tune $I_{sw,n1}$, a characteristic that is not present in JJ-based networks, which need splitters to have fan-out higher than 3.

For all the following analysis and simulations the parameters of the NW neurons do not follow table 4.1, so it was not simulated the decrease of spiking frequency with increasing switching current. In the model of the nanowires, only L_{NW} and I_{sw} differ from the default values and no other experimental parameters are set, so the neurons can still work at high I_{sw} without latching. This is not correct but considering that the synapses can not be charged by multiple spikes from NW neurons, the change of spiking frequency does not modify their behavior, so the results are consistent. It is important also to say that the output current of a NW main neuron with $R_{sh,main} = 50\ \Omega$ in these simulations is $\sim 0.7 I_{sw,n1}$ instead of $\sim 0.8 I_{sw,n1}$ (value obtained in section 4.2.2).

Figure 6.5 shows the structure of the hybrid system and the results of the time-domain simulation with $I_{sw,n1} = 200\ \mu A$ and $M = 11$.

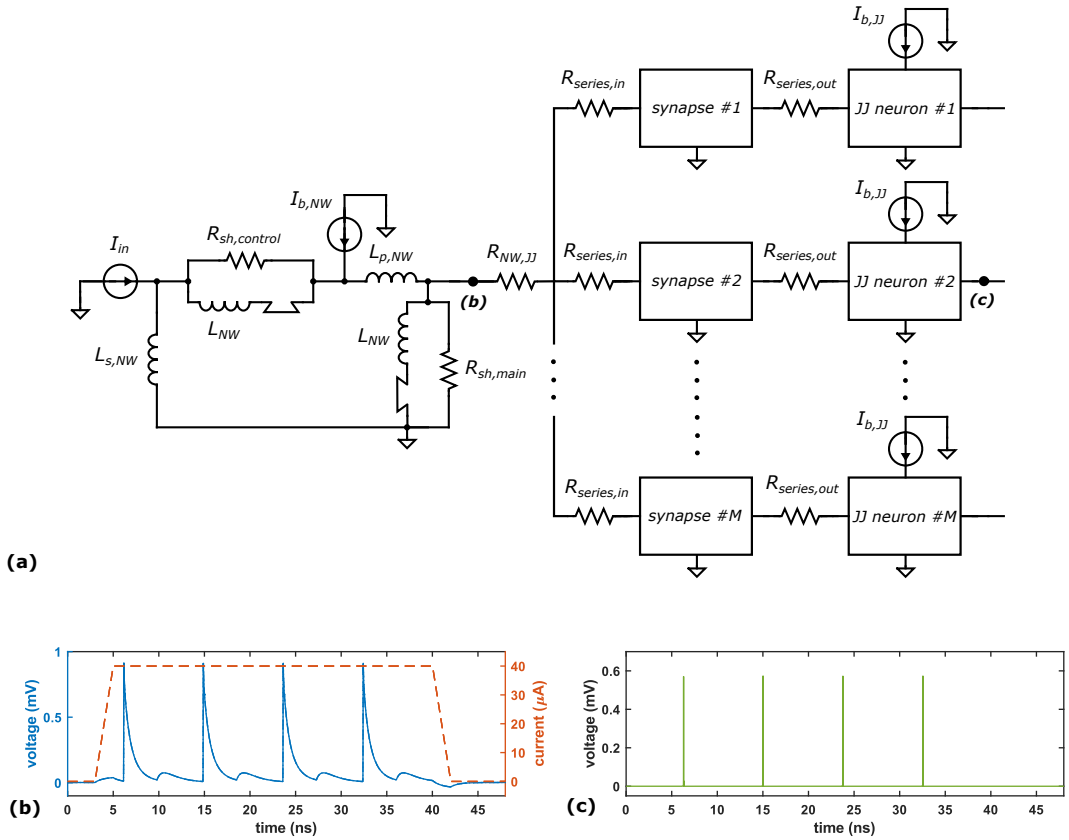

Figure 6.5: Maximum fan-out for the NW-JJ system. **(a)** Schematic of the simulated circuit. The parameters of the NW neuron do not follow the rules of table 4.1: only L_{nw} and I_{sw} are set in the model of the nanowires, all the other parameters are the default ones. Synapses parameters: $R_{series,in} = 4\ \Omega$, $R_{syn,1} = 30\ \Omega$, $L_{syn} = 1\ nH$, $R_{syn,2} = 30\ \Omega$, $R_{series,out} = 1\ \Omega$, $R_{NW,JJ} = 10\ \Omega$. **(b)** Time domain of the output voltage of the main NW neuron ($I_{sw,n1} = 200\ \mu A$, high $R_{sh,main} = 50\ \Omega$) that drives 11 JJ neurons ($I_{bias} = 1.9 I_c$) through inductive synapses.

Using equation 4.9 it is possible to estimate the maximum fan-out as function of the circuit parameters, with the structure of figure 6.5A. It was found by simulations that the junctions of a JJ neuron need at least $1.01 I_c$ of current to fire. This means that the input current of the JJ neuron has to be higher than $I^{min}_{in,JJ} = 2.02 I_c - I_{b,JJ}$. Considering that W is equal to 0.9 with the synaptic parameters used in figure 6.5,

and $I_{out,NW} = 0.7 I_{sw,NW}$, the expression of the maximum fan-out is:

$$FO_{max} \approx 0.31 \left(\frac{I_{sw,NW}}{I_{c,JJ}} \right) \frac{1}{1 - \left(\frac{I_{b,JJ}}{2.02 I_{c,JJ}} \right)} \tag{6.1}$$

In order to verify this equation is correct simulations were performed integrating LTspice and Matlab, as done in section 4.2.2. Figure 6.6 shows the maximum fan-out of system (2) as function of $I_{sw,NW}$ at different bias current of the JJ neurons.

