3.5 Silicon Neuron Models

Learning objectives:

In this chapter, I would like to show you how to implement the previously discussed neuron models within electrical circuits. Here, silicon technology offers the unique possibility of extremely energy-efficient integrated circuits. Especially if the transistors are operated at extremely low voltages below the threshold that describes the formation of the inversion channel.

In this chapter I would like to answer the following questions:

- What is the current-voltage characteristics of MOSFETs in the sub-threshold range?
- What is a differential pair and how does it work?
- How can weighted input currents be emulated?
- Why do you use currents instead of voltage?
- What is the Axon-Hillock neuron?
- How can refractory times be implemented in terms of circuitry?

3.5 Silicon Neuron Models

- In the last chapters we got to know the most diverse models for modelling a neuron.
  - We mainly looked at the implementation of these models in software and paid little attention to hardware implementation (apart from the van der Pol oscillator).
  - But in neuromorphics we are interested in hardware: thus, we need a way to implement the presented neuron models in hardware.
  - In the past, we had already seen small circuit models in the most diverse places (LIF neuron, Hodgkin Huxley, …), but how can we build them in hardware?

- Above all, this requires a good technology that enables us to transfer the equations as precisely as possible into a electrical circuit and the technology should also be scalable.
  - This means we are not only interested in one neuron, but especially a neural network that contains many neurons.
  - One such technology is offered by silicon technology - which also forms the origin of the neuromorphic engineering.
  - In particular, the VLSI (Very Large Scale Integration) technology enables the creation of an integrated circuit (IC) by combining thousands of transistors in a single chip and is the backbone of modern CMOS (Complementary Metal Oxide Semiconductor) Technology

In this chapter we want to consider how neurons can be implemented within analogue CMOS circuits.

3.5 Silicon Neuron Models

- As we discussed at the beginning of the semester (hopefully one or the other may remember it) the idea of making custom analogue VLSI implementations of neural networks dates back to the late ’80’s - early ’90s.
  - It was first proposed by carver Mead and builds the starting point for neuromorphic engineering.
  - The relevant Silicon neural network characteristics are:

1. Above threshold (strong inversion)
2. Mixed analog/digital
3. Rate-based
4. Real-time
5. Conductance-based
6. Large-scale, event-based networks
7. Below threshold (weak inversion)
8. Fully analog
9. Spiking
10. Accelerated-time
11. Integrate-and-Fire
12. Small-scale, hard-wired

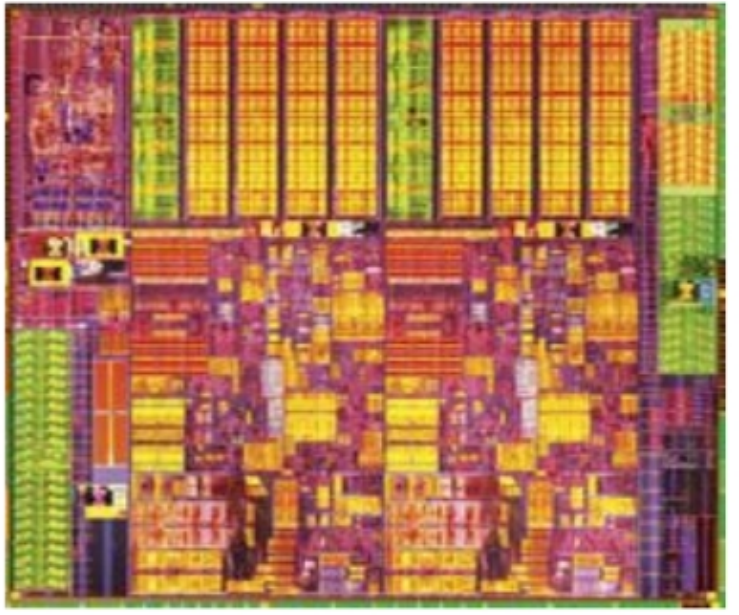
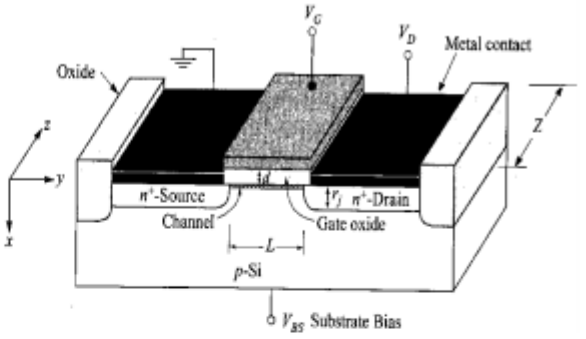


