Artificial neural networks and superconducting basic elements beyond von Neumann computer

Superconducting neural networks for broadband signal receiving and processing

Klenov N.V., Soloviev I.I., Schegolev A.E.,

Ruzhickiy V.I., Bakurskiy S.V., Kupriyanov M.Yu.

Lomonosov Moscow State University,

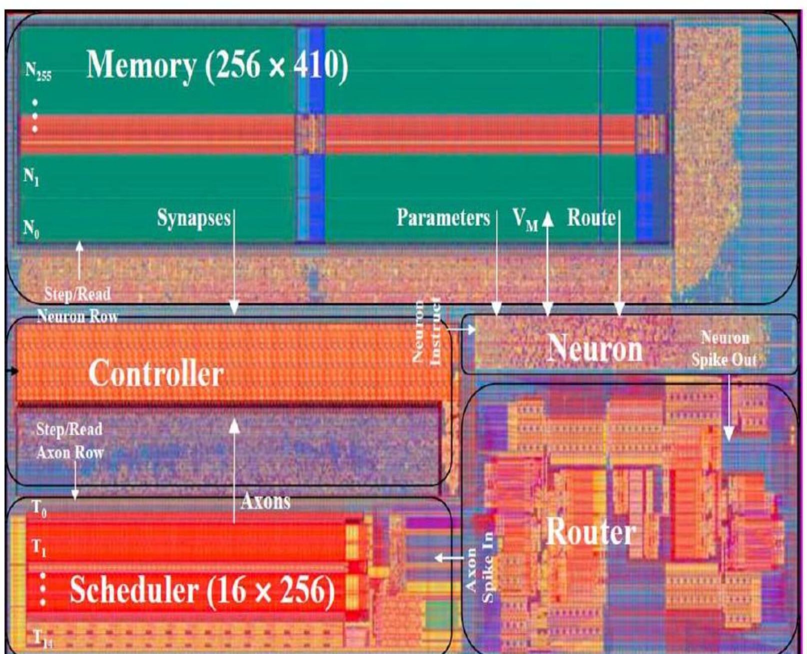
Dukhov All-Russia Research Institute of Automatics

State of the Art

Artificial Neural Networks

hardware — software

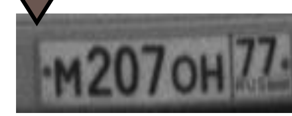
Synaptic chip IBM (TrueNorth)



ANN

1) clustering

2) recognition



As well as:

3) *forecasting*
4) *approximation*
5) *optimization*

Increase energy efficiency and speed!

Relevance

Adiabatic superconducting logic demonstrates the highest energy efficiency

Orbital satellite communication system

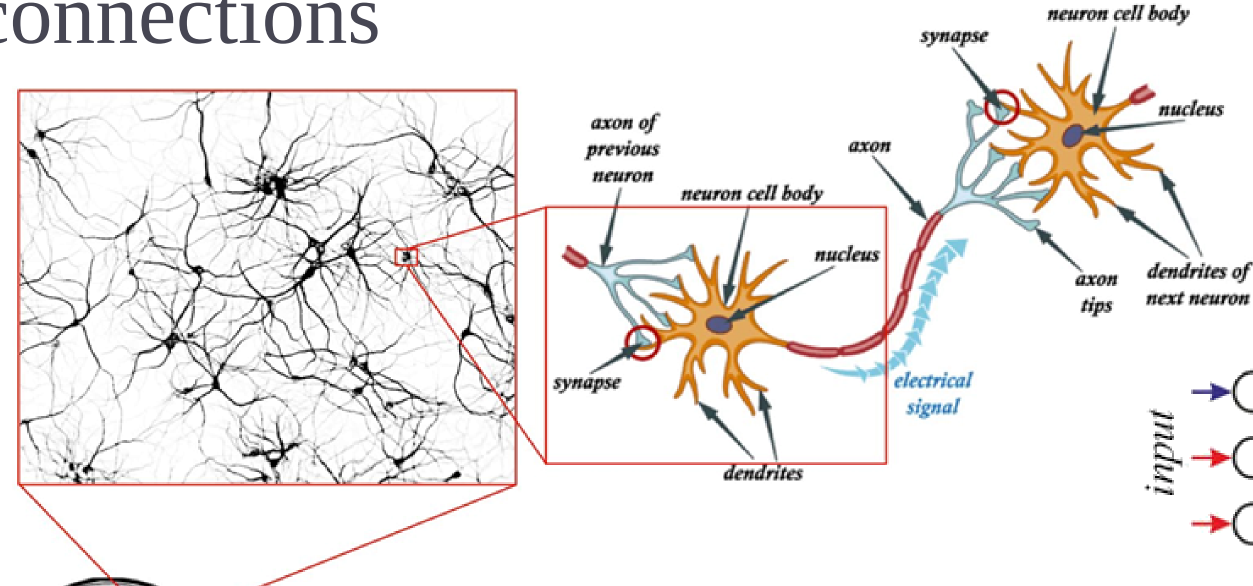
Reception and processing of weak broadband signals

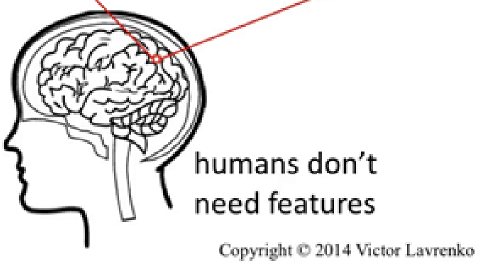
An adiabatic superconducting processor was created that consumes 15 fJ at clock speeds of 5 GHz

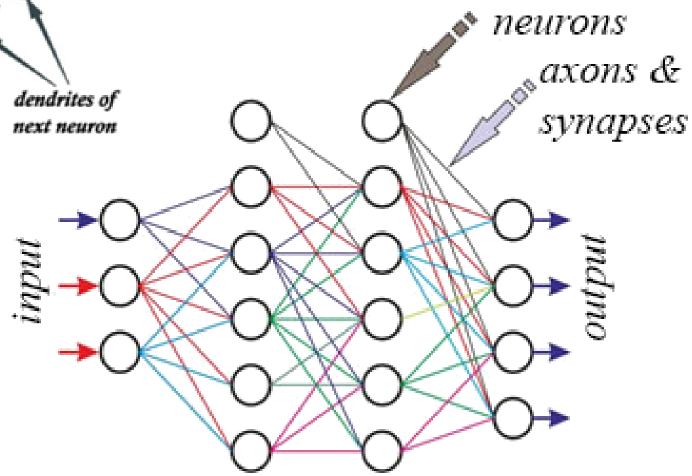
Deep space exploration

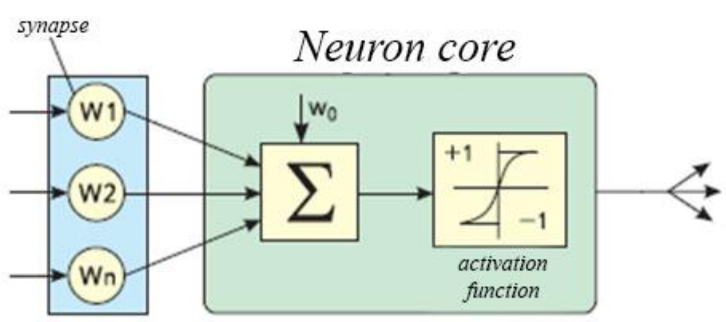
Physical problem statement:investigation of the adiabatic evolution of systems based on the S-quantron for the construction of an energy-efficient element base of the SNN.

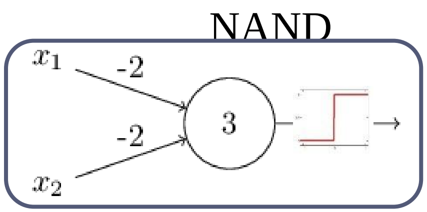
Subject: physical implementation of neurons and connections











analog

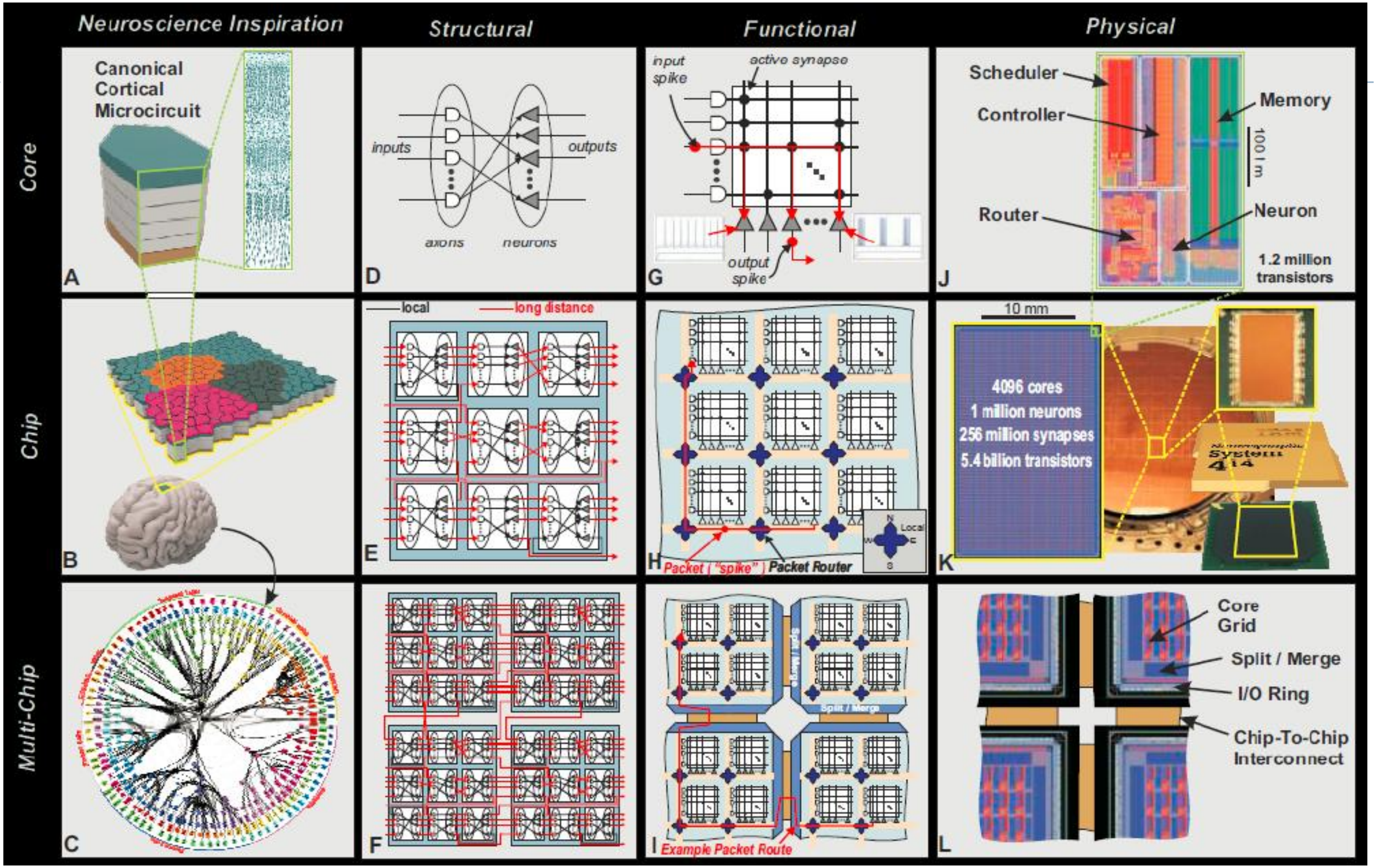
IBM TrueNorth (1M neurons, 256M synapses)
Intel Loihi (130K neurons, 130M synapses)

digital

Google TPU (105 x 4 = 420 TFLOS)
Huawei Ascend 910 (256 TFLOPS [FP16], 512 TOPS (INT8))

IBM TrueNorth

- Science 345, 668 (2014)
- Front Neurosci. 9, 141 (2015)

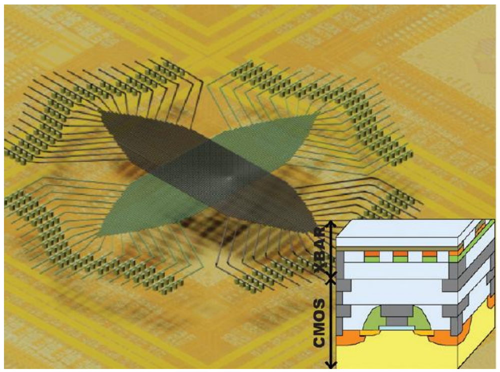


1M neurons
256M synapses, ~ 400 Mbit SRAM
46 x 10^9 OPS (70 - 100 mW)

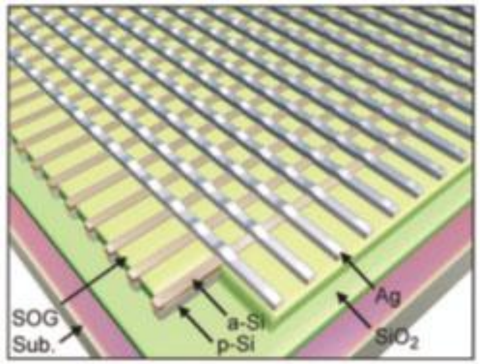
4096 cells (64 x 64)
100 Kbit SRAM for synapse memory (2 bits/synapse)
256 neurons (f_c = 1 KGz)

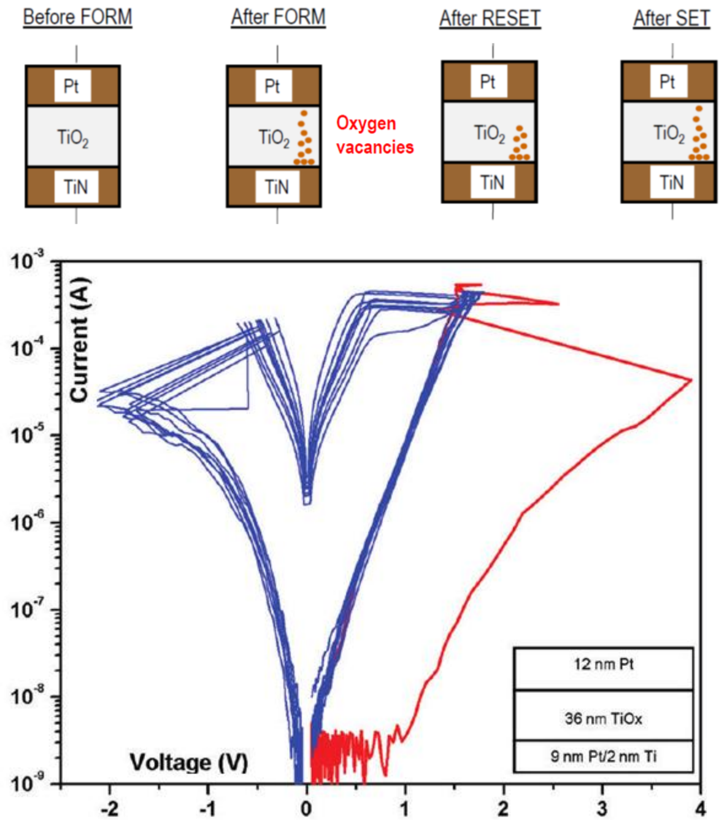
Memristors

CMOS + Memristors



top grid

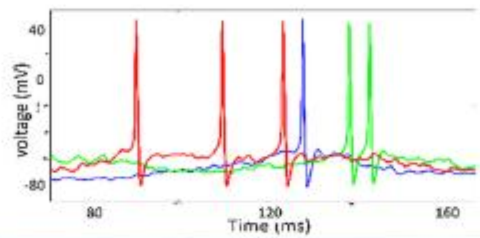
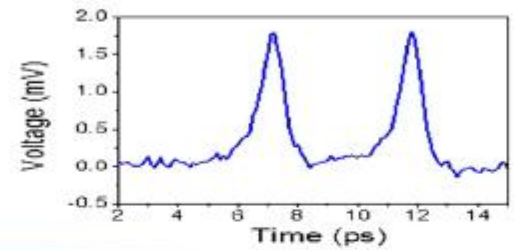
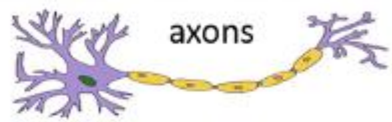
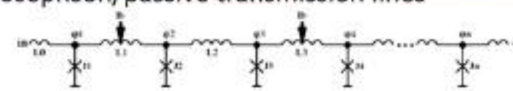
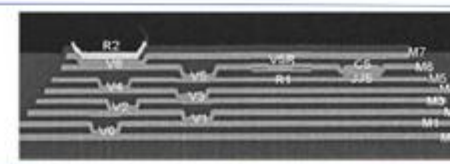
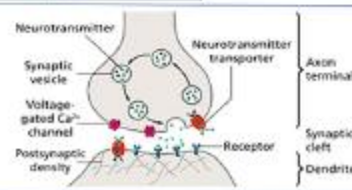
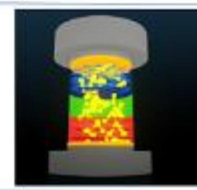