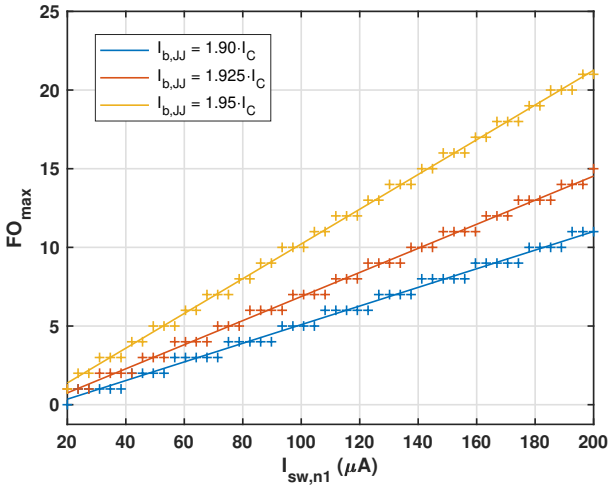

Figure 6.6: Maximum fan-out for a NW neuron that drives JJ neurons through inductive synapses as function of the switching current of the main neuron $I_{sw,n1}$, at different bias current of the JJ neurons $I_{b,JJ}$. The parameters of the NW neuron do not follow the rules of table 4.1: only L_{nw} and I_{sw} are set in the model of the nanowires, all the other parameters are the default ones; ($R_{sh,main} = 50\ \Omega$). Synapses parameters: $R_{series,in} = 4\ \Omega$, $R_{syn,1} = 30\ \Omega$, $L_{syn} = 1\ nH$, $R_{syn,2} = 30\ \Omega$, $R_{series,out} = 1\ \Omega$, $R_{NW,JJ} = 10\ \Omega$.

Equation 6.2 correctly describe the result. With $I_{b,JJ} = 1.9 I_c$ the maximum fan-out is:

$$FO_{max} \approx 5.25 \left(\frac{I_{sw,NW}}{I_{c,JJ}} \right) \tag{6.2}$$

Using parameters of table 4.1, FO_{max} would be 14% higher (with $I_{sw,NW} = 200\ \mu A$ it would be 12 instead of 11).

The same simulation was performed also for the system (1). Figure 6.7 shows the comparison between system (1), system (2) and the NW-based system with nTron synapses (in all the systems the target neurons are biased at $1.9I_c$ or $1.9I_{sw}$).

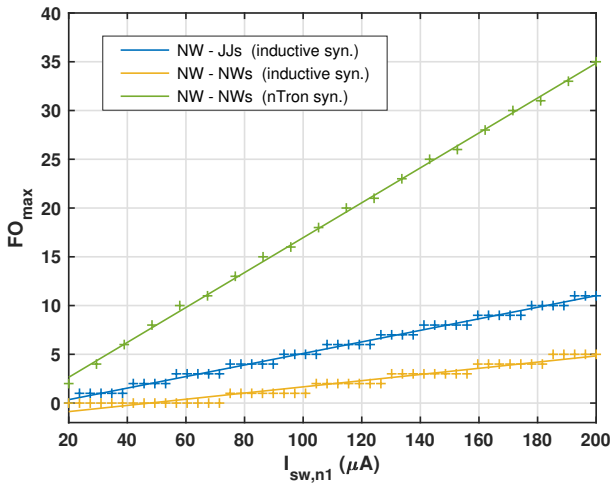

Figure 6.7: Maximum fan-out for different systems. All the target neurons are biased with $1.90I_c (or 1.90I_{sw})$. NW-JJs system: NW neuron driving JJ neurons through inductive synapses. Synapses parameters of figure 6.6. NW-NWs system: NW neuron driving NW neurons through inductive synapses. Synapses parameters of figure 6.6 with $L_{syn} = 15\ nH$. NW-NWs (nTron syn.) system: NW neuron driving NW neurons though nTron synapses. Parameters of figure 4.10 and table 4.1.

Obviously the nTron-based system is the most performing, but its power consumption is high. The hybrid system can reach a double fan-out with a better isolation from back-propagation respect to the NW-based system that does not exploit nTrons. Moreover here only JJ neurons with $I_c = 90\ \mu A$ were used, but it would be possible to exploit Josephson junctions with critical currents even lower than 20 μA, as done by [11]. With $I_c = 10\ \mu A$ the maximum fan-out could be higher than 100.

6.3.2 Fan-in

The fan-in from NW to JJ neurons was not analyzed because it can certainly reach high values, as already said. Here it was studied the fan-in from N JJ neurons to one NW neuron through inductive synapses. The circuit schematic of the fan-in structure is shown in figure 6.8.

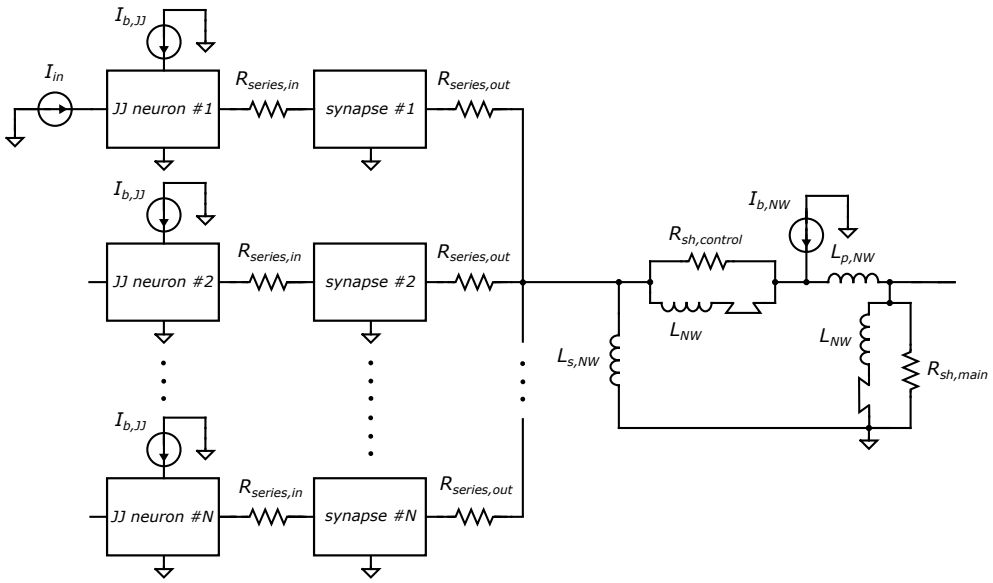

Figure 6.8: Schematic of the circuit simulated to study fan-in performances. N JJ neurons fan into one target NW neuron through inductive synapses, and only one of the N neurons is activated through I_{in}.