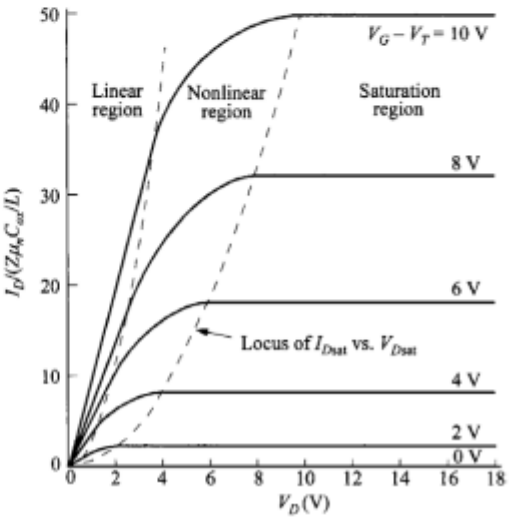
Figure of an integrated an analogue VLSI chip

3.5 Silicon Neuron Models

- Before we start, however, we have to repeat something about the central component of the silicon technology, the MOS transistor and its basic circuits:
  - The MOS transistor is a field effect transistor (FET) and is a 3-terminal device consisting of a channel resistor that connects two of the two connections and a third contact that is used to control the channel resistance (see figure).
  - The highlight of an FET is that the channel resistance is controlled capacitive by an electrical field.

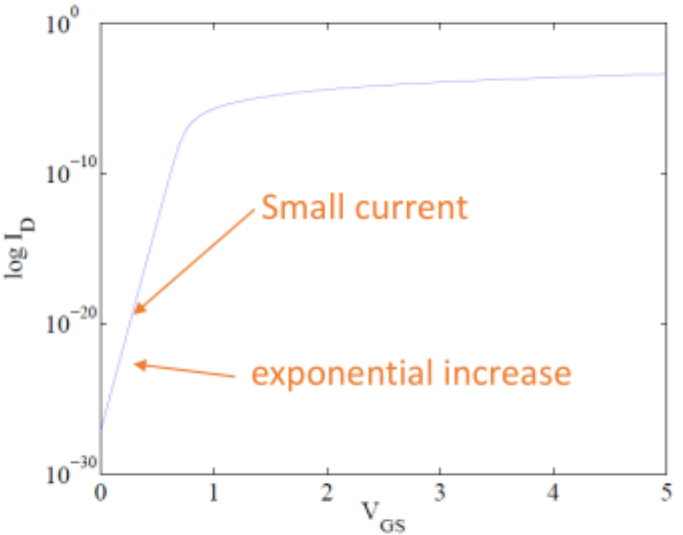


- The transistor works as follows:
  - If the gate voltage is in the "on" range, the transistor is also "on" - current flows from drain to source (electrons from source to drain!) with almost no loss. In other words, the source-drain resistance is small.
  - If the gate voltage is in the "off" range, the transistor is also "off"; no drain-source current flows. In other words, the source-drain resistance is very high.
  - The I-V transfer characteristic of a MOSFETis shown on the right and can be divided into three parts:
    - the drain current first increases linearly with drain voltage (the linear region),
    - then gradually levels off (the nonlinear region),
    - and finally approaching a saturated value (the saturation region).



3.5 Silicon Neuron Models

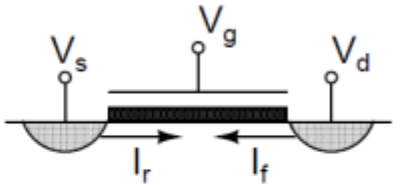
- So far we had not mentioned the operation region of a MOSFET that is particularly important for neuromorphic circuit technology: the sub-threshold region.
  - In the sub threshold region the channel between source and drain has not yet fully developed and the transistor is in a weak inversion (*voltage range between built-in potential and threshold for the experts*).
  - The figure shows the sub-threshold current I_{DS} in dependency of gate voltage V_{GS} on a logarithmic scale.
  - It is not used in today's CMOS technologies, since it owns a strong non-linearity and the charge carrier transport is dominated by diffusion. **But, in neuromorphics we need exactly that!**
  - Further the subthreshold current is very low, means energy consumption is very low. But this also places great demands on the circuit technology: low power and analogue circuits

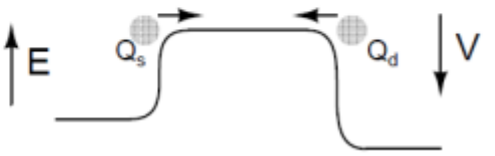


- Since the current is essentially a drift current, it can be calculated similar to the collector current of a bipolar transistor.
  - In the following we want to consider the contributions of the forward and backward diffusion current.