	ITRI, IEDM 2008	NEC, VLSI 2010	Panasonic, IEDM 2008	Univ. + IMEC, IMW 2010	Fujitsu, IEDM 2007
Device	TiN/Ti/HfO_x/TiN	Ru/TiO_x/TaO_x/Ru	Pt/TaO_x/Pt	Au/NiO_x/TiN	Pt/Ti-doped NiO/Pt
Reset	2V, 25uA	0.65V, 200uA	1.5V, 100uA	0.5V DC, 9.5uA	1.9V, 100uA
Set	2.3V	2.8V	2V	2.7V DC	2.8V
Form Voltage	3V	?	?	3.7V DC	3V
Switching Time	<10ns	<1us	<100ns	NA	10ns

P ~ 10^{-4} W

ANN implementation

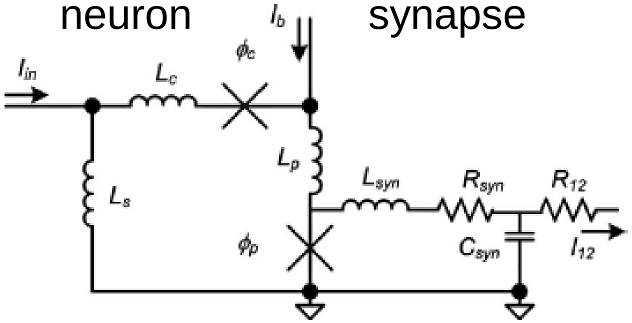
	Neural	*Single Flux Quantum*
a) Information transfer: quantized pulse trains		
b) Long distance "lossless" pulse transmission		Josephson/passive transmission lines 
c) 3D architecture	 White matter neural tracks	
d) Memory/ plasticity	 synapse	 magnetic Josephson junction
e) Number of neurons/fanout	10^{11} /1000	10^{7} / 100
f) Speed (pulses/neuron/s)	10^{3}	10^{10}
g) Power	20 W	0.1 mW

- M.L. Schneider et al., J. Appl. Phys. 124, 161102 (2018)

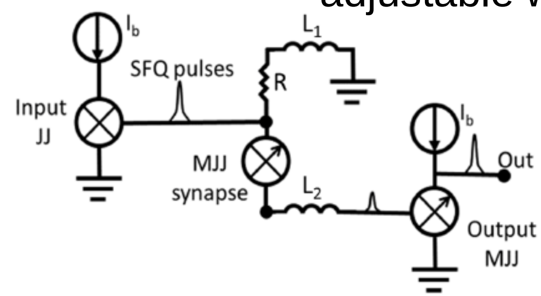
SFQ-based ANNs

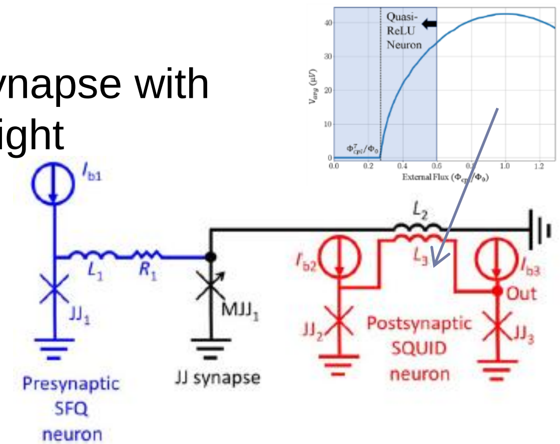
artificial neuron

artificial synapse

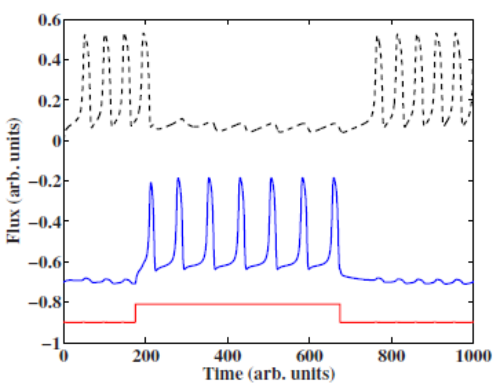


MJJ based artificial synapse with adjustable weight

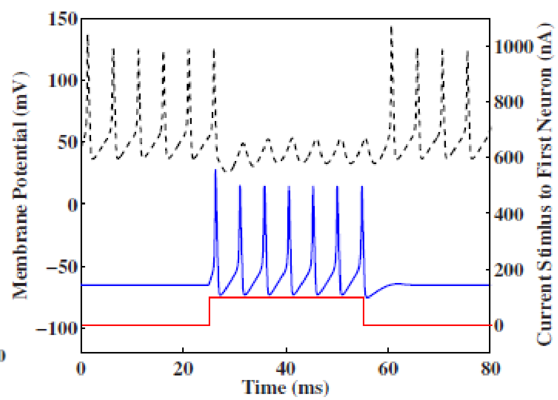




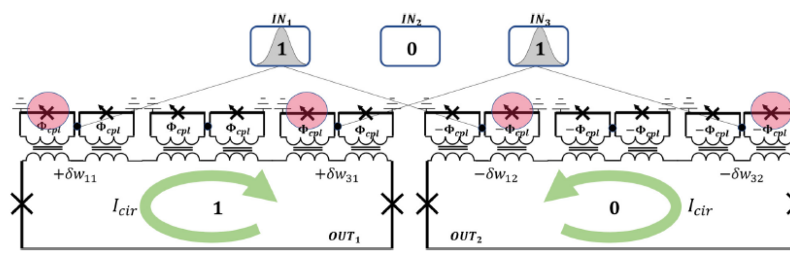
Inhibitory synaptic coupling of two SFQ-based neurons



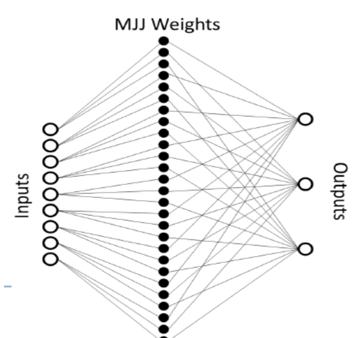
Behavior of two Hodgkin-Huxley neurons coupled with an inhibitory synapse model

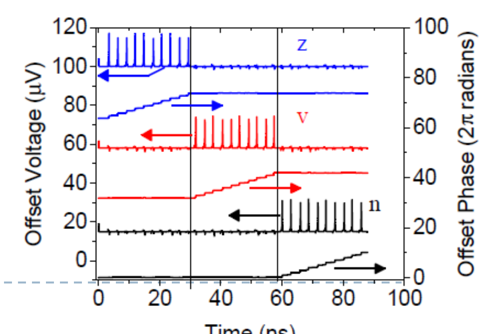


Synaptic connections with neurons



Example of operation

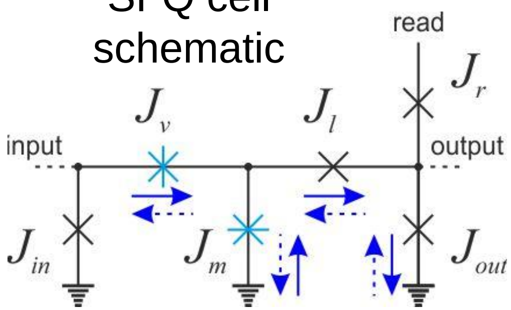




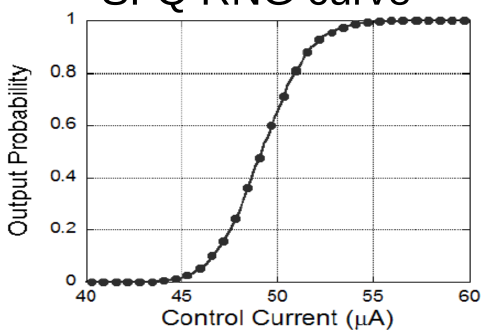
- P. Crotty et al., Phys. Rev. E 82, 011914 (2010)
- M.L. Schneider et al., IEEE ICRC, 17413500 (2017)
- M.L. Schneider et al., Sci. Adv. 4, e1701329 (2018)
- M.L. Schneider et al., J. Appl. Phys. 124, 161102 (2018)

Superconducting tensor processor unit

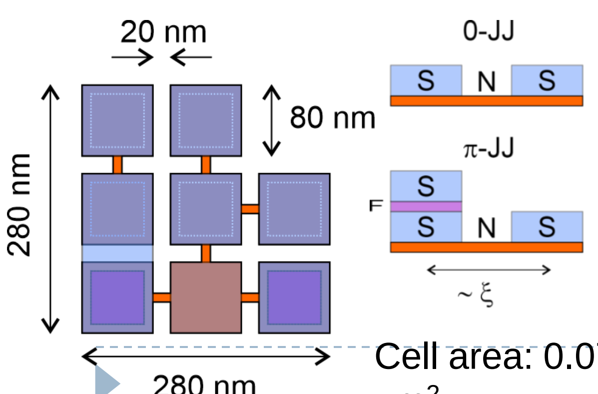
SFQ cell schematic



SFQ RNG curve



SFQ cell topology



Cell area: 0.078 μm^2

<u>Estimations (SC TPU):</u>

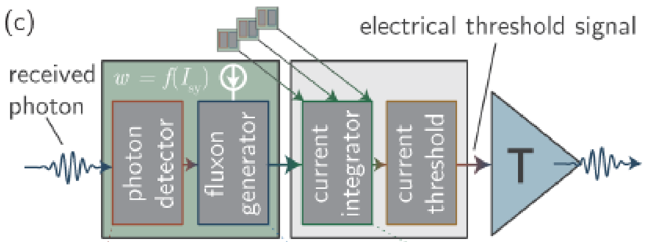
- Stream length = 256
- Cell area with = 0.3x0.3 ~ 0.1 μ^2
- ALU area = 2x256x0.1 ~ 50 μ^2
- Chip area = 1x1 cm^2 = 10^8 μ^2
- 2x2x128x128 ~ 65x10^3 ALUs/chip
- Available memory ~ 10^8/0.1 = 1 Gbit/layer
- f_c = 40 GHz (t_c = 25 ps)
- E_{bit} = 2 aJ (E_{op} = 2x256 = 512 aJ): **E_{SCE}/E_{CMOS} ~ 10^{-3}**
- 1chip = 2 active layers / 2 cores / 2 systolic arrays 128x128 ALUs
- 2x2x2x128x128x2x40x10^9 = **10.5 POPS** (~ 40 POPS four-chip module): **Th_{SCE}/Th_{CMOS} ~ 10^2**
- 25 SC TPU ~ **1 EOPS**

✓ High clock frequency and 3D architecture allow *2 orders of magnitude performance improvement*

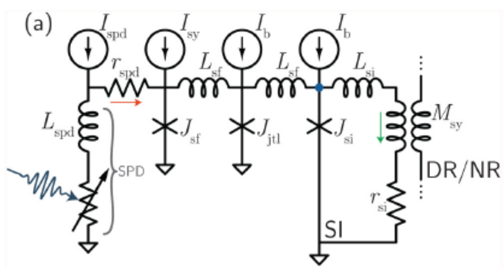
✓ Low energy dissipation allows *an order of magnitude improvement in energy efficiency* (including a cooling penalty)

Superconducting optoelectronic ANN

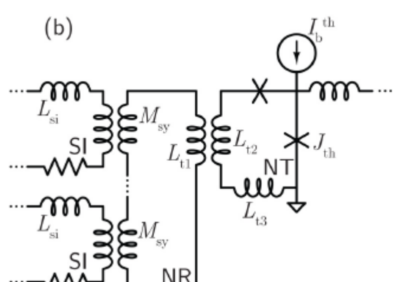
signal processing



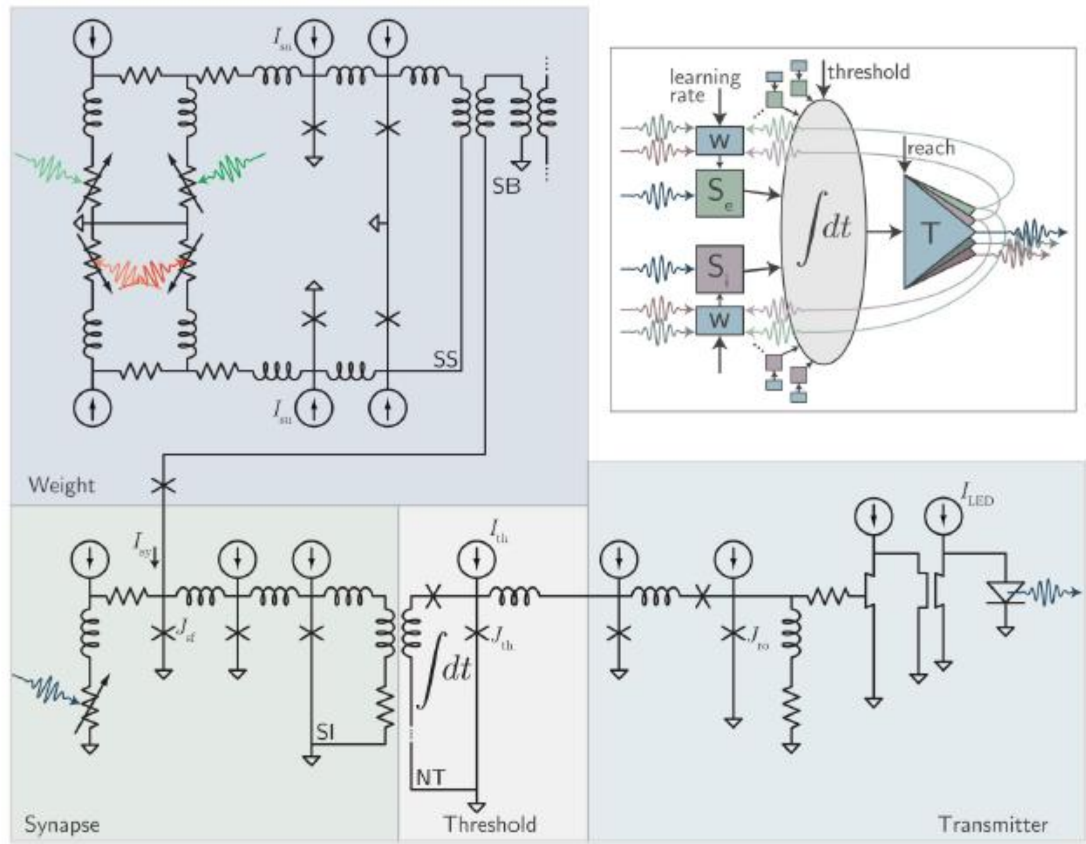
synaptic connection



integration and threshold circuits



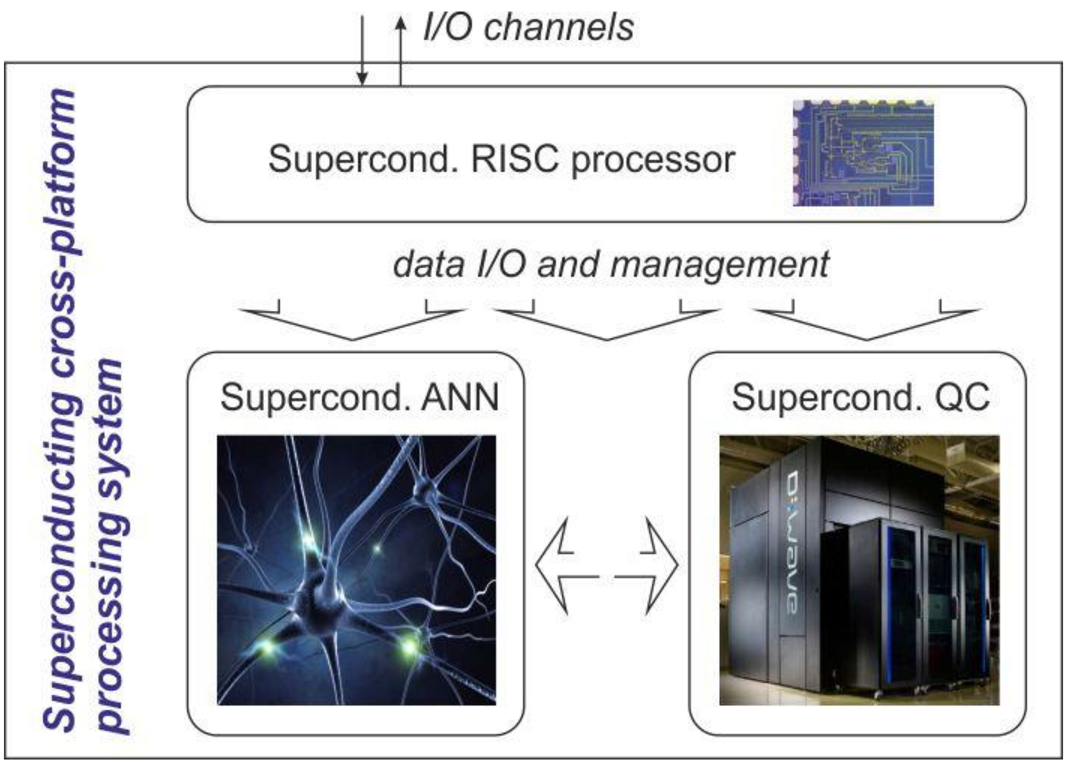
neural block with synaptic weight update