As explained in section 6.2.2, the JJ neuron can activate a NW neuron through inductive synapse only if a train of spikes charge the inductance L_{syn}, generating an almost DC current. The expression for the input current of the target neuron $I_{in,NW}$ is similar to the one of section 4.2.2:

$$I_{in,NW} \approx W I_{out,JJ} \frac{1}{1 + N\left(\frac{Z_{in,NW}}{Z_{out,syn}}\right)} \tag{6.3}$$

and even in this case, $WI_{out,JJ}$ depends on the number of spikes that charge the synapse. The parameters of the NW neuron were modified to decrease $Z_{in,NW}$: $L_{s,NW}$ and $L_{p,NW}$ were set to 5 nH, and the synaptic parameters are the ones of section 6.2.2 ($L_{syn} = 20\ nH$). The obtained input current as function of N, at different t_{on} (time during which the neuron continuously fires) of the JJ neuron is shown in figure 6.9.

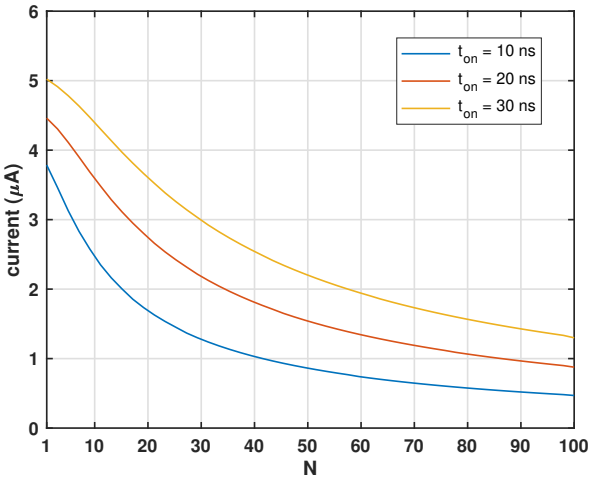

Figure 6.9: Input current $I_{in,NW}$ of the target neuron of figure 6.8, as function of the number of neurons in the first layer N, for different number of pulses generated by the NW neuron 1 (or different t_{on} of the pulse generated by I_in). NW neuron parameters: $L_{s,NW} = L_{p,NW} = 5$ nH, $L_{s,NW} = 4\ nH$, $R_{sh} = 15\ \Omega$, $I_{sw} = 30\ \mu A$, $I_{b,NW} = 0\ \mu A$. JJ neurons parameters: $I_{b,JJ} = 171\ \mu A$, $L_{s,JJ} = L_{s,JJ} = 50\ pH$, $I_{in} = 18\ \mu A$. Synapses parameters: $R_{series,in,1} = 5\ \Omega$, $R_{syn1,1} = 20$, $L_{syn,1} = 20\ nH$, $R_{syn2,1} = 10$, $R_{series,out,1} = 1\ \Omega$.

Fitting the three curves, it results that $Z_{out,syn}/Z_{in,NW}$ is ~ 14 for $t_{on} = 10\ ns$, ~ 25 for $t_{on} = 20\ ns$ and ~ 35 for $t_{on} = 30\ ns$. Using the same equation 4.16 of the NW-based system, and considering that the NW neuron is biased with $1.95 I_{sw,NW}$, otherwise it can not be activated, with $t_{on} = 30\ ns$ we obtain a maximum fan-in of:

$$FI_{max} = 0.39 \cdot \left(\frac{Z_{out,syn}}{Z_{in,NW}} \right) \approx 13 \tag{6.4}$$

which is very low compared to fan-in of NW-based and JJ-based systems. Moreover the bias current is higher so the noise can alter the operations. It is also important to say that making a JJ neuron constantly fire for more than 10 ns with a single action potential of a NW neuron is impossible, so in a real hybrid network it should be considered the case for which $t_{on} < 10\ ns$, obtaining a maximum fan-in close to 1.
Structures different from the one of figure 4.14 were exploited always trying to not increase the power consumption, introducing the inductive coupling to increase $Z_{out,syn}/Z_{in,NW}$, but they are discarded because they cause the decrease of the synapse output current. The alternative is to introduce additional energy consumption, amplifying the current in the synapses (for example using nTron synapses), or restoring the current at the input of the target neuron, with amplification. In [11]

the fan-in problem of the JJ-based system is solved with the inductive coupling, and a JTL that restores the current on the input of the target. Further work would be needed to improve the fan-in performances, probably trying to integrate the fan-in structure of [11] with a NW neuron. It is a goal difficult to achieve, because of the huge difference between the impedance of JJ neurons (order of pH) and the one of NW neurons (order of nH).