3.5 Silicon Neuron Models

- As shown in the figure opposite, a voltage diffusion of charge carriers into the channel area is created by applying potential (V_g, V_s, V_d) to the 3 terminals of the MOSFET.
- These are two pn junctions: In the example of an n-MOS transistor, doies would be an n-p junction at the source contact and a p-n junction at the drain.
- Without going too much into the details (those who have heard of semiconductor devices know this), we can write the drain-source current over the two drift currents of the pn-diodes at the contacts:



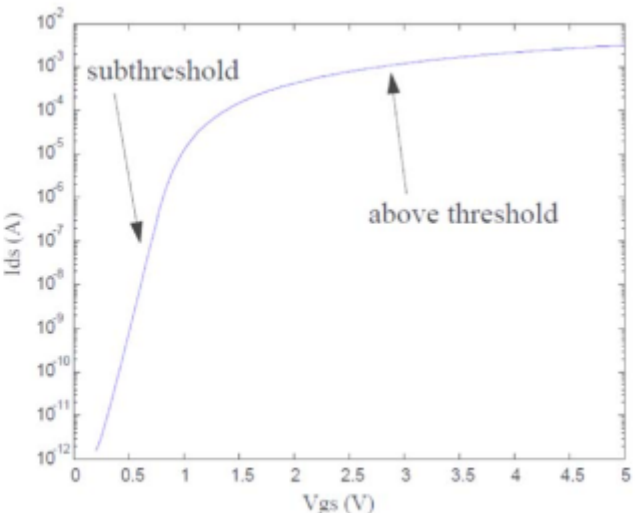


$$I_{DS} = I_F - I_R = I_0 \exp\frac{-\kappa(V_g - V_s)}{k_B T/q} - I_0 \exp\frac{-\kappa(V_g - V_d)}{k_B T/q}$$

$$= I_0 \exp\frac{-\kappa(V_{GS})}{k_B T/q} - I_0 \exp\frac{-\kappa(V_{GD})}{k_B T/q}$$

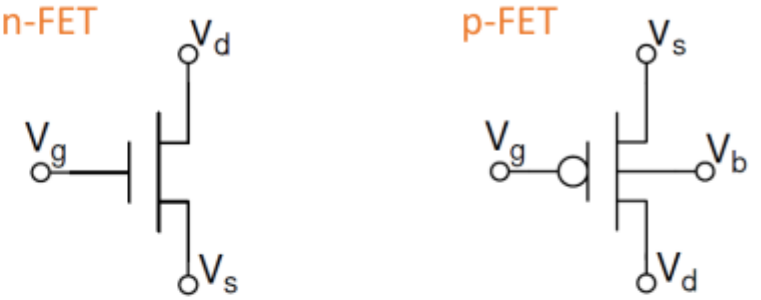
- With I_0 denotes the n-FET current-scaling parameter and κ the n-FET subthreshold slope factor (subthreshold swing).
- For $V_{DS} > k_B T/q$ the revers current I_R is negligible and the current is only described by the forward current of a n-FET:

$$I_{DS} = I_0 \exp\frac{-\kappa q(V_{GS})}{k_B T}$$



3.5 Silicon Neuron Models

- The symbols for the two MOSFET basic types –p-FET and n-FET- are shown in the figure.
- Hopefully everyone knows the difference between both types: n-FET is initially off and the p-FET is initially on…
- In traditional CMOS circuits, all n-FETs have the common bulk potential (Vb) connected to Ground (Gnd), and all p-FETs have a common bulk potential (typically) connected to the power supply rail (Vdd).
- The corresponding (complementary) equation for the p-FET is:



$$I_{DS} = I_0 \exp\frac{-\kappa_p(V_{dd} - V_g)}{k_B T/q}\left(\exp\frac{-(V_{dd} - V_s)}{k_B T/q} - \exp\frac{-(V_{dd} - V_d)}{k_B T/q}\right)$$

- Next we want to look at the essential basic circuits, i.e. subthreshold MOSFET analogue circuits. An essential basic type is the differential pair which is used in many other circuits:
  - A differential pair is used whenever voltages or currents are compared and builds the basis of integrated-circuit amplifier.
  - This circuit is the basis of many more complex circuits. Let's see how it works on the next slide….