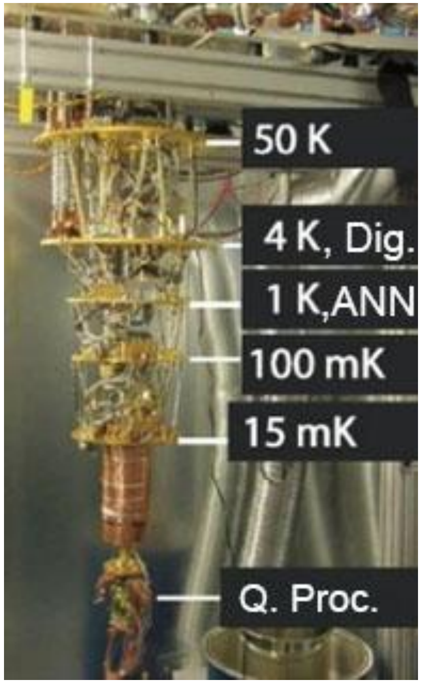
- J. M. Shainline et al., J. Appl. Phys. 124, 152130 (2018)
- J. M. Shainline et al., J. Appl. Phys. 126, 044902 (2019)

Superconducting "hybrid" solutions

Fast and energy-efficient platform for quantum and neural network computing

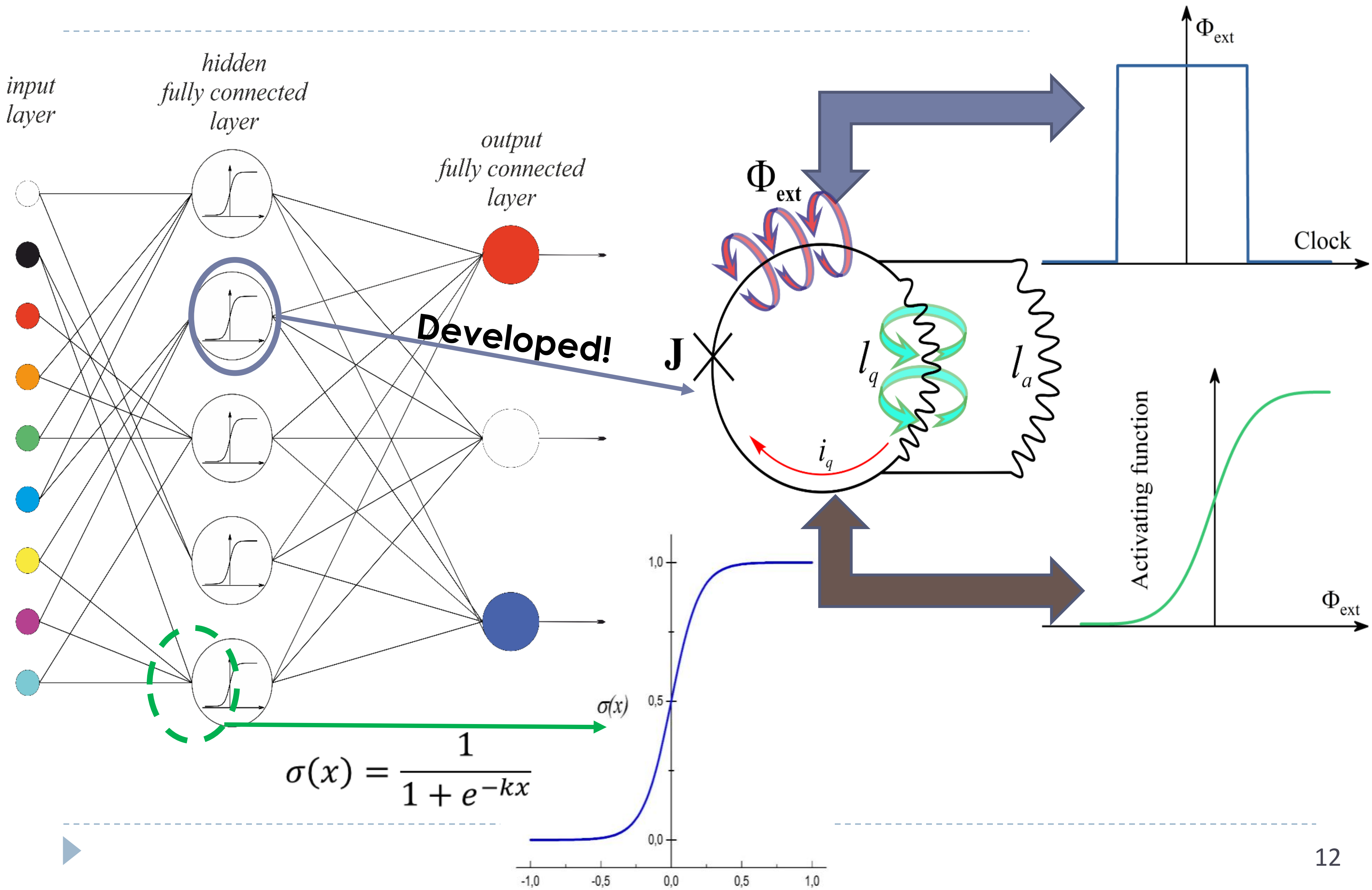


Similar element bases!



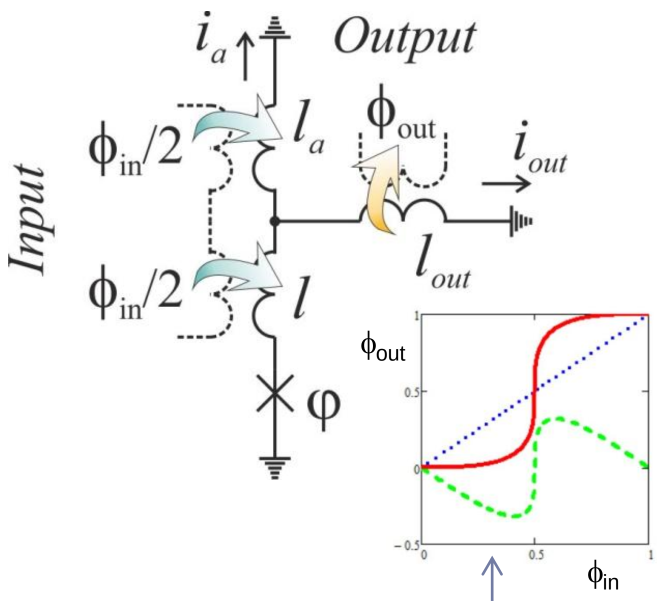
Temperature levels

Sigma-neuron for perceptron



$$\sigma(x) = \frac{1}{1 + e^{-kx}}$$

Sigma-neuron: main idea



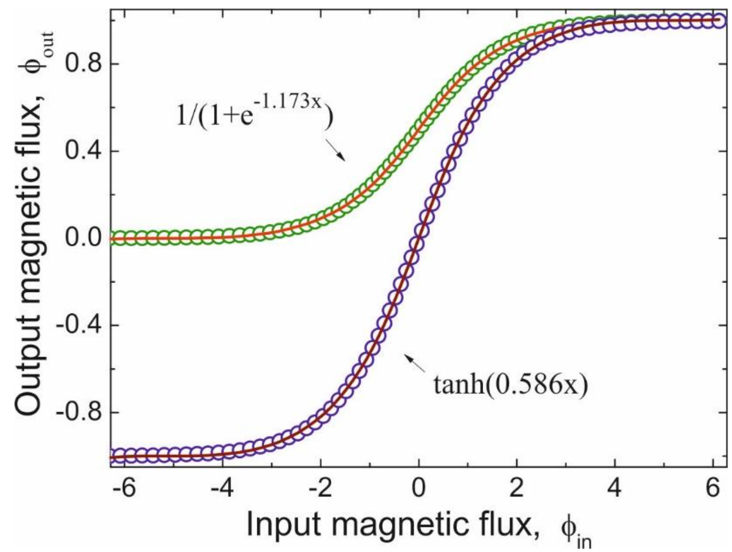
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

or

$$\tau(x) = \tanh(x)$$

$$l_a = 1 + l.$$

$$l = 0.125, \; l_{out} = 0.3$$



$$\varphi + l\sin\varphi = \phi_{in} + l_a i_a,$$

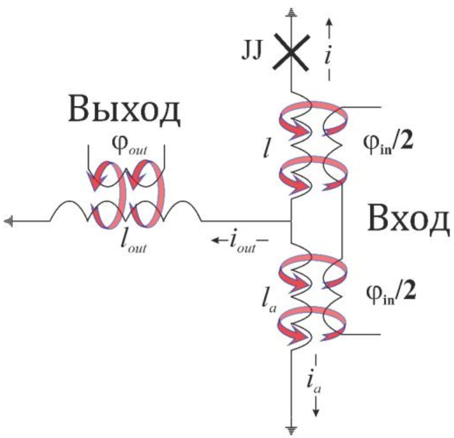
$$\varphi + l\sin\varphi = \phi_{in}/2 + l_{out} i_{out},$$

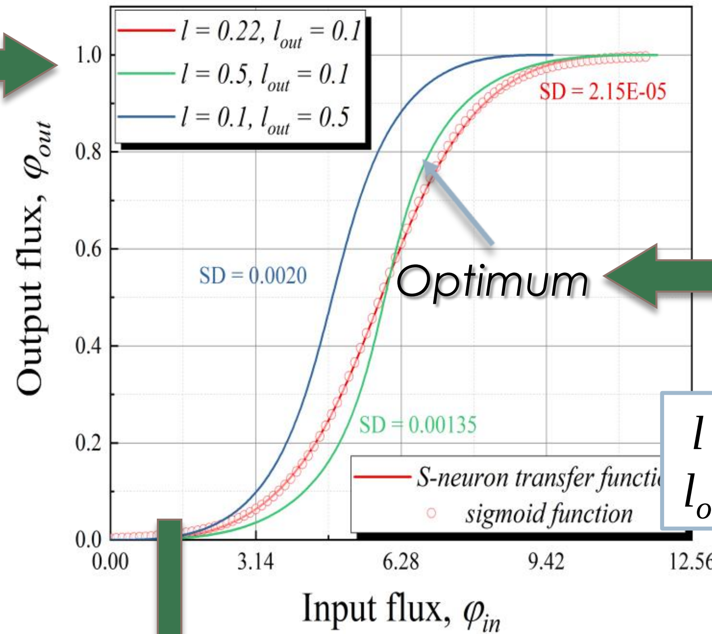
$$\phi_{in} = 2\left(\frac{l_a + l_{out}}{l_a + 2l_{out}}\right)\left[\varphi + \left(l + \frac{l_a l_{out}}{l_a + l_{out}}\right)\sin\varphi\right]$$

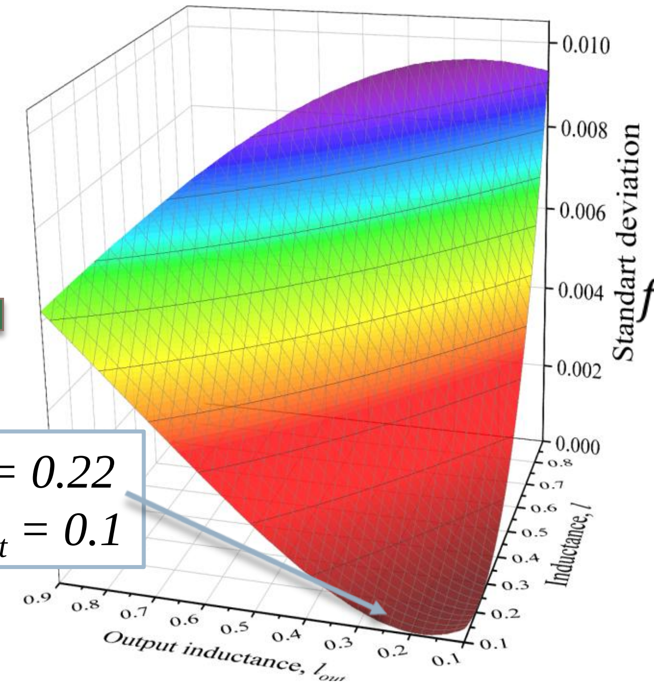
$$\phi_{out} = l_{out}\frac{\phi_{in} - 2l_a\sin\varphi}{2(l_a + l_{out})}$$

Soloviev I. I., Schegolev A. E., Klenov N. V., Bakurskiy S. V., Kupriyanov M. Y., Tereshonok M. V., Golubov A. A. (2018). Adiabatic superconducting artificial neural network: Basic cells. *Journal of Applied Physics*, 124(15), 152113.

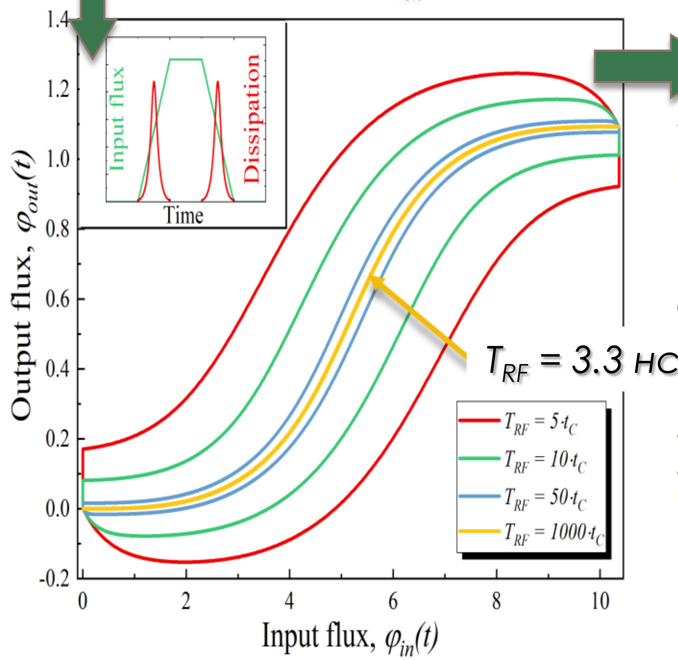
Sigma-neuron: optimization

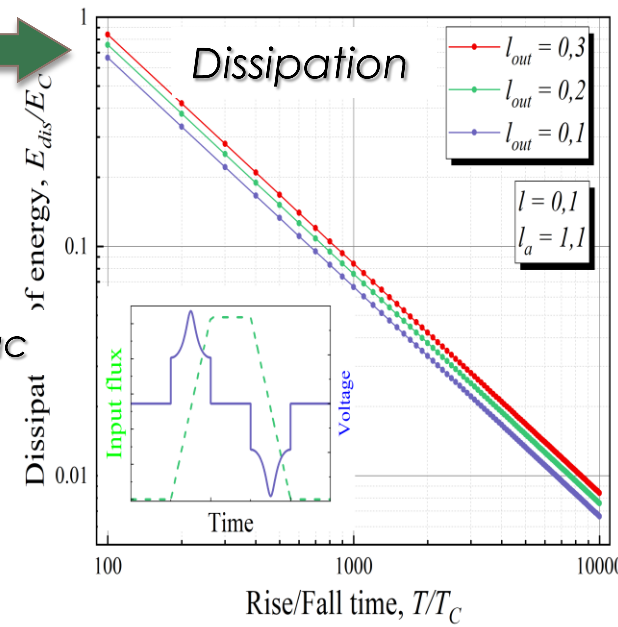






$$\varphi_{out} = l_{out} \frac{\varphi_{in} - 2 l_a \sin\varphi}{2\left(l_a + l_{out}\right)}$$