6.3.3 Generalized Fan-out in a complete JJ-NW neural network

A hypothetical 3-layers network formed by a first layer of NW neurons, a second one of JJ neuron and a last one of NW neurons, using only inductive synapses, would be useful for for application like pattern recognition. It was already demonstrated that NW neurons could be used to classify simple 9-pixels images [29], but the exploited structure can be formed by only two layers, so probably it would struggle to recognize patterns on larger images. Introducing an additional hidden layer the network would be more flexible and accurate. Thanks to the better fan-out performances of the hybrid system, it would be possible to realize it. Respect to a NW-based system the energy consumption would be lower. Moreover it would be better respect to JJ-based systems to interface the network with external environment because the input and output neurons would be NW neurons. The problem is that the low JJ-to-NW fan-in would limit the number of neuron that can be put on the last layer, and in addition it would not be possible to realize networks with more than 3 layers. For this reason, here the NW neurons of the last layers were replaced by JJ neurons. The fan-in of a JJ-based network that exploits inductive synapses can be much higher because the ratio $Z_{out,syn}/Z_{in,n2}$ is increased by two orders of magnitude. In this way we exploit both the advantages of the NW-to-JJs fan-out and JJs-to-JJ fan-in. The simplified structure that was analyzed here is shown in figure 6.10. For simplicity only one JJ neuron was put in the last layer. The drawback of this configuration obviously is that the output is generated by a JJ neuron, so it is more difficult to read it.

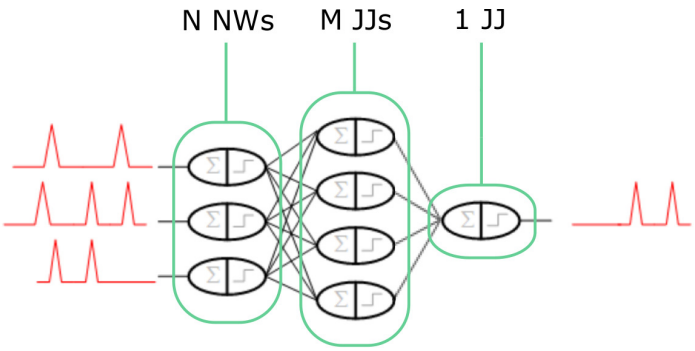

Figure 6.10: Simplified schematic of the analyzed network, with N NW neurons in the first layer, M JJ neurons in the hidden layer and 1 JJ neuron in the last one. The inductive synapses are not shown.

It is important to estimate how many NW neurons N_{on} are necessary to activate M JJ neurons in the second layer, to understand the real power of the hybrid system in terms of fan-out. Therefore the structure was simulated to find the generalized fan-out defined in section 4.2.2, and verify that the JJ neurons of the second layer can provide enough current to the output JJ neuron.

From equation 6.5 we know how to compute the generalized fan-out. Considering the value of $I_{in,n2}^{min}$, the generalized fan-out can be written as:

$$GFO_{max} = \left(\frac{M_{on}}{N_{on}}\right)_{max} \approx \overline{W}\left(\frac{I_{out,NW}}{2.02I_{c,JJ} - I_{b,JJ}}\right)\frac{1}{1 + N\left(\frac{Z_{in,JJ}}{Z_{out,syn}}\right)} \tag{6.5}$$

This is equal to the expression of the maximum fan-out multiplied by the factor that depends on $Z_{in,JJ}/Z_{out,syn}$, introduced to consider the fan-in from the first to the second layer. In the case of NW-to-JJ fan-in $N \ll Z_{out,syn}/Z_{in,JJ}$ so the expression of GFO_{max} results to be equal to the one of FO_{max}. In this section, in order to make a more realistic analysis it was supposed that W is equally distributed in the range [0-1] for all the synapses, so $\overline{W}$ is 0.5 (previously it was set to 0.9 to maximize the fan-out). Moreover it is known that $I_{out,NW} \approx 0.7I_{sw}$. With these assumptions we can say that:

$$GFO_{max} \approx 0.17\left(\frac{I_{sw,NW}}{I_{c,JJ}}\right)\frac{1}{1 - \left(\frac{I_{b,JJ}}{2.02I_{c,JJ}}\right)} \tag{6.6}$$

This expression shows that higher are the ratios $I_{sw,NW}/I_{c,JJ}$ and $I_{b,JJ}/2.02I_{c,JJ}$, higher is the number of JJ neurons that can be activated by N_{on} NW neurons

firing, in average. Choosing as bias current $I_{b,JJ} = 1.9I_c$, the final expression is:

$$GFO_{max} \approx 2.9 \left(\frac{I_{sw,NW}}{I_{c,JJ}} \right) \tag{6.7}$$

The network was simulated with $I_{sw,NW} = I_{c,JJ}$, $N = 50$ and variable M to validate this expression. Figure 6.11 shows the input current of the output JJ neuron as function of M/N_{on} at different value of N_{on}.

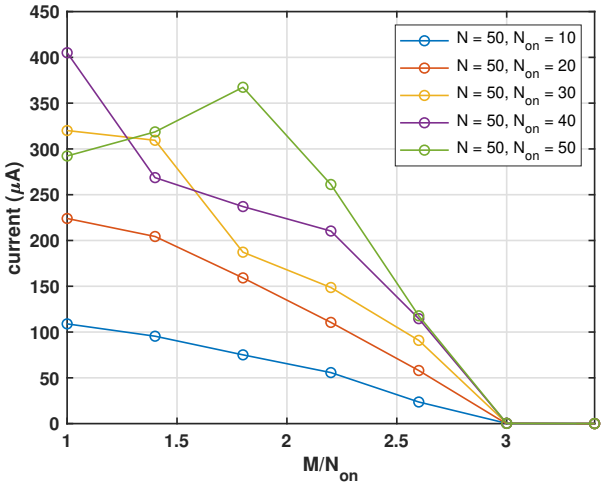

Figure 6.11: Input current of the JJ neuron of the last layer (figure 6.10), as function of M/N_{on} (number of JJ neuron in the second layer over number of firing NW neurons), at different value of N_{on}/N ($N = 50$). NW neurons parameters: $L_{s,NW} = L_{p,NW} = 5$ nH, $L_{s,NW} = 4\ nH$, $R_{sh,control} = 10\ \Omega$, $R_{sh,main} = 50\ \Omega$, $I_{sw} = 90\ \mu A$, $I_{b,NW} = 57\ \mu A$, $R_{NW,JJ} = 5\ \Omega$, $I_{in} = 6\ \mu A$. JJ neurons parameters: $I_{b,JJ} = 171\ \mu A$, $L_{s,JJ} = L_{s,JJ} = 50\ pH$. Synapses parameters (NW-to-JJ): $R_{series,in} = 4\ \Omega$, $R_{syn,1} = R_{syn,2} = 30$, $L_{syn} = 10\ nH$, $R_{series,out} = 1\ \Omega$. For the synapses between the second and last layers $L_{syn} = 1\ nH$.

L_{syn} of the synapses between the first and the second layer is set in order to have $W = 0.5$. All the N_{on} neurons are firing contemporary, so if they are able to activate JJ neurons in the second layer, they make them fire (generating more spikes) all a the same time. Therefore the amount of current provided to the output neuron is very high as expected. This means that it could be possible to add more JJ neurons to the last layer. Obviously for the same M, increasing the number of firing NW neurons , an higher current is generated, but in a real network it is very unlikely to have $N_{on} > 10$ for $N = 50$. In biological systems usually only $\sqrt{N}$ neurons fire. In the simulation for $N_{on} \geq 30$, at certain values of M/N_{on} lower than 2, the current seems to be lower. Probably it happens because too much

current is provided to JJ neurons. A more detailed analysis should be done to understand why. However this is a secondary problem, the graph firstly shows that the output current becomes zero approximately when $M/N_{on} = 3$. Considering that $I_{sw,NW} = I_{c,JJ}$, the GFO_{max} is ~ 3, respecting equation 6.7. This value can be increased with the same method used to boost the NW-to-JJs fan-out. These results show that it is possible to realize an hybrid network, in which the number of neurons in the hidden layer is higher of the number of neurons in the first one. In order to do a more detailed study, also the inductive coupling in the synapses should be introduced to reproduce the inhibitory behavior.

Chapter 7

Conclusion and outlook

Neuromorphic computing with alternative hardware structures have the potentiality to really revolutionize information processing for innovative applications like autonomous vehicle, robotics and internet of things (IoT). One of the most promising architectures is the artificial spiking neural network, which ideally could be able to reach the high performances of the brain. Many technologies has been proposed for the realization of SNNs, but the most efficient in terms of energy is based on superconducting electronics.
In this work, it was studied the possibility to create a SNN composed by superconducting nanowire-based devices, analyzing its limitations and introducing innovations in terms of design. All the most interesting results obtained in this thesis with associated future outlooks are listed below:

From chapter 3:

- For the first time, easily usable LTspice models of Josephson junction, nTron and hTron were created. The hTron model still need to be improved, however with these models and the integration between Matlab and LTspice, it was proposed a platform that facilitates the simulation of complex systems like large neural networks. In the future it can be used to design other innovative structures based on superconducting electronics.

From chapter 4:

- It was proposed and designed the inductive synapse with inductive coupling, which allowed the realization of a neural network for image recognition and a stochastic system based on the Winner-Takes-All (WTA) theory [29].
- The performances of the inductive synapse, firstly used for NW neurons, were improved introducing an innovative structure (nTron synapse) that exploits

the potentialities of the nTron (or hTron). The nTron synapse can reproduce both the inhibitory and excitatory behavior with variable strength, obtaining higher output current and tunability of the weight (ideally 120%) respect to standard inductive synapses. The level of tunability and other electrical characteristics of the nTron synapse still need to be improved eventually modifying the geometry of the nTron, and a complementary circuit should be included to introduce the unsurprising learning as done by [7].

- The fan-out limit of a network based on NW neurons and nTron synapses was analyzed through simulations finding a maximum fan-out of ~ 35, which can not be reached with standard inductive synapses, but it is still much lower than the values in the brain. Therefore it was proposed a structure inspired by a recent work [11], based on a new circuit named nTron splitter, that ideally could allow to reach fan-out of thousands. The nTron splitter was introduced here for the first time, so a more accurate analysis would be necessary to understand its limitations in terms of fabrication and occupied area. It would be also necessary to demonstrate experimentally that a single nTron can drive multiple nTrons.

- The fan-in limit was analyzed through simulations finding a maximum fan-in of ~ 56. This value is far from the one of biological systems, but it was observed that should be possible to easily increase it playing with all the parameters of the network. Probably it will be necessary to introduce a fan-in structure based on inductive coupling to further increase the maximum fan-in, as done by [11] and [7].

- Due to the characteristic non-linear input impedance and leakage currents of the nTron, the operation of a large neural network exploiting nTron synapses could be compromised for limited isolation. Therefore it was here proposed to replace it with the hTron synapse, for its better isolation and fan-out characteristics. It was demonstrated by simulations that it can couple two neurons but further work will be needed to understand if it can be realized and to find its limitations.

From chapter 5:

- Relaxation oscillators (shunted nanowire) with different values of shunt resistance were fabricated and characterized to study the spiking behavior, obtaining spiking frequencies even higher than 1.2 GHz. These kind of oscillators in the future could be exploited for applications different from SNNs. NW neurons were also tested, and resulted to not work properly with an input current pulse, so more devices need to be characterized to understand and solve the problem. Moreover it would be important to find a theoretical

description of the NW neuron, that would help the design, and the develop algorithms.

- nTrons and inductive synapses with different specifications were designed and fabricated in order to study experimentally the potentiality of the nTron synapse, and as final goal demonstrate the possibility to couple two neurons though it. Only some of the fabricated nTrons were tested obtaining an estimated value of tunability lower than expected. Therefore probably the geometry of the nTron has to be rethought. The next step would be to fabricate and test new nTrons, inductive synapses, and finally nTron synapses. The final and crucial step would be to connect two neurons.

From chapter 6:

- Here for the first time, a possible integration between JJ neurons and NW neurons was theorized and analyzed through simulations. It was demonstrated that the two technologies can communicate with both the nTron and inductive synapses, even if they work at very different spiking frequencies.