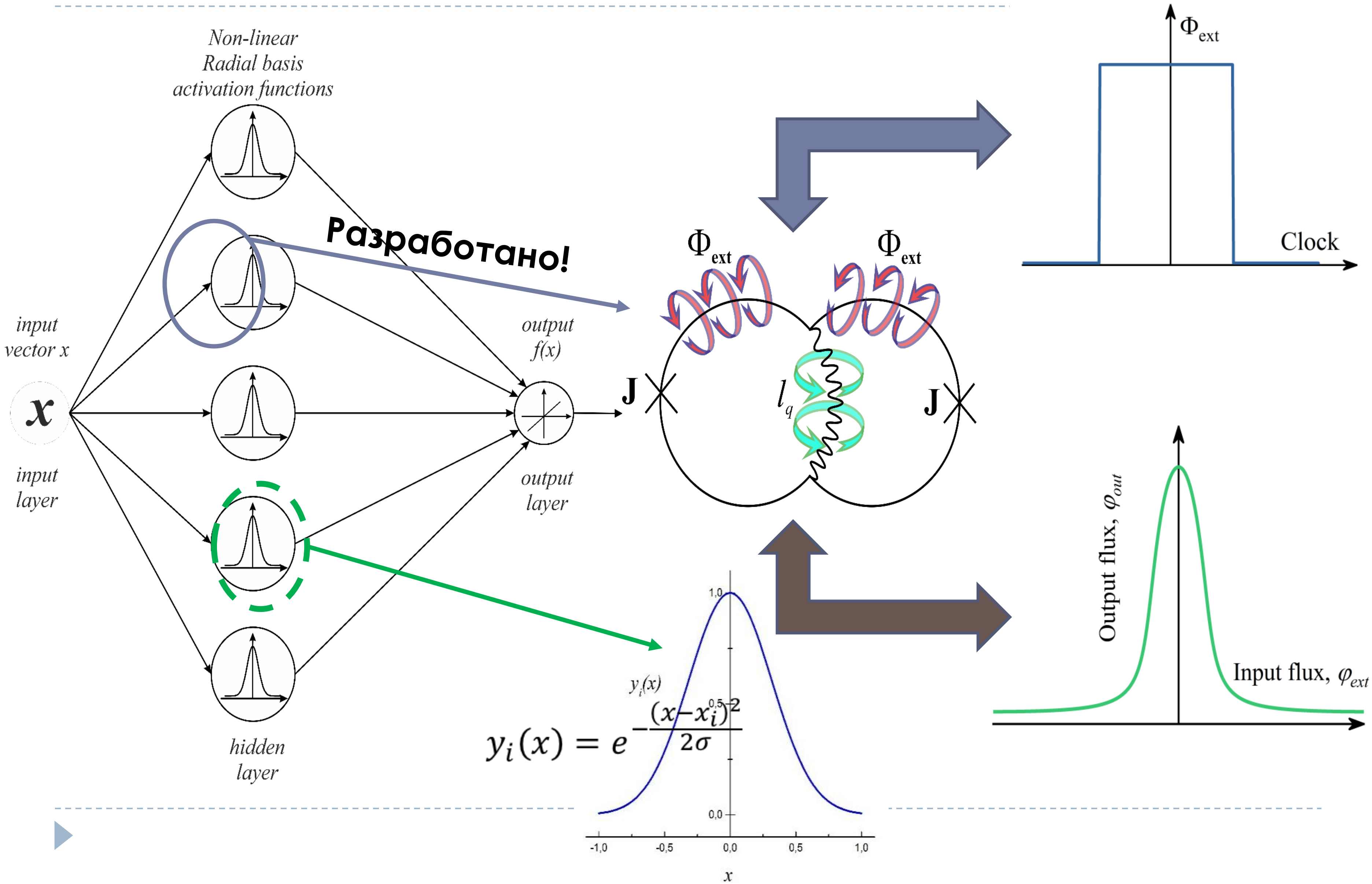
$I_C = 100\ \mu A \quad R_N = 1\ \Omega$

$T_{RF} = 3.3$ ns

$E_{dis} \approx 2.6 \times 10^{-21}$ J = 2.6 zJ

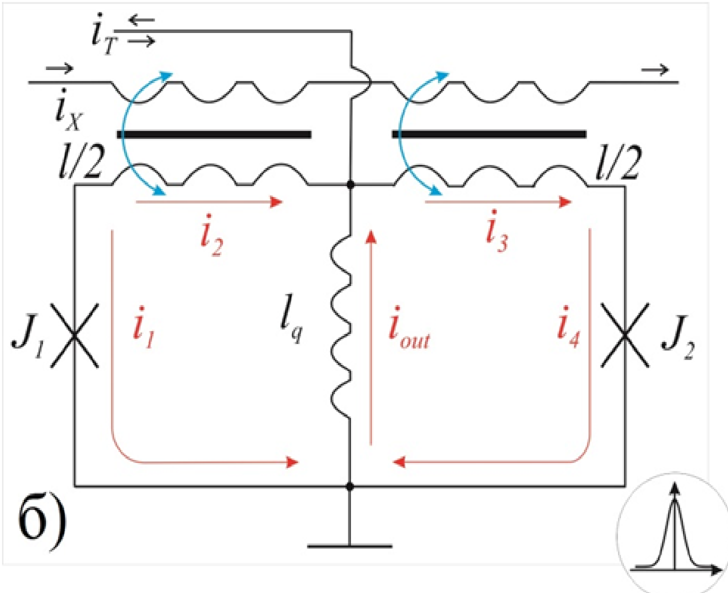
($E_L = k_B T \times \ln 2 = 4 \times 10^{-23}$ J)

Gauss-neuron for RBF-network



$$y_i(x) = e^{-\frac{(x-x_i)^2}{2\sigma}}$$

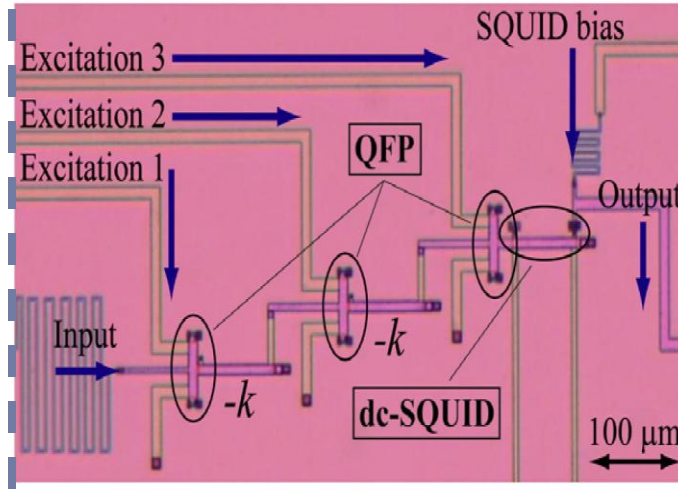
Gauss-neuron: main idea

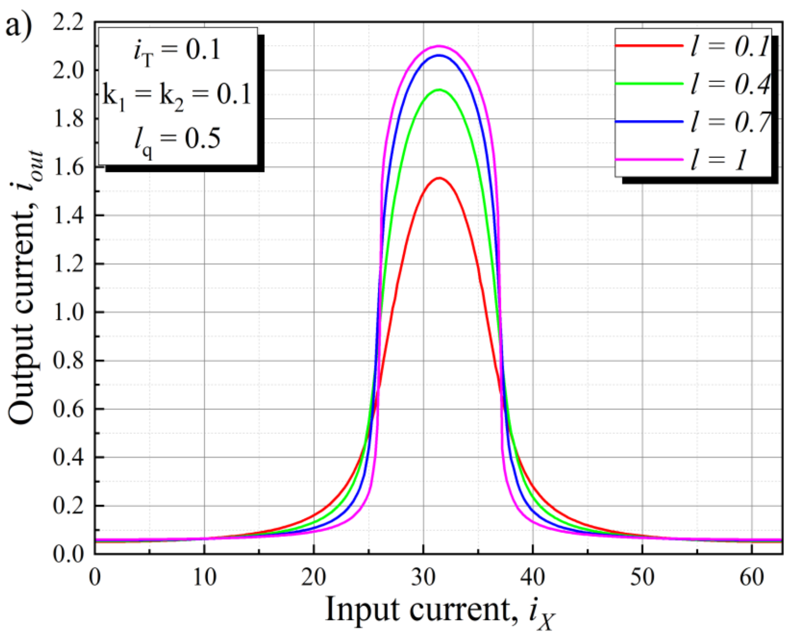


E_{dis} = 10 zJ

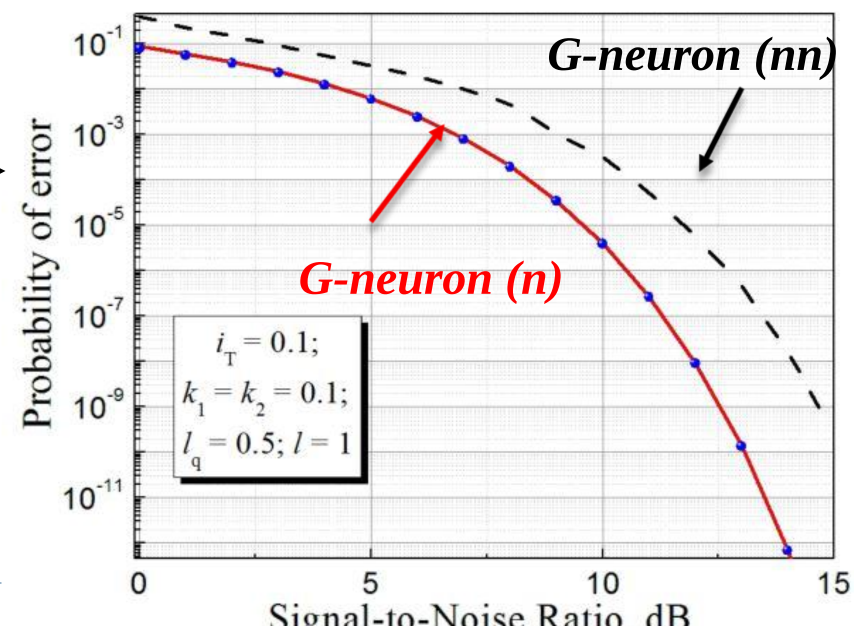
t = 0.2 ns

N. Takeuchi, D. Ozawa, Y. Yamanashi, and N. Yoshikawa

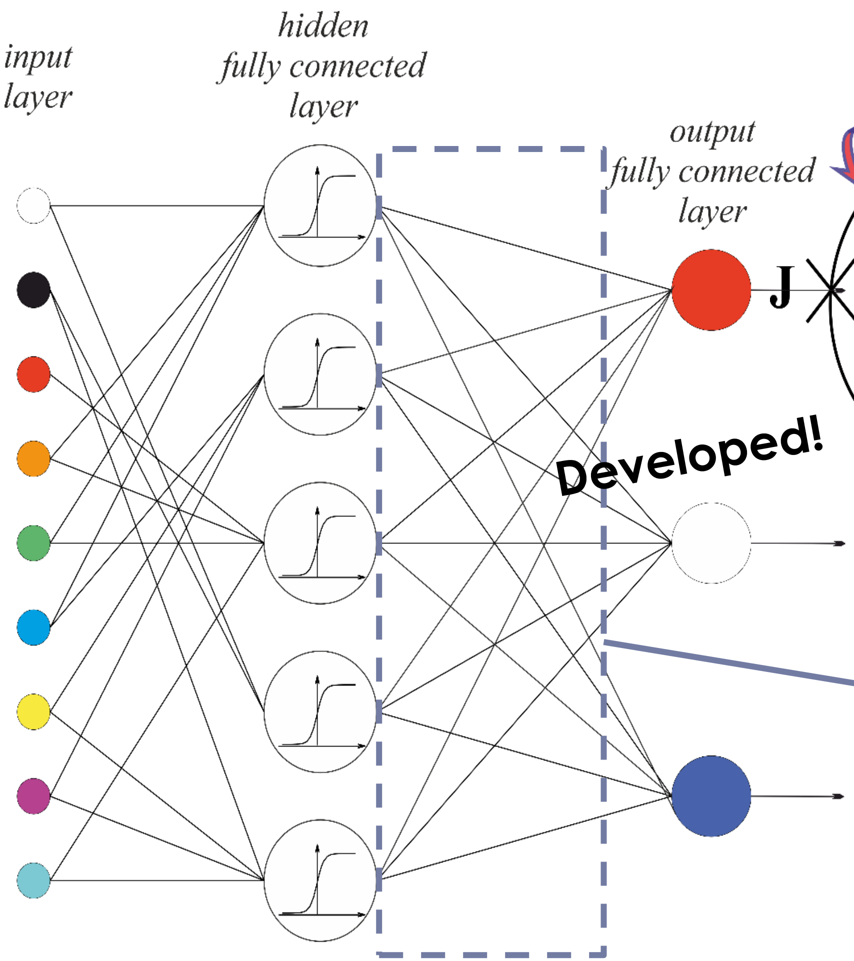


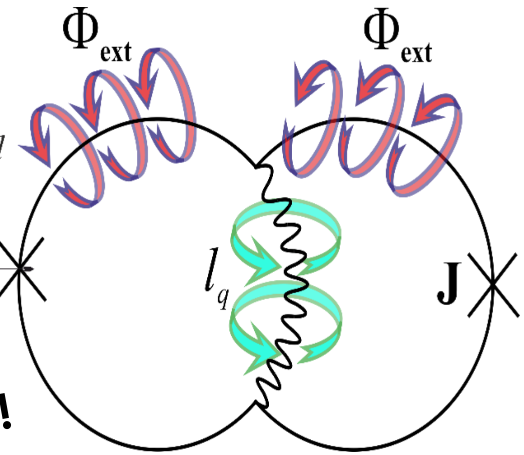


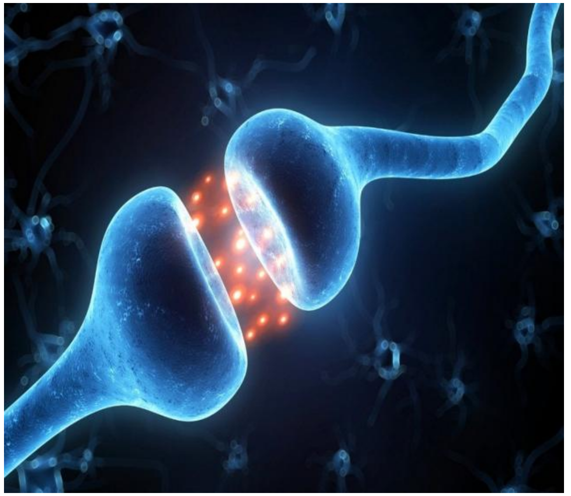
Test

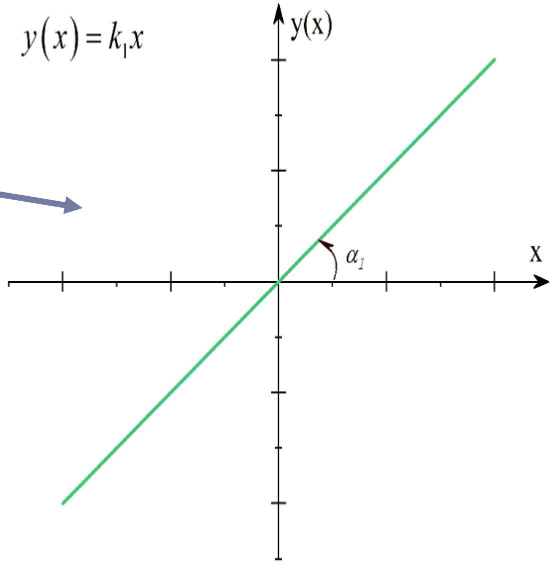


Synapse: main idea

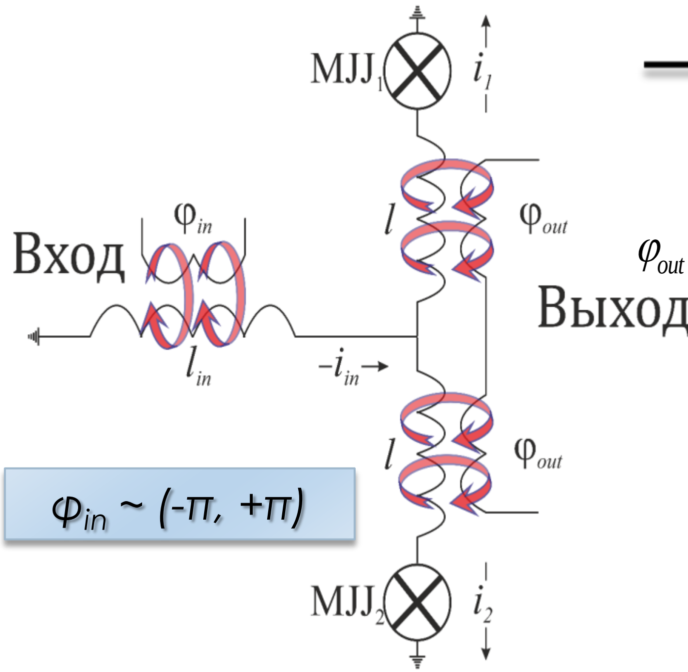






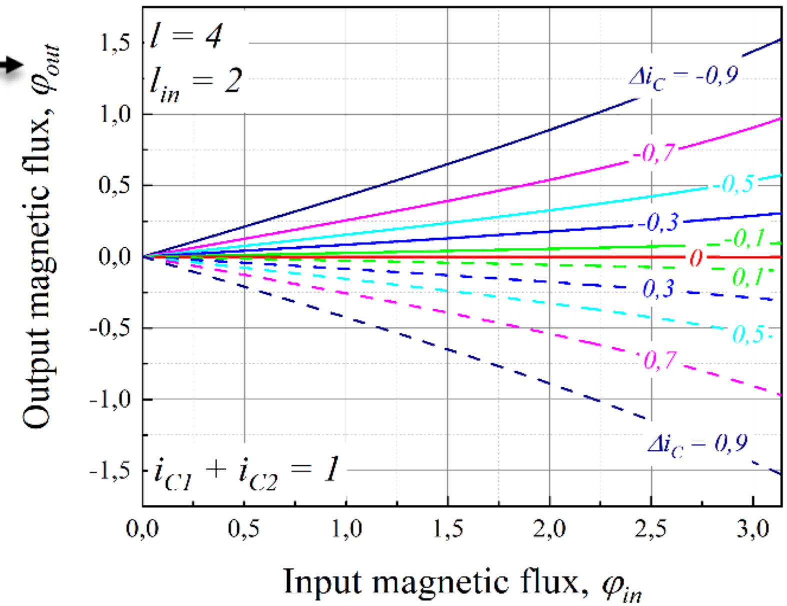


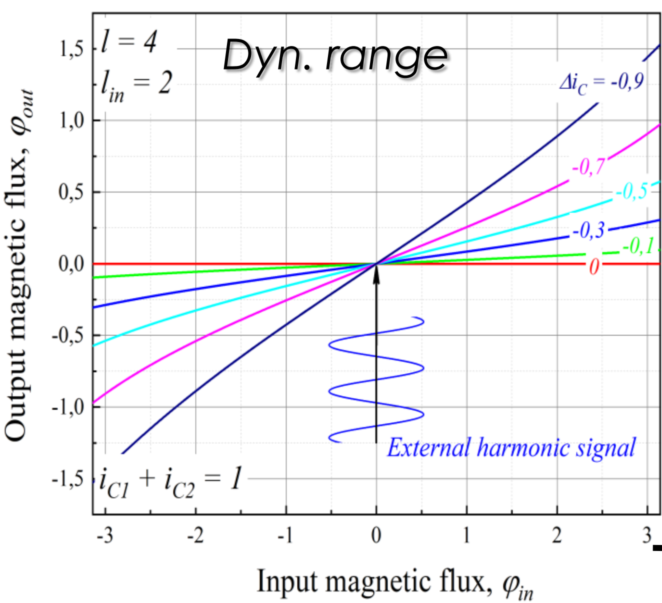
Synapse: realization

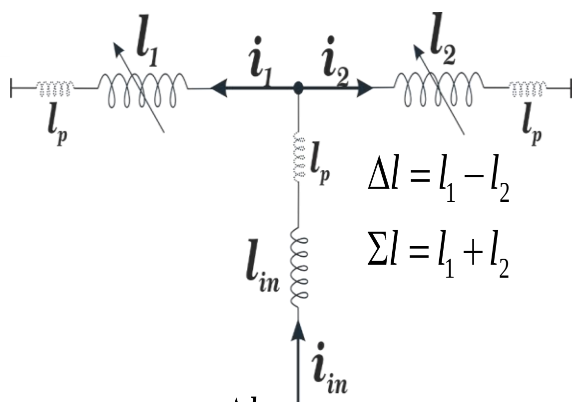


φ_{in} ~ (-π, +π)