- The proposed hybrid system, that exploits inductive synapses, takes advantage of the high driving capability of NW neurons, and low input impedance of JJ neurons to obtain a maximum fan-out (from NW neuron to JJ neurons) higher than the one of JJ-based system without using JJ splitters and also higher than a NW-based systems with inductive synapses, still maintaining a low cross-talk current.

- The maximum fan-in from JJ neurons to NW neuron is very low, so an innovative circuital structure need to be found in order to make possible the realization of large networks with multiple alternate layers. Probably even in this case the solution would be a system based on inductive coupling.

- It was presented a hybrid system made by a first layer of NW neurons, a hidden layer of JJ neurons and the last one of JJ neurons, to exploit the good fan-out performances of the NW-to-JJ connection and the high fan-in achievable by a JJ-to-JJ connection. This network might be useful to increase the accuracy of pattern-recognition tasks (maintaining low power consumption) when the number of input pixels exceeds 9 [29] and the complexity of the images increases.

This work demonstrates that hypothetically it could be possible to realize an artificial spiking neural network based on superconducting nanowires with supervised learning, that can be integrated with CMOS technologies and Josephson junctions.

The fabrication variability and noise effects were not considered in the performed analysis and still a lot of work will be necessary to fabricate a real neural network. Moreover it would be very interesting to try to implement rate or temporal encoding of information in the system, to really exploit the advantages of spiking computation in innovative applications.

Bibliography

[1] Patrick Crotty, Dan Schult, and Ken Segall. «Josephson junction simulation of neurons». In: *Phys. Rev. E* 82 (1 July 2010), p. 011914. DOI: 10.1103/PhysRevE.82.011914. URL: https://link.aps.org/doi/10.1103/PhysRevE.82.011914 (cit. on pp. i, 4, 9, 15, 17, 22, 39, 100).

[2] Emily Toomey, Ken Segall, and Karl K. Berggren. «Design of a Power Efficient Artificial Neuron Using Superconducting Nanowires». In: *Frontiers in Neuroscience* 13 (2019), p. 933. ISSN: 1662-453X. DOI: 10.3389/fnins.2019.00933. URL: https://www.frontiersin.org/article/10.3389/fnins.2019.00933 (cit. on pp. i, 13–15, 17–21, 42, 52, 55, 99, 100, 102).

[3] Adam N. McCaughan and Karl K. Berggren. «A Superconducting-Nanowire Three-Terminal Electrothermal Device». In: *Nano Letters* 14.10 (2014). PMID: 25233488, pp. 5748–5753. DOI: 10.1021/nl502629x. eprint: https://doi.org/10.1021/nl502629x. URL: https://doi.org/10.1021/nl502629x (cit. on pp. i, 10, 27–30, 45, 54, 55, 58, 65).

[4] Reza Baghdadi et al. «Heater Nano-Cryotron: A Superconducting Nanowire-Based Thermal Switch». In: *Submitted* (2020) (cit. on pp. i, 10, 31–35, 48, 65, 75).

[5] von Neumann J. *The Computer and the Brain*. Yale University Press, 2012. URL: https://yalebooks.yale.edu/book/9780300181111/computer-and-brain (cit. on p. 1).

[6] Wulfram Gerstner, Werner M. Kistler, Richard Naud, and Liam Paninski. «Neuronal Dynamics». In: Cambridge Univ. Press, 2014. Chap. 1. URL: https://neuronaldynamics.epfl.ch/ (cit. on pp. 2–5, 15).

[7] Jeffrey M. Shainline et al. «Superconducting optoelectronic loop neurons». In: *Journal of Applied Physics* 126.4 (2019), p. 044902. DOI: 10.1063/1.5096403. eprint: https://doi.org/10.1063/1.5096403. URL: https://doi.org/10.1063/1.5096403 (cit. on pp. 2, 23, 24, 51, 56, 99, 120).

[8] Steve Furber and Steve Temple. «Neural Systems Engineering». In: *Computational Intelligence: A Compendium*. Ed. by John Fulcher and L. C. Jain. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 763–796. ISBN: 978-3-540-78293-3. DOI: 10.1007/978-3-540-78293-3_18. URL: https://doi.org/10.1007/978-3-540-78293-3_18 (cit. on p. 2).

[9] Emily Toomey. «Superconducting nanowire electronics for alternative computing». PhD thesis. Cambridge: Massachusetts Institute of Technology, May 2020. URL: https://www.rle.mit.edu/qnn/wp-content/uploads/2020/05/EToomey-PhD-EECS-May2020.pdf (cit. on pp. 4, 22, 65).

[10] Purves D, Augustine GJ, and Fitzpatrick D et al. *Neuroscience. 2nd edition*. Sunderland (MA): Sinauer Associates, 2001. URL: https://www.ncbi.nlm.nih.gov/books/NBK10799/ (cit. on p. 4).

[11] M. L. Schneider and K. Segall. *Fan-out and Fan-in properties of superconducting neuromorphic circuits*. 2020. arXiv: 2008.06409 [physics.app-ph] (cit. on pp. 6, 62, 71, 74, 107, 111, 113, 114, 120).

[12] I. M. Comsa, K. Potempa, L. Versari, T. Fischbacher, A. Gesmundo, and J. Alakuijala. «Temporal Coding in Spiking Neural Networks with Alpha Synaptic Function». In: *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 8529–8533 (cit. on p. 7).

[13] Si Wu, Shun-ichi Amari, and Hiroyuki Nakahara. «Population Coding and Decoding in a Neural Field: A Computational Study». In: *Neural Computation* 14.5 (2002), pp. 999–1026. DOI: 10.1162/089976602753633367. eprint: https://doi.org/10.1162/089976602753633367. URL: https://doi.org/10.1162/08997660275363336 (cit. on p. 7).