$$\varphi_{out} = -2\psi = (i_1 - i_2)l$$



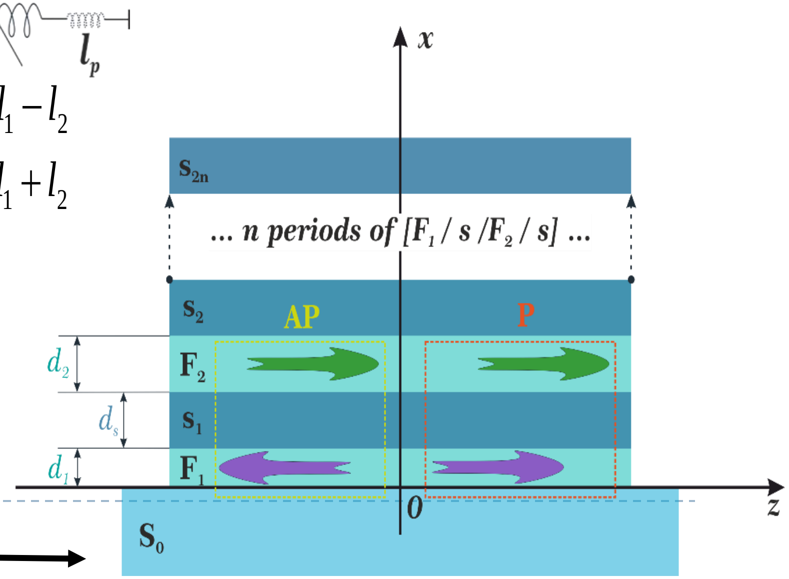




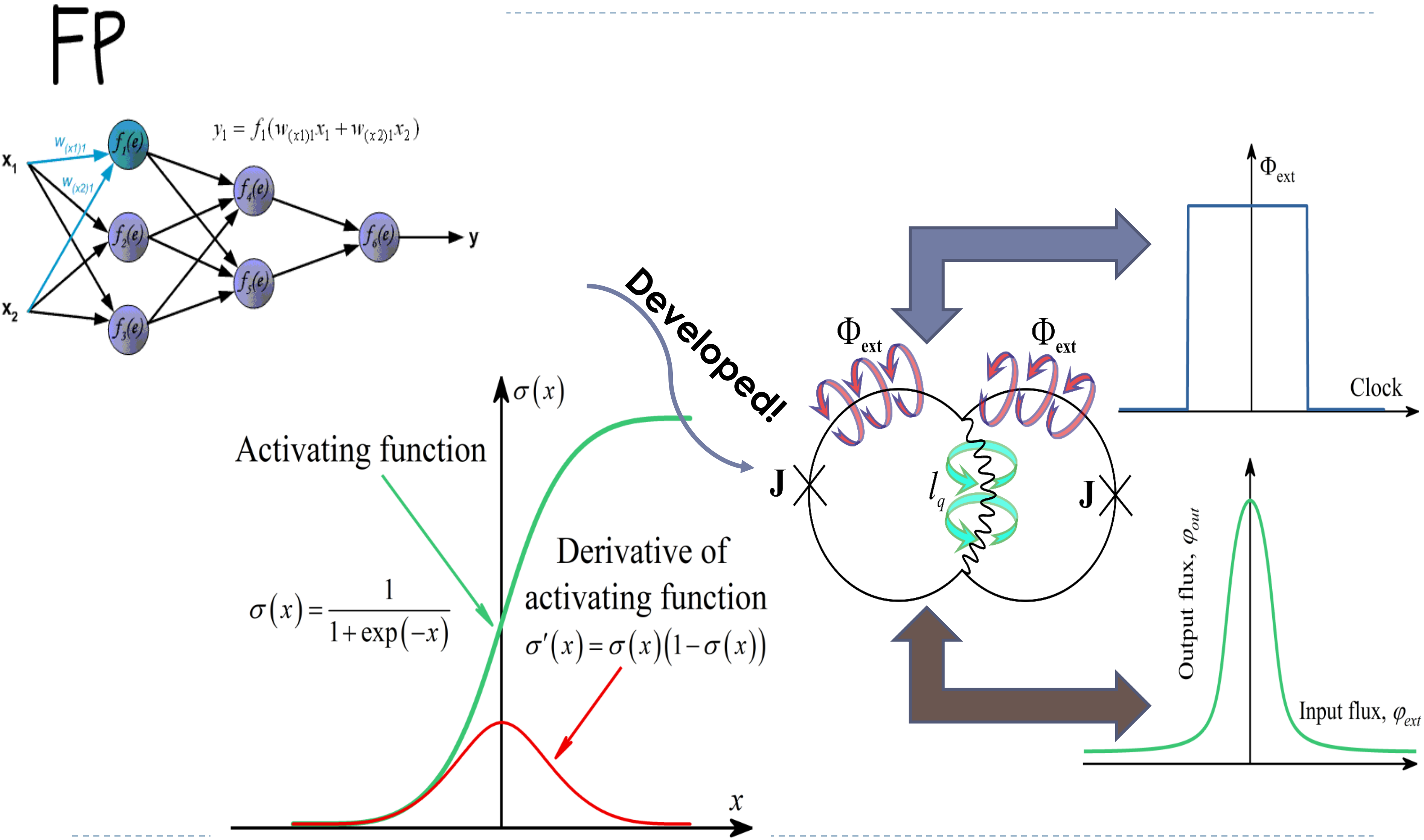
$$\Delta l = l_1 - l_2$$

$$\Sigma l = l_1 + l_2$$

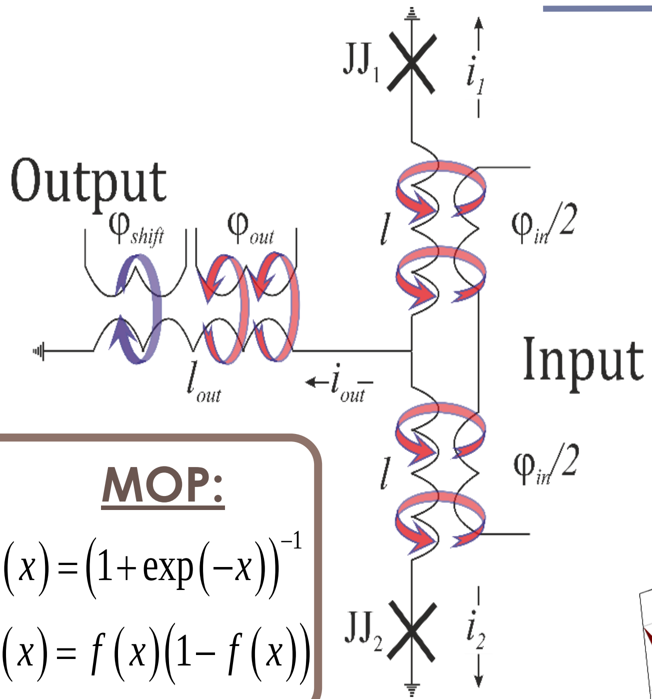
$$i_{out} = \varphi_{in} \cdot \frac{\Delta l}{l_{in}\left(\Sigma l + 2l_p\right)}$$



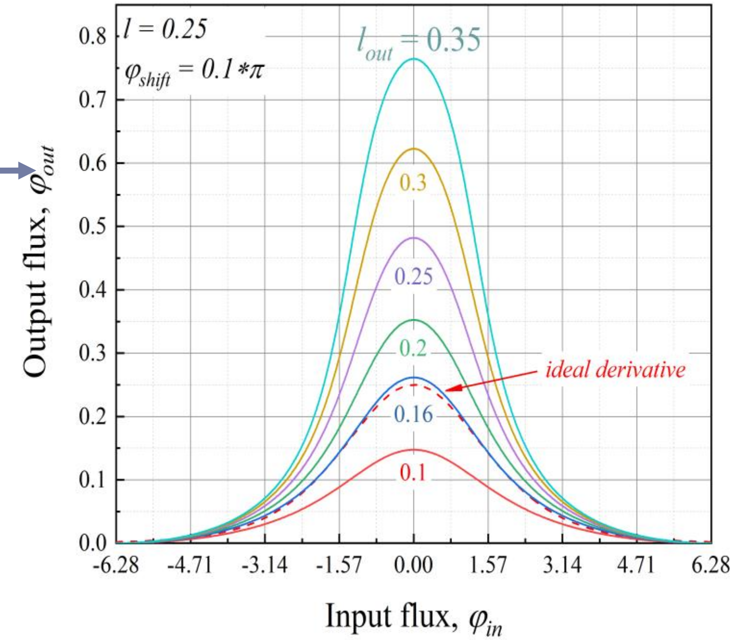
Learning cell: main idea



Learning cell: realization



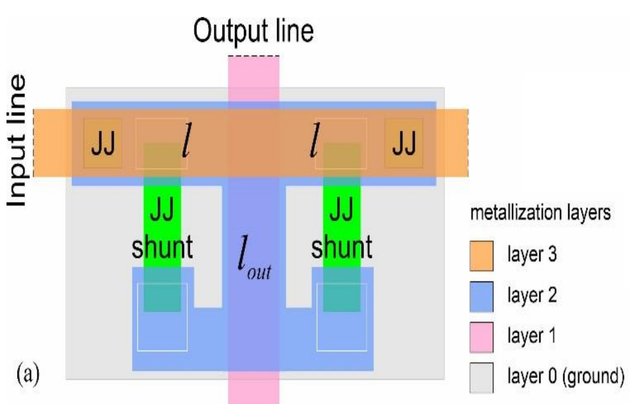
$$
\begin{cases}
i_{1,2} = \dot{\varphi}_{1,2} + i_{C_{1,2}} \cdot \sin\varphi_{1,2} \\
i_{out} + i_1 + i_2 = 0 \\
i_1 l + \varphi_1 + \dfrac{\varphi_{in}}{2} = i_{out} l_{out} + \varphi_{shift} \\
i_2 l + \varphi_2 - \dfrac{\varphi_{in}}{2} = i_{out} l_{out} + \varphi_{shift}
\end{cases}
$$



MOP:

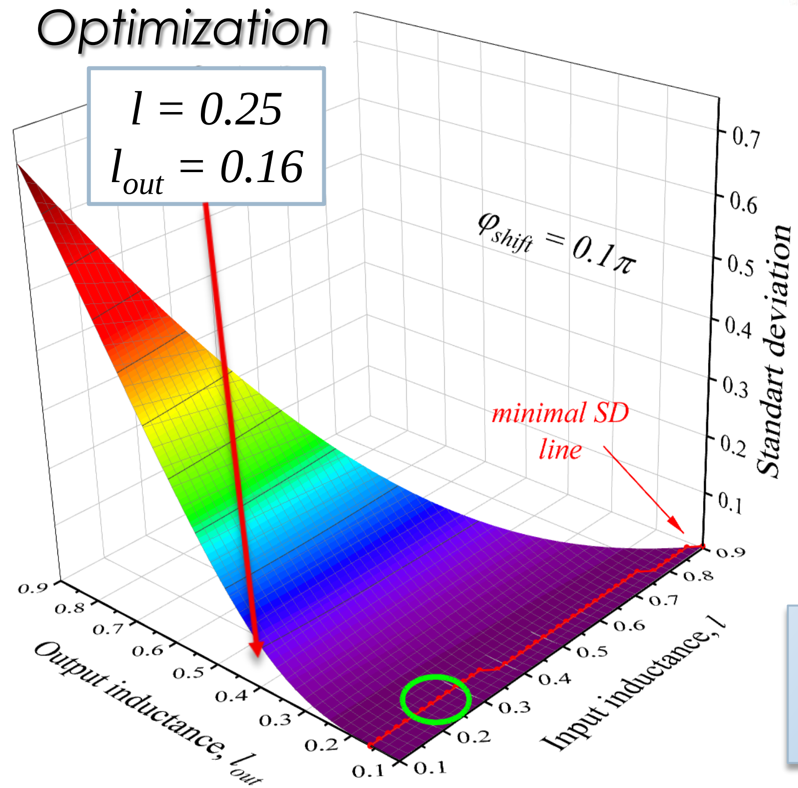
$$f(x) = \left(1 + \exp(-x)\right)^{-1}$$

$$f'(x) = f(x)\left(1 - f(x)\right)$$



Optimization

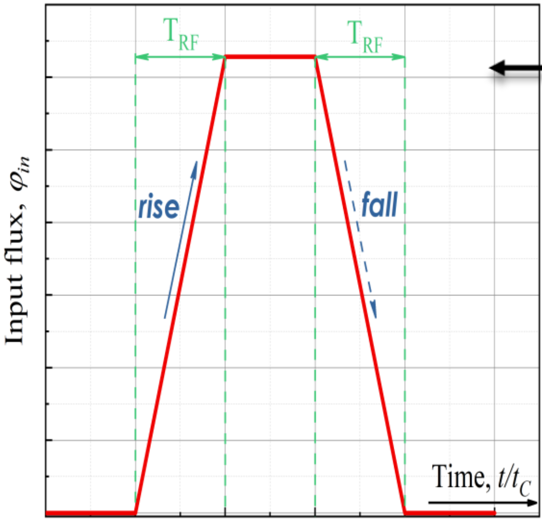
$l = 0.25$
$l_{out} = 0.16$



$$SD = \frac{\sum_{n=1}^{N}\left[f'_x\left(\varphi_{in}^{(n)}\right) - \varphi_{out}\left(\varphi_{in}^{(n)}\right)\right]^2}{N}$$

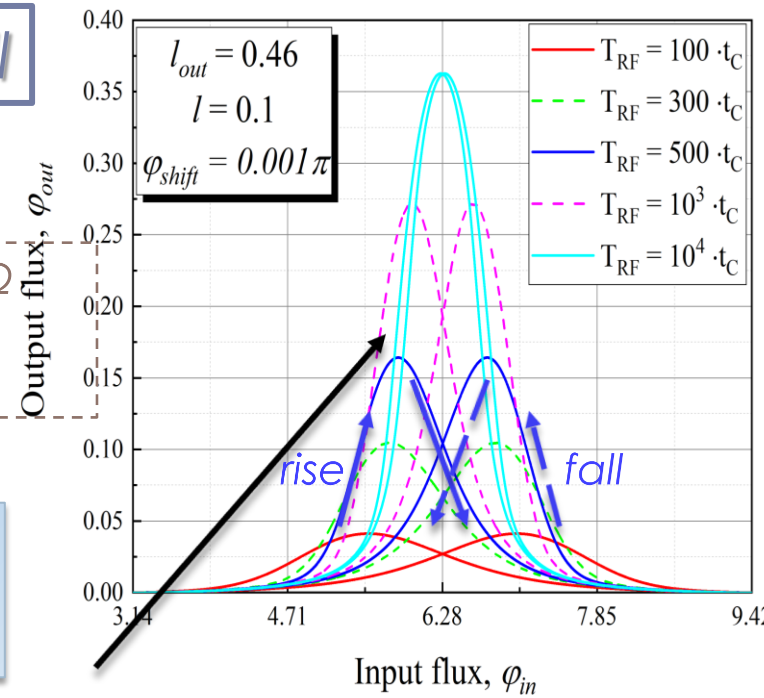
Learning cell: 10×10 µm

Learning cell



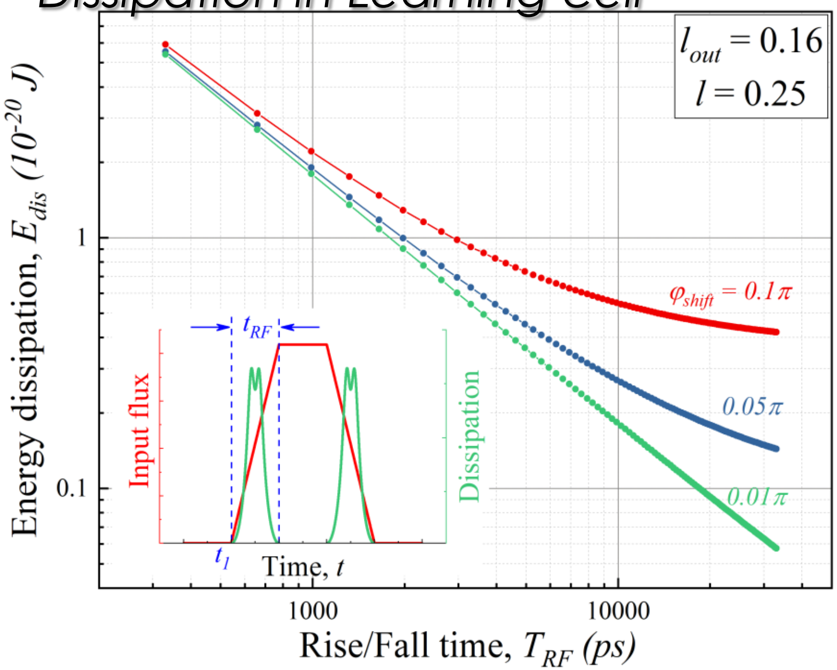
Tact

$I_C = 100\ \mu A$ $R_N = 1\ \Omega$
$T_{RF} = 33$ ns
$E_{dis} \approx 1$ zJ

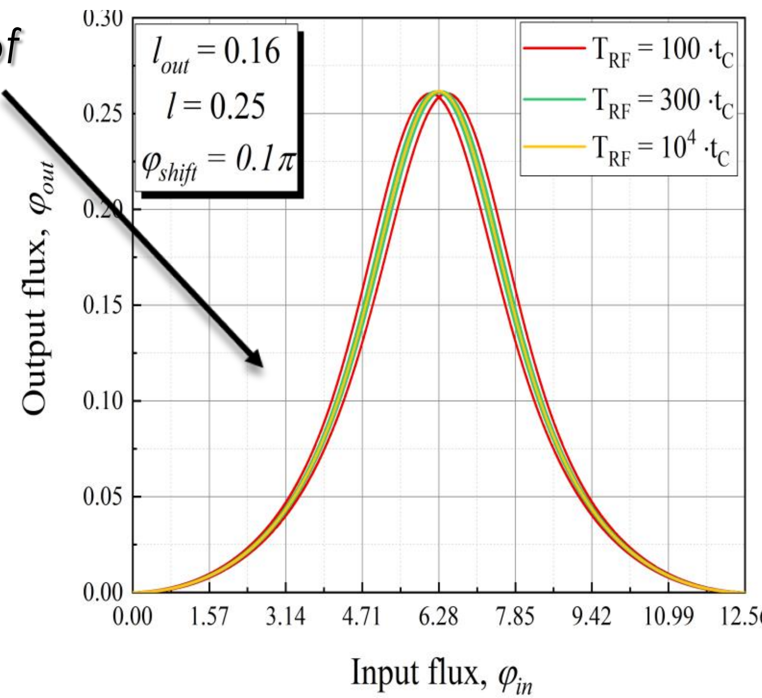


$t \approx 3.3$ ns
$l = 0.25$ $l_{out} = 0.16$
$E_{dis} \approx 4 - 60$ zJ

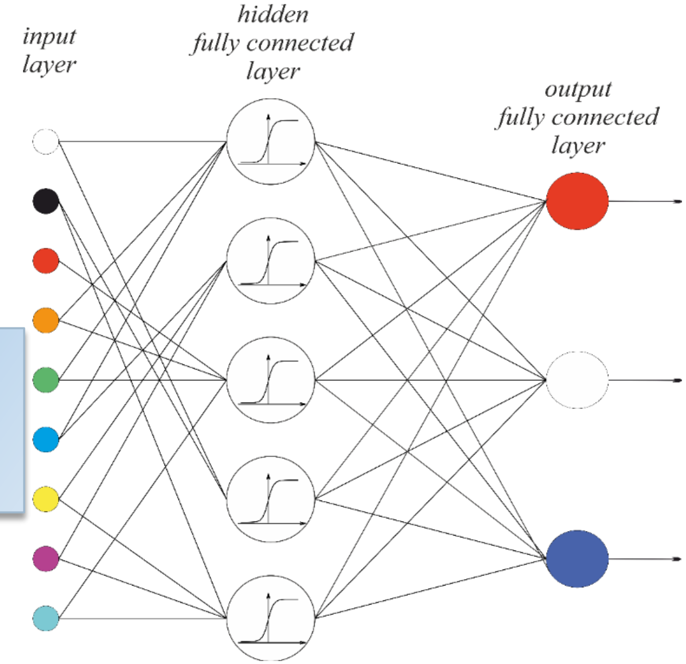
Dissipation in Learning cell

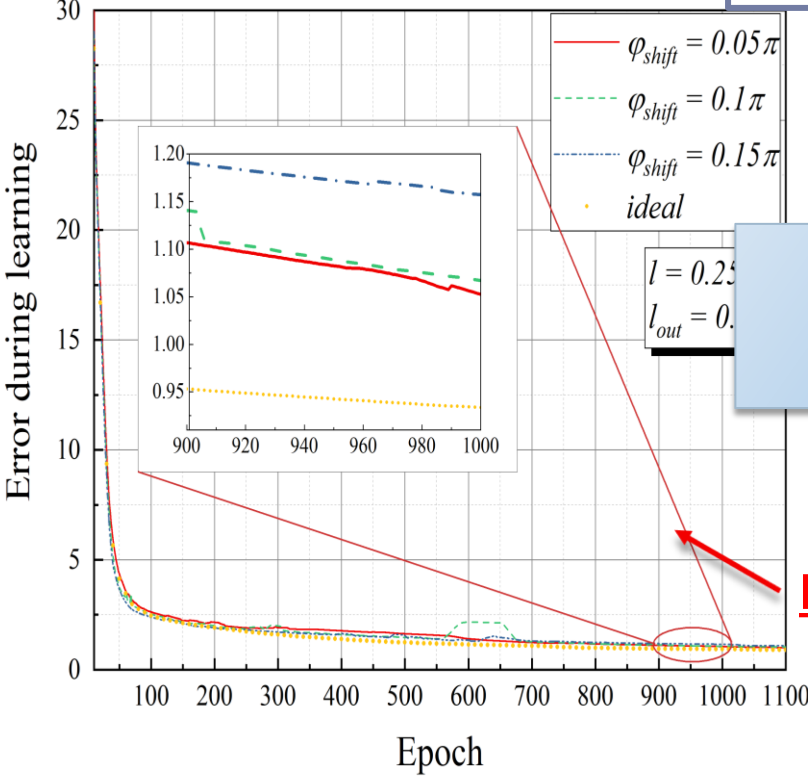


Impact of φ_{shift}



Learning cell

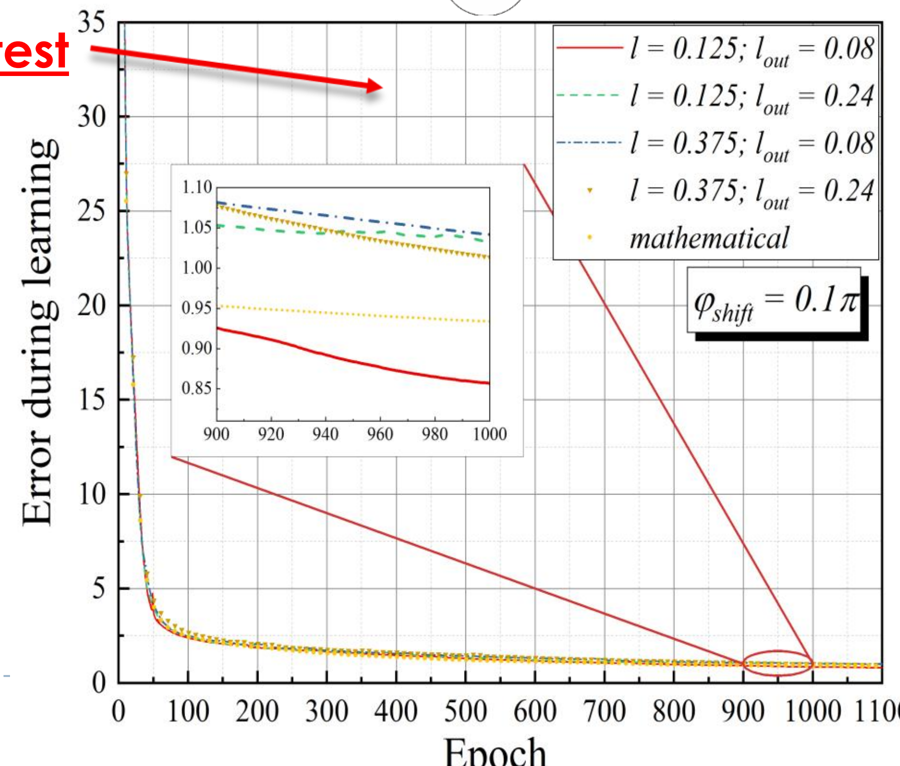




$t_{epoch} \approx 1$ ns

$E_{dis} \approx 10$ zJ

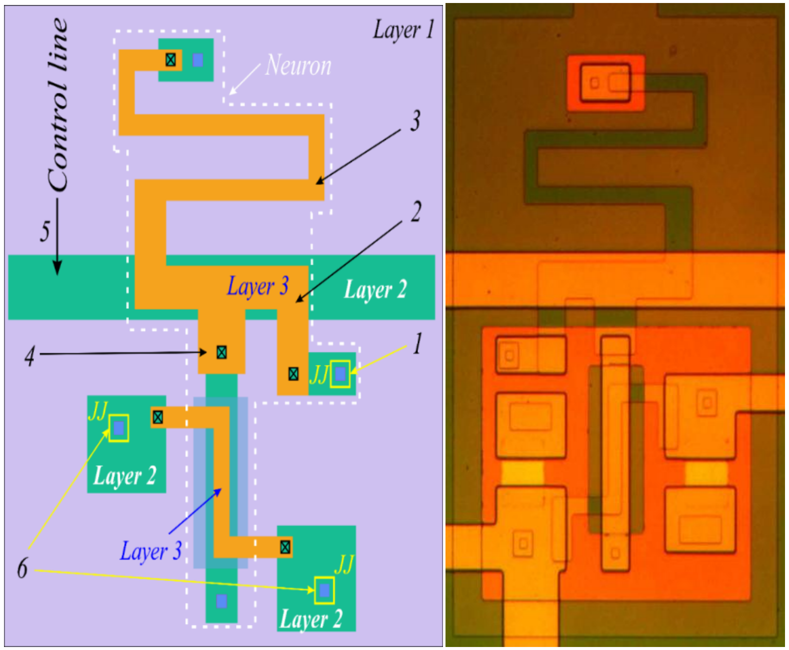
Fisher test



A perceptron network of 16 neurons was modeled:
input layer – 4 neurons,
hidden layer – 9 neurons,
output layer – 3 neurons

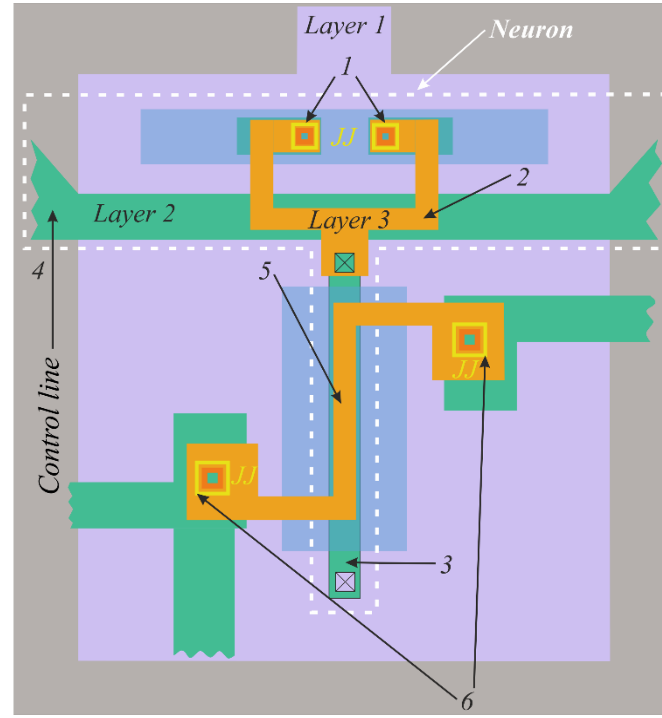
Realization

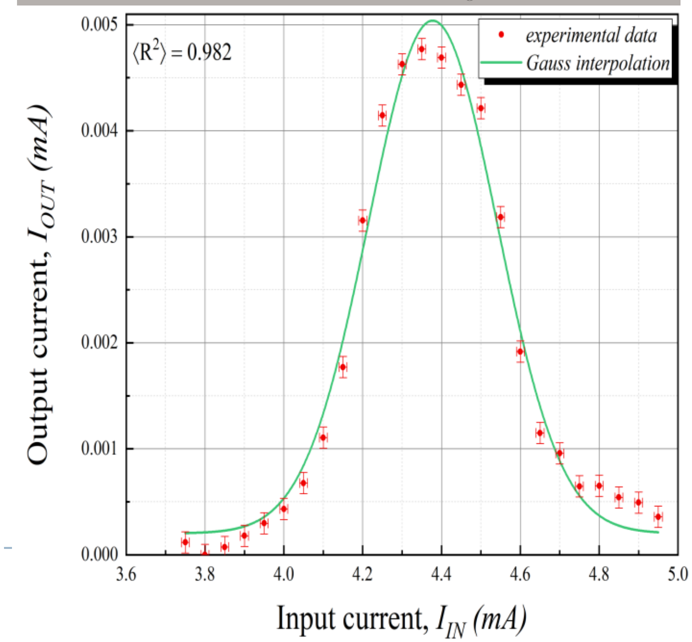
S-neuron



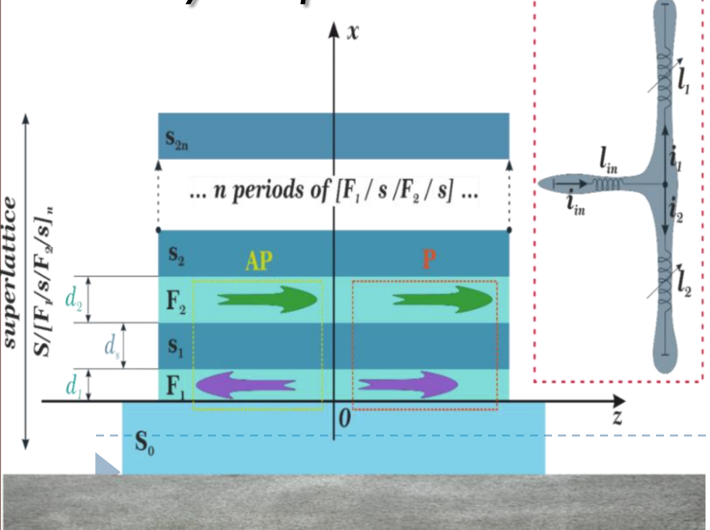
Thanks for V.M.K. & A.S.S. groups!