[14] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. «Theory of Superconductivity». In: *Phys. Rev.* 108 (5 Dec. 1957), pp. 1175–1204. DOI: 10.1103/PhysRev.108.1175. URL: https://link.aps.org/doi/10.1103/PhysRev.108.1175 (cit. on p. 7).

[15] V. Michal, E. Baggetta, M. Aurino, S. Bouat, and J. Villegier. «Superconducting RSFQ logic: Towards 100GHz digital electronics». In: *Proceedings of 21st International Conference Radioelektronika 2011*. 2011, pp. 1–8 (cit. on p. 8).

[16] B.D. Josephson. «Possible new effects in superconductive tunnelling». In: *Physics Letters* 1.7 (1962), pp. 251–253. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(62)91369-0. URL: http://www.sciencedirect.com/science/article/pii/0031916362913690 (cit. on p. 8).

[17] M. T. Levinsen, R. Y. Chiao, M. J. Feldman, and B. A. Tucker. «An inverse ac Josephson effect voltage standard». In: *Applied Physics Letters* 31.11, 776 (Dec. 1977), p. 776. DOI: 10.1063/1.89520 (cit. on p. 8).

[18] Emily Toomey. «Microwave Response of Nonlinear Oscillations in Resistively Shunted Superconducting Nanowires». MA thesis. Cambridge: Massachusetts Institute of Technology, 2017. URL: https://dspace.mit.edu/bitstream/handle/1721.1/113924/1023498717-MIT.pdf?sequence=1 (cit. on pp. 8, 10, 13, 42).

[19] R. Gross, A. Marx, and F. Deppe. *Applied Superconductivity: Josephson Effect and Superconducting Electronics*. De Gruyter Textbook Series. Walter De Gruyter Incorporated, 2016. ISBN: 9783110417067. URL: https://books.google.com/books?id=4SIzrgEACAAJ (cit. on p. 9).

[20] A. Vl. Gurevich and R. G. Mints. «Self-heating in normal metals and superconductors». In: *Rev. Mod. Phys.* 59 (4 Oct. 1987), pp. 941–999. DOI: 10.1103/RevModPhys.59.941. URL: https://link.aps.org/doi/10.1103/RevModPhys.59.941 (cit. on pp. 9, 27).

[21] G. N. Gol'tsman et al. «Picosecond superconducting single-photon optical detector». In: *Applied Physics Letters* 79.6 (2001), pp. 705–707. DOI: 10.1063/1.1388868. eprint: https://doi.org/10.1063/1.1388868. URL: https://doi.org/10.1063/1.1388868 (cit. on p. 10).

[22] Anthony J Annunziata, Daniel F Santavicca, Luigi Frunzio, Gianluigi Catelani, Michael J Rooks, Aviad Frydman, and Daniel E Prober. «Tunable superconducting nanoinductors». In: *Nanotechnology* 21.44 (Oct. 2010), p. 445202. DOI: 10.1088/0957-4484/21/44/445202. URL: https://doi.org/10.1088%2F0957-4484%2F21%2F44%2F445202 (cit. on p. 11).

[23] Karl K Berggren, Qing-Yuan Zhao, Nathnael Abebe, Minjie Chen, Prasana Ravindran, Adam McCaughan, and Joseph C Bardin. «A superconducting nanowire can be modeled by using SPICE». In: *Superconductor Science and Technology* 31.5 (Apr. 2018), p. 055010. DOI: 10.1088/1361-6668/aab149. URL: https://doi.org/10.1088%2F1361-6668%2Faab149 (cit. on pp. 11, 12, 25, 41).

[24] Andrew J. Kerman, Joel K. W. Yang, Richard J. Molnar, Eric A. Dauler, and Karl K. Berggren. «Electrothermal feedback in superconducting nanowire single-photon detectors». In: *Phys. Rev. B* 79 (10 Mar. 2009), p. 100509. DOI: 10.1103/PhysRevB.79.100509. URL: https://link.aps.org/doi/10.1103/PhysRevB.79.100509 (cit. on pp. 12, 13).

[25] Paul Bunyk, konstantin Likharev, and Dmitry Zinoviev. «RSFQ Technology: Physics and Devices». In: *International Journal of High Speed Electronics and Systems* 11.01 (2001), pp. 257–305. DOI: 10.1142/S012915640100085X. eprint: https://doi.org/10.1142/S012915640100085X. URL: https://doi.org/10.1142/S012915640100085X (cit. on pp. 15, 71).

[26] E. M. Izhikevich. «Which model to use for cortical spiking neurons?» In: *IEEE Transactions on Neural Networks* 15.5 (Sept. 2004), pp. 1063–1070. ISSN: 1941-0093. DOI: 10.1109/TNN.2004.832719 (cit. on p. 17).

[27] Patrick Crotty, Dan Schult, and Ken Segall. «Josephson junction simulation of neurons». In: *Phys. Rev. E* 82 (1 July 2010), p. 011914. DOI: 10.1103/PhysRevE.82.011914. URL: https://link.aps.org/doi/10.1103/PhysRevE.82.011914 (cit. on pp. 17, 23, 106).

[28] Geoffrey Hinton and Terrence J. Sejnowski. *Unsupervised Learning: Foundations of Neural Computation*. The MIT Press, May 1999. ISBN: 9780262288033. DOI: 10.7551/mitpress/7011.001.0001. URL: https://doi.org/10.7551/mitpress/7011.001.0001 (cit. on p. 20).

[29] E. Toomey, K. Segall, M. Castellani, M. Colangelo, N. Lynch, and K. K. Berggren. «Superconducting Nanowire Spiking Element for Neural Networks». In: *Nano Letters* 0.0 (0). PMID: 32965119, null. DOI: 10.1021/acs.nanolett.0c03057. eprint: https://doi.org/10.1021/acs.nanolett.0c03057. URL: https://doi.org/10.1021/acs.nanolett.0c03057 (cit. on pp. 22, 24, 58, 77, 78, 84, 86, 114, 119, 121).