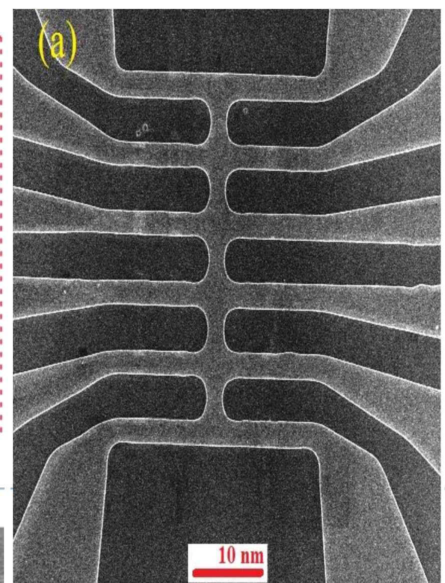
G-neuron



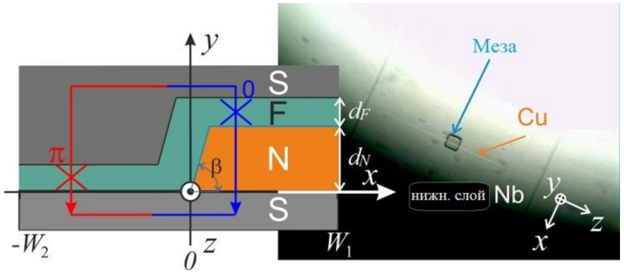


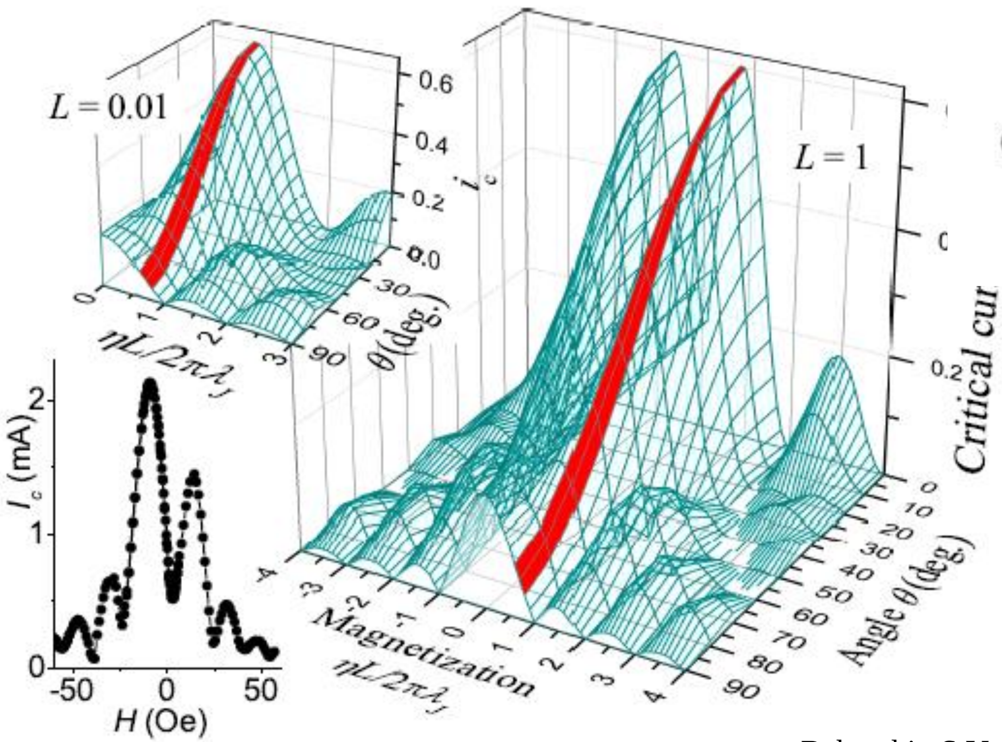
Ind. synapse

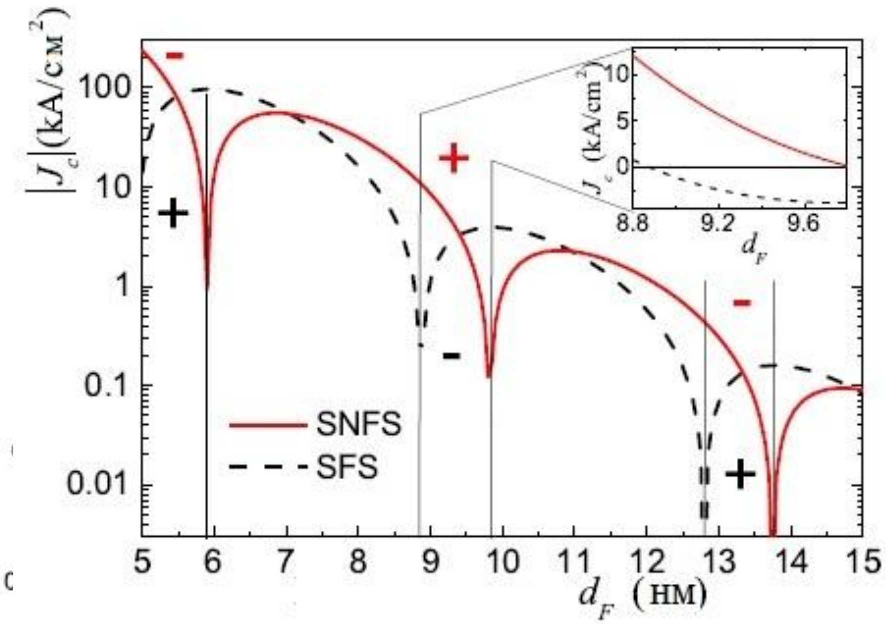




Magnetic rotary valve (thanks to A.A.G. and V.V.R. groups)







SF-NFS → **SIsF-NFS**

- ✓ up to 100 GHz
- ✓ strong I_C modulation

Bakurskiy S.V., **Klenov N.V.**, Soloviev I.I., Bol'ginov V.V., Ryazanov V.V., Vernik I.V., Mukhanov O.A., Kupriyanov M.Yu., Golubov A.A. Theoretical model of superconducting spintronic SIsFS devices // Applied Physics Letters. — 2013. — Vol. 102. — P. 192603.

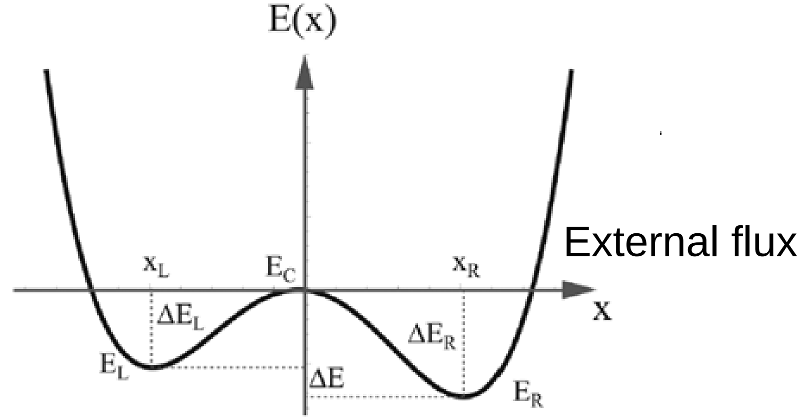
Qubit as a neuron?

$$f(x) = \frac{1}{1+e^{-x}}$$

$$\frac{dp_L}{dt} = -\Gamma_{LR}p_L + \Gamma_{RL}p_R,$$

$$\frac{dp_R}{dt} = -\Gamma_{RL}p_R + \Gamma_{LR}p_L,$$

$$\frac{dp_L}{dt} = -(\Gamma_{LR}+\Gamma_{RL})p_L + \Gamma_{RL}$$

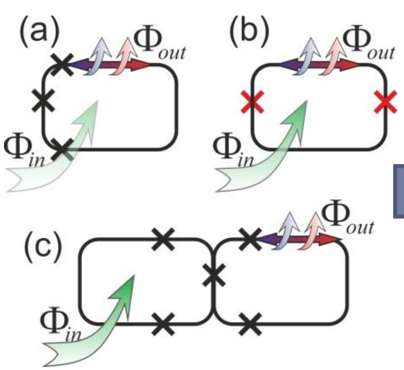


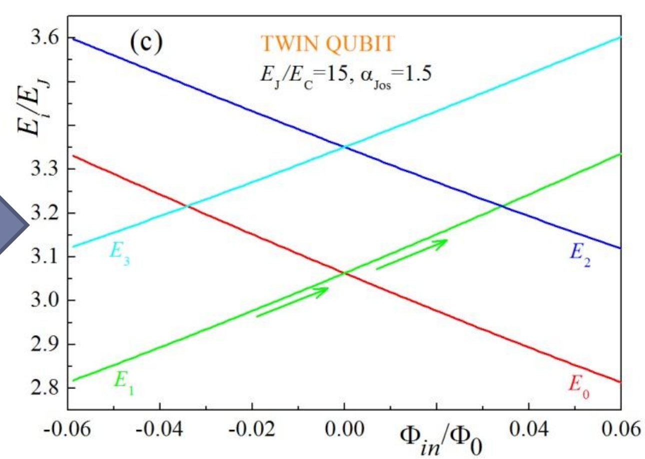
$$p_L = \left(1 - \frac{\Gamma_{RL}}{\Gamma_{LR}+\Gamma_{RL}}\right)e^{-(\Gamma_{LR}+\Gamma_{RL})t} + \frac{\Gamma_{RL}}{\Gamma_{LR}+\Gamma_{RL}}$$

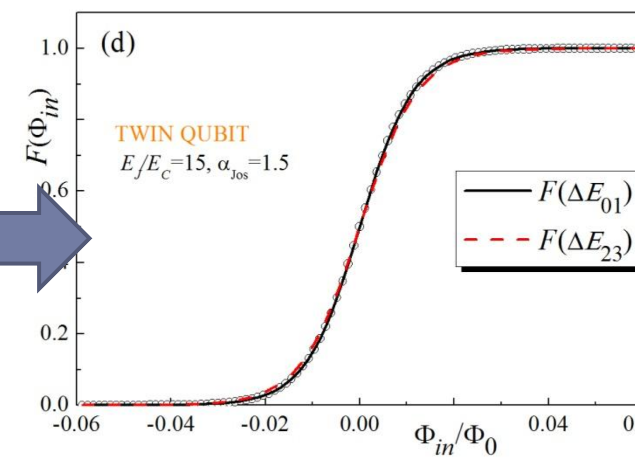
$$p_L = \frac{\Gamma_{RL}}{\Gamma_{LR}+\Gamma_{RL}} = \frac{1}{1+\Gamma_{LR}/\Gamma_{RL}}$$

$$\Gamma = \frac{\omega_0}{2\pi}e^{-\frac{\Delta U}{k_B T}}$$

$$p_L = \frac{1}{1+e^{-\frac{\Delta E}{k_B T}}}$$







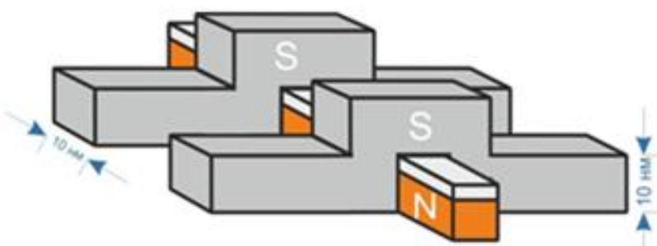
N. V. Klenov, et al. *Low Temperature Physics*, 45(7):769–774, 2019.

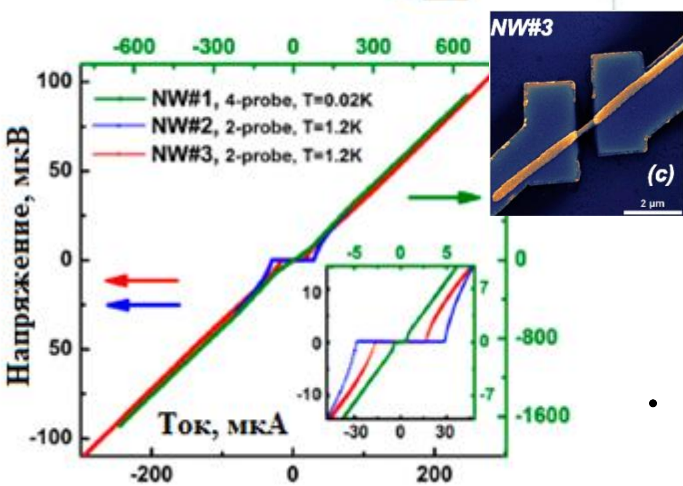
A. V. Bogatskaya, N. V. Klenov, et al. *Laser Physics Letters*, 16(5):056006, 2019.

THANK YOU FOR YOUR ATTENTION

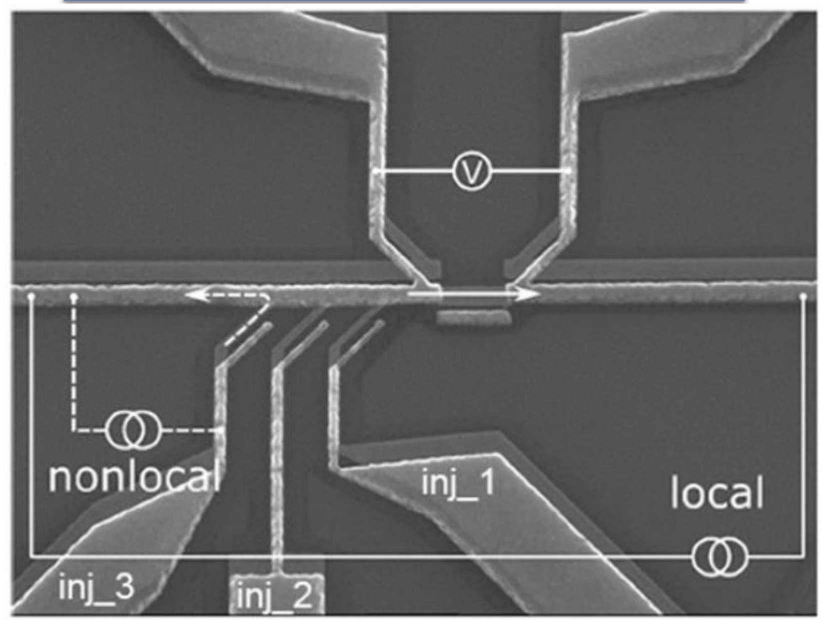
Практические реализации: уменьшение размеров базовых элементов (I)

Научная Новизна. Созданы и апробированы методики для анализа процессов переноса заряда в компактных джозефсоновских элементах и фазовых батареях (с учетом особенностей влияния топологии при переходе к наноразмерным структурам), входящих как в состав ШП АЦП, так и в состав сигнального процессора, нейросетевого и квантового блока обработки сигнала



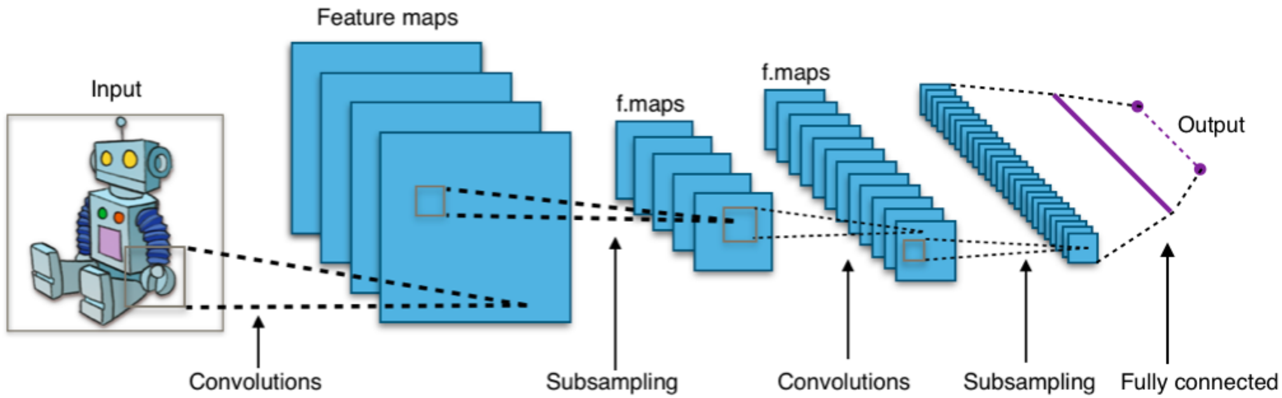


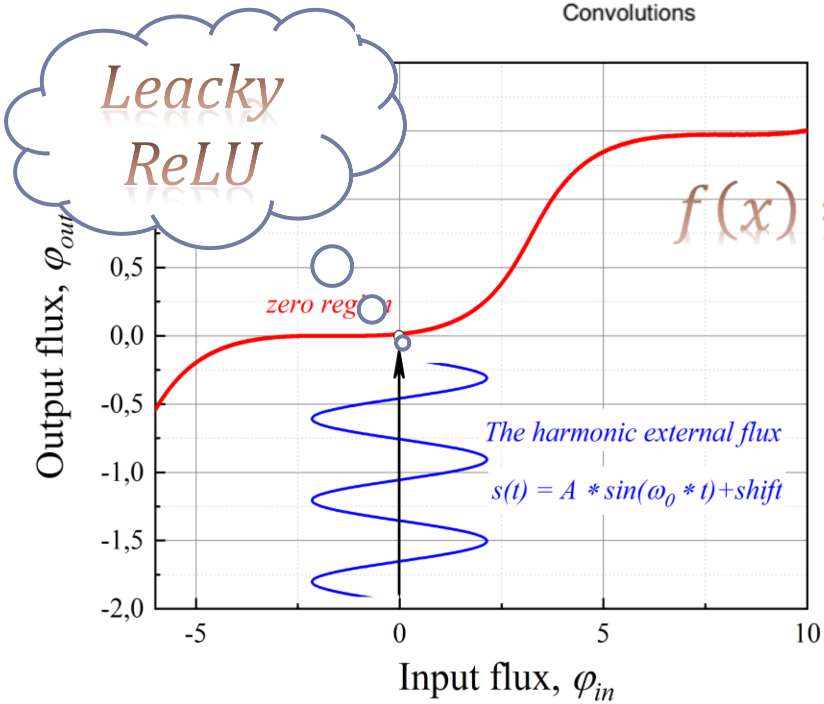
Отработка технологических решений



- Golikova T.E., Hubler F., Beckmann D., **Klenov N.V.**, Bakurskiy S.V., Kupriyanov M.Yu., Batov I.E., Ryazanov V.V. Critical current in planar SNS Josephson junctions // JETP Letters. — 2013. — Vol. 96, no. 10. — P. 668–673.