[30] M. L. Schneider, C. A. Donnelly, S. E. Russek, B. Baek, M. R. Pufall, P. F. Hopkins, and W. H. Rippard. «Energy-efficient single-flux-quantum based neuromorphic computing». In: *2017 IEEE International Conference on Rebooting Computing (ICRC)*. Nov. 2017, pp. 1–4. DOI: 10.1109/ICRC.2017.8123634 (cit. on p. 22).

[31] Jonathan Binas, Ueli Rutishauser, Giacomo Indiveri, and Michael Pfeiffer. «Learning and stabilization of winner-take-all dynamics through interacting excitatory and inhibitory plasticity». In: *Frontiers in Computational Neuroscience* 8 (2014), p. 68. ISSN: 1662-5188. DOI: 10.3389/fncom.2014.00068. URL: https://www.frontiersin.org/article/10.3389/fncom.2014.00068 (cit. on p. 22).

[32] Michael L. Schneider, Christine A. Donnelly, Stephen E. Russek, Burm Baek, Matthew R. Pufall, Peter F. Hopkins, Paul D. Dresselhaus, Samuel P. Benz, and William H. Rippard. «Ultralow power artificial synapses using nanotextured magnetic Josephson junctions». In: *Science Advances* 4.1 (2018). DOI: 10.1126/sciadv.1701329. eprint: https://advances.sciencemag.

org/content/4/1/e1701329.full.pdf. URL: https://advances.sciencemag.org/content/4/1/e1701329 (cit. on pp. 23, 51).

[33] D. A. Buck. «The Cryotron-A Superconductive Computer Component». In: *Proceedings of the IRE* 44.4 (Apr. 1956), pp. 482–493. ISSN: 2162-6634. DOI: 10.1109/JRPROC.1956.274927 (cit. on p. 27).

[34] Adam Nykoruk McCaughan. «Superconducting thin film nanoelectronics». PhD thesis. Cambridge: Massachusetts Institute of Technology, Sept. 2015. URL: https://dspace.mit.edu/handle/1721.1/101576 (cit. on p. 28).

[35] Qing-Yuan Zhao, Emily A Toomey, Brenden A Butters, Adam N McCaughan, Andrew E Dane, Sae-Woo Nam, and Karl K Berggren. «A compact superconducting nanowire memory element operated by nanowire cryotrons». In: *Superconductor Science and Technology* 31.3 (Feb. 2018), p. 035009. DOI: 10.1088/1361-6668/aaa820. URL: https://doi.org/10.1088%2F1361-6668%2Faaa820 (cit. on p. 31).

[36] A. N. McCaughan, V. B. Verma, S. M. Buckley, J. P. Allmaras, A. G. Kozorezov, A. N. Tait, S. W. Nam, and J. M. Shainline. «A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors». In: *Nature Electronics* 2.10 (Oct. 2019), pp. 451–456. ISSN: 2520-1131. DOI: 10.1038/s41928-019-0300-8. URL: https://doi.org/10.1038/s41928-019-0300-8 (cit. on p. 31).

[37] Steven B. Kaplan. «Acoustic matching of superconducting films to substrates». In: *Journal of Low Temperature Physics* 37.3 (Nov. 1979), pp. 343–365. ISSN: 1573-7357. DOI: 10.1007/BF00119193. URL: https://doi.org/10.1007/BF00119193 (cit. on p. 32).

[38] M. Bhushan, J. Timmerwilke, D. Amparo, G. W. Gibson, and M. B. Ketchen. «Silicon Technology Inspired Test Structures and Methodology for SFQ Model-to-Hardware Correlation». In: *IEEE Transactions on Applied Superconductivity* 29.5 (Aug. 2019), pp. 1–7. ISSN: 1558-2515. DOI: 10.1109/TASC.2019.2898410 (cit. on pp. 37, 38).

[39] Sergey Tolpygo, Vladimir Bolkhovsky, Terence Weir, Alex Wynn, Daniel Oates, Leonard Johnson, and Mark Gouker. «Advanced Fabrication Processes for Superconducting Very Large Scale Integrated Circuits». In: *IEEE Transactions on Applied Superconductivity* (2016), pp. 1–1. ISSN: 1558-2515. DOI: 10.1109/tasc.2016.2519388. URL: http://dx.doi.org/10.1109/TASC.2016.2519388 (cit. on p. 38).

[40] *WRspice Reference Manual*. Consulative Committee for Space Data Systems (CCSDS). Sunnyvale, CA, 2020. URL: http://www.wrcad.com/manual/wrsmanual.pdf (cit. on p. 38).

[41] A. B. Cawthorne, C. B. Whan, and C. J. Lobb. «Influence and evaluation of parasitic inductance in shunted Josephson junctions». In: *IEEE Transactions on Applied Superconductivity* 7.2 (June 1997), pp. 2355–2358. ISSN: 1558-2515. DOI: 10.1109/77.621712 (cit. on p. 38).

[42] Nikhil Kumar, C B Winkelmann, Sourav Biswas, H Courtois, and Anjan K Gupta. «Controlling hysteresis in superconducting constrictions with a resistive shunt». In: *Superconductor Science and Technology* 28.7 (May 2015), p. 072003. DOI: 10.1088/0953-2048/28/7/072003. URL: https://doi.org/10.1088%2F0953-2048%2F28%2F7%2F072003 (cit. on p. 79).

[43] Qing-Yuan Zhao, Adam N McCaughan, Andrew E Dane, Karl K Berggren, and Thomas Ortlepp. «A nanocryotron comparator can connect single-flux-quantum circuits to conventional electronics». In: *Superconductor Science and Technology* 30.4 (Mar. 2017), p. 044002. DOI: 10.1088/1361-6668/aa5f33. URL: https://doi.org/10.1088%2F1361-6668%2Faa5f33 (cit. on pp. 99, 106).