Сверхпроводниковый ReLU (Rectifier Linear Unit) сверточных ИНС



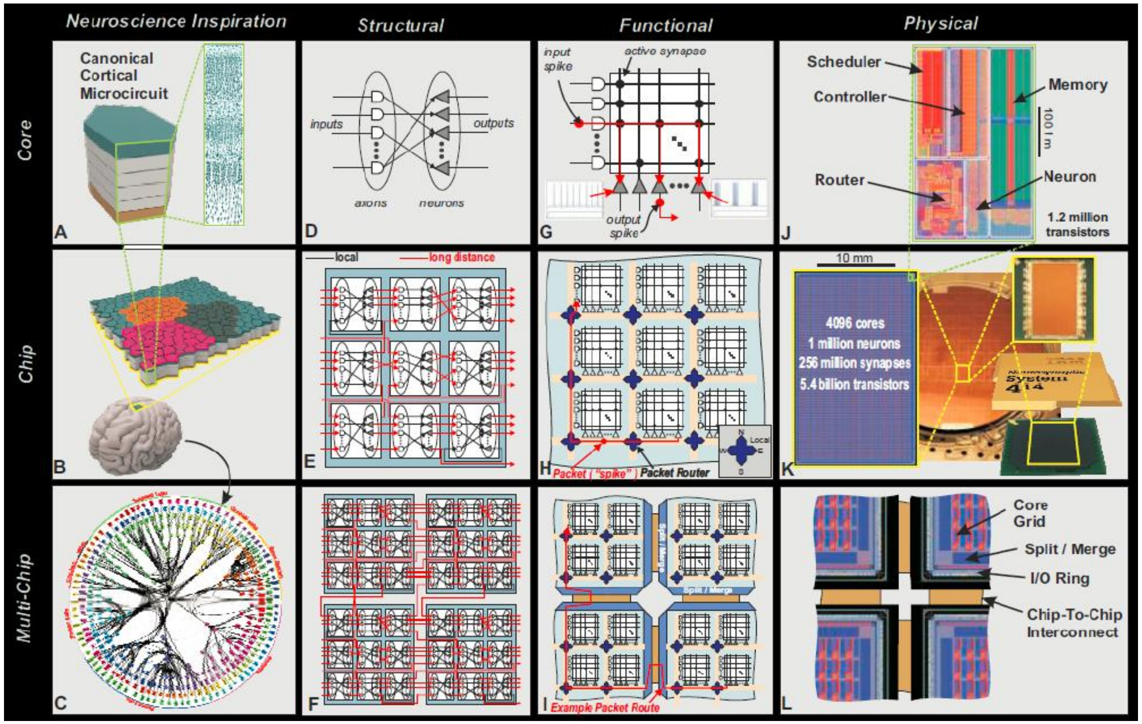


$f(x) = max\{0, x\}$ - математическая запись

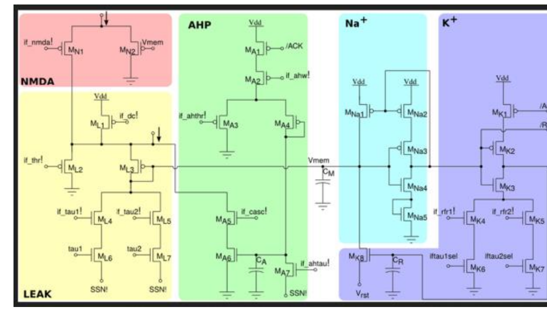
Позволяет проводить обучение сверточной сети в несколько раз быстрее функций гиперболического тангенса или сигмоидной функции, без ущерба обобщающим свойствам сети

Современные подходы к созданию элементной базы нейронных сетей

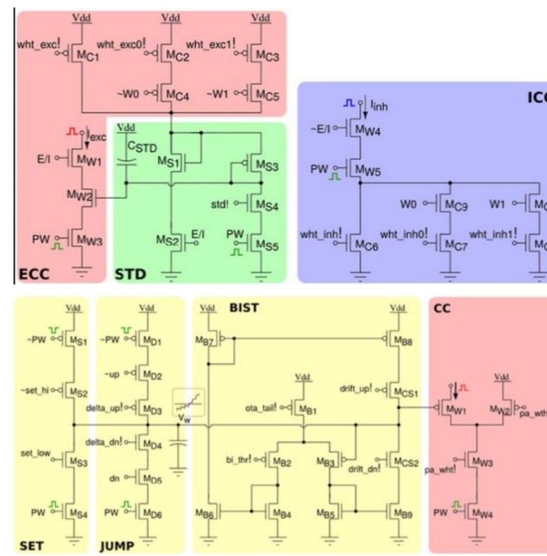
IBM TrueNorth



КМОП нейрон



Схемы для синапса



- Science 345, 668 (2014)
- Front Neurosci. 9, 141 (2015)

Состояние исследований. Джозефсоновские структуры

<u>Основные физические эффекты:</u>

- Эффекты Джозефсона

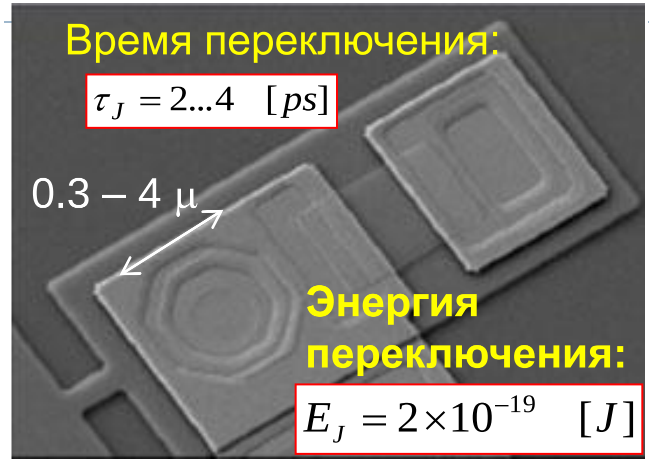
$$I = I_c \sin(\varphi)$$

$$V = \frac{\eta}{2e}\frac{d\varphi}{dt} = \frac{\Phi_0}{2\pi}\frac{d\varphi}{dt}$$

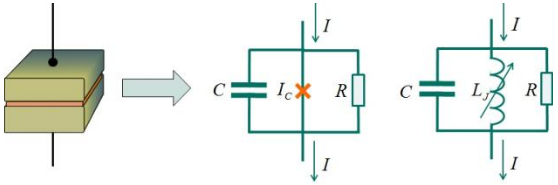
- Квантование магнитного потока

$$\Phi_0 = \frac{h}{2e} = \int V dt = 2.07 \quad [мВ \cdot пс]$$

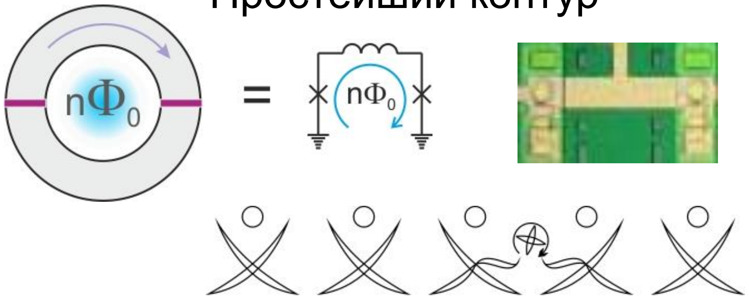
или $2.07 \times 10^{-15} \quad [Вб]$



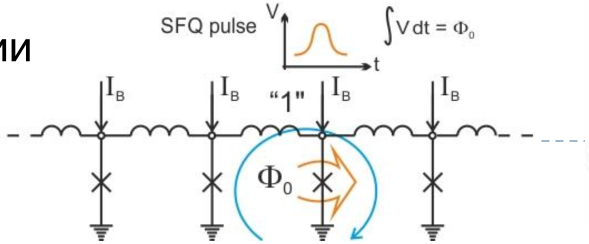
Джозефсоновский контакт с резистивным шунтом

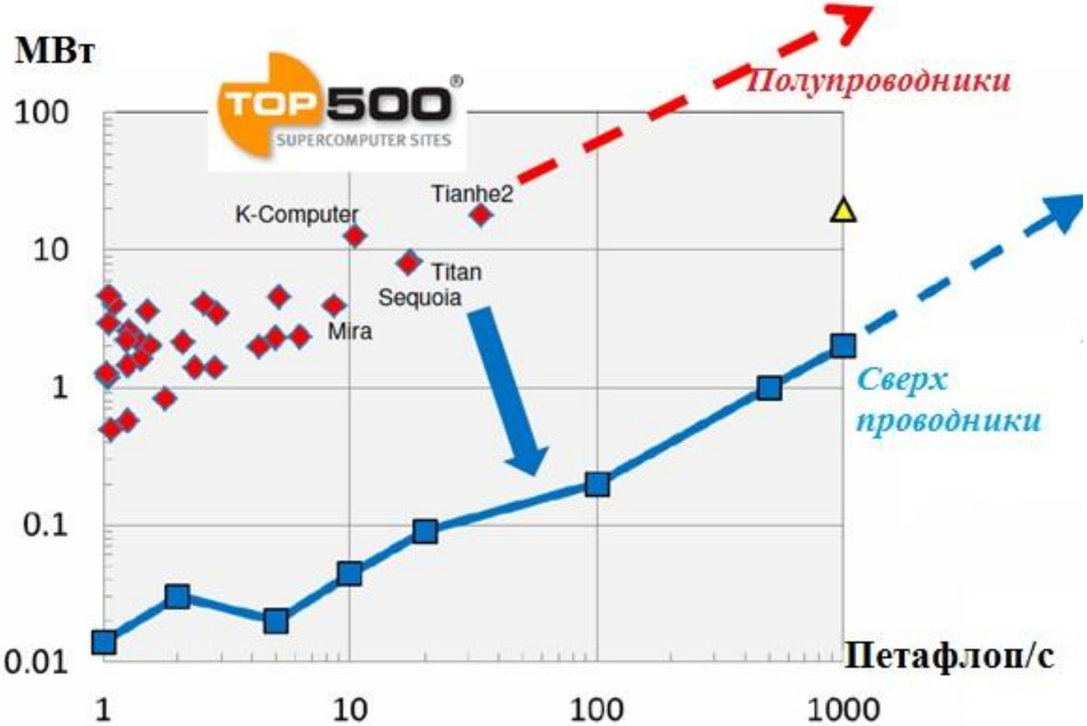


Простейший контур

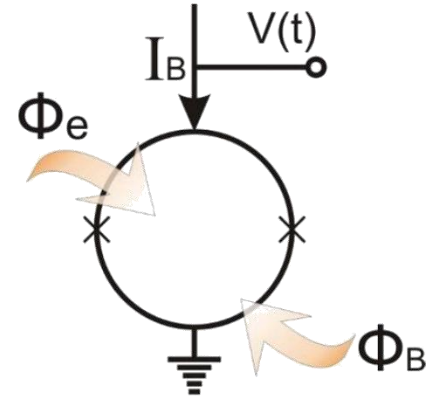


Передача информации



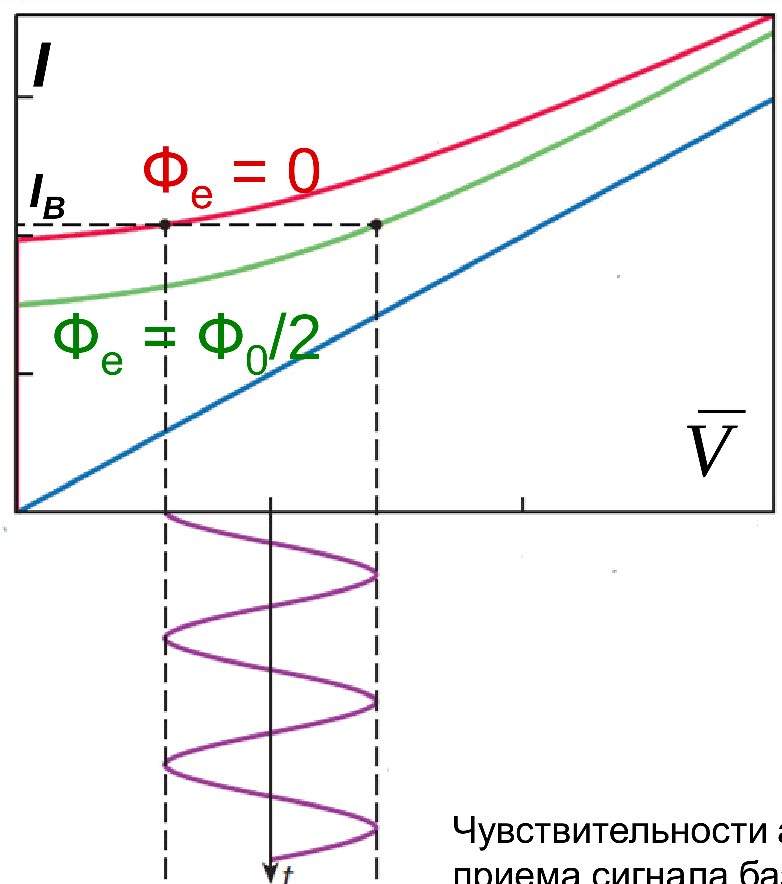


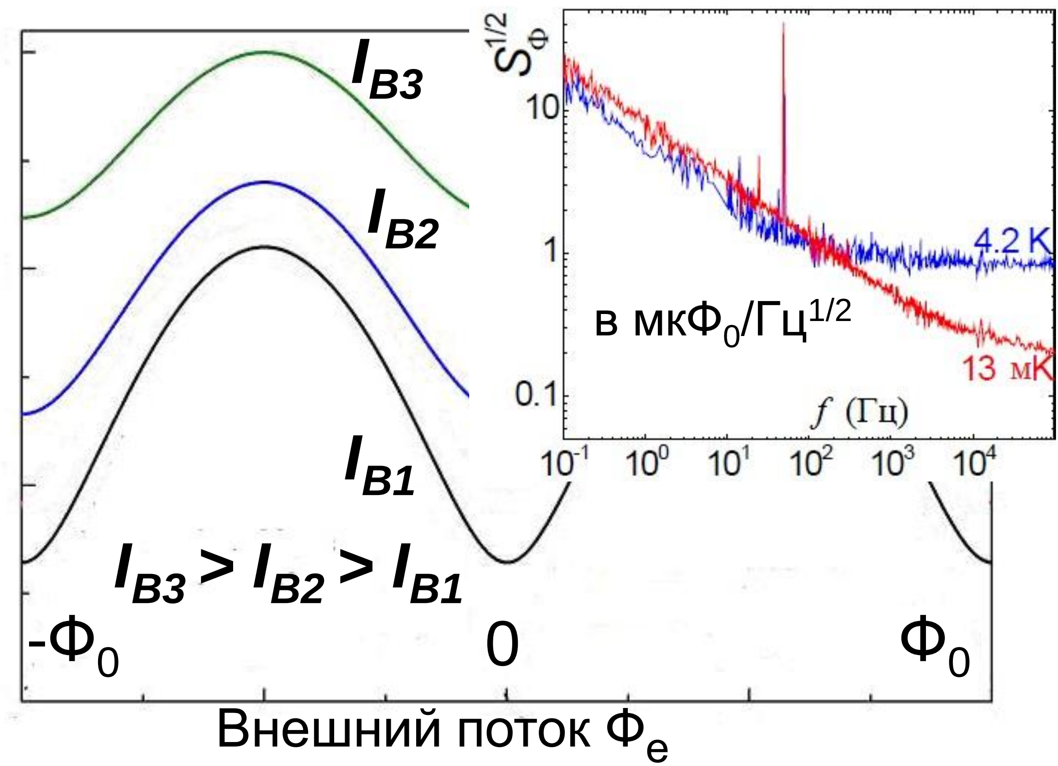
Состояние исследований.
Активный элемент приемной системы



R. C. Jaklevic, J. Lambe, A. H. Silver, and J. E. Mercereau "Quantum Interference Effects in Josephson Tunneling" *Phys. Rev. Letters* **12** (7), 159–160 (1964)

<u>СКВИД</u> – сверхпроводящий квантовый интерферометр





Чувствительности антенны, занимающей площадь 3.3x3.3 мм^2 достаточно для приема сигнала базовой станции сотовой связи на границе соты (35 км)

- Общая идея работы:

разрабатываем принципы создания «проблемных» элементов перспективной когнитивной широкополосной системы приема и обработки сигнала

Структура физ. реализации КШСС