3.5 Silicon Neuron Models

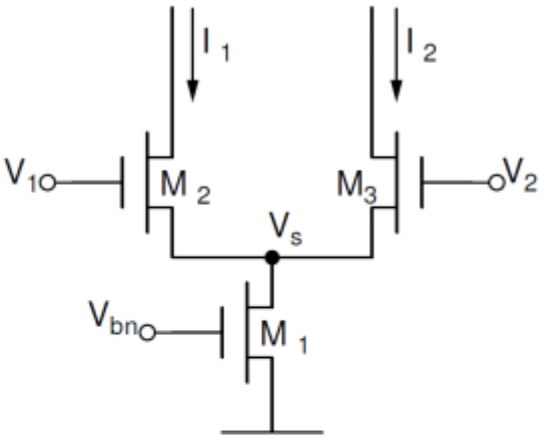
- The simplest circuitry of a differential pair is shown in the figure.
  - Now our goal is to determine the dependencies of the two input currents I_1 and I_2 on the difference in the input voltages $(V_1 - V_2)$.
  - Using Kirchhoff's law:

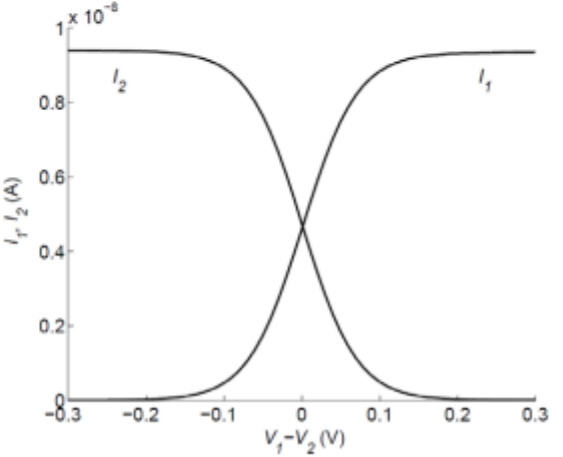
$$\begin{cases} I_1 = I_0 \exp\dfrac{-\kappa q(V_1 - V_s)}{k_B T} \\ I_2 = I_0 \exp\dfrac{-\kappa q(V_2 - V_s)}{k_B T} \\ I_b = I_0 \exp\dfrac{-\kappa q(V_{bn})}{k_B T} \end{cases}$$

With $I_b = I_1 + I_2$

$$\Rightarrow \exp\frac{-\kappa q(V_s)}{k_B T} = \frac{I_b}{I_0} \frac{1}{\exp\dfrac{-\kappa q(V_1)}{k_B T} + \exp\dfrac{-\kappa q(V_2)}{k_B T}}$$

$$I_1 = I_b \frac{\exp\dfrac{-\kappa q(V_1)}{k_B T}}{\exp\dfrac{-\kappa q(V_1)}{k_B T} + \exp\dfrac{-\kappa q(V_2)}{k_B T}} \quad \text{and} \quad I_2 = I_b \frac{\exp\dfrac{-\kappa q(V_2)}{k_B T}}{\exp\dfrac{-\kappa q(V_1)}{k_B T} + \exp\dfrac{-\kappa q(V_2)}{k_B T}}$$





  - The resulting course of the two currents as a function of the voltage differences is shown in the figure.
  - The ratio of the input currents I_1/I_2 only depends on the difference in the input voltages $(V_1 - V_2)$. If this difference is zero, then we get an ideal current source, both currents are of equal size.

3.5 Silicon Neuron Models

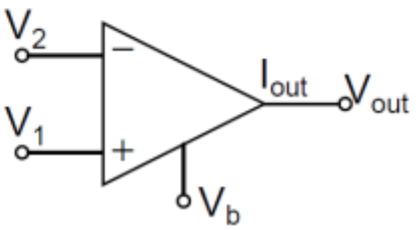
- Well, in order to examine the current gain a little more closely, it makes sense to do the following transformation:

$$\frac{I_1}{I_2} = \exp\frac{-\kappa q(V_1 - V_2)}{k_B T} \text{ with } I_1 + I_2 = I_b \Rightarrow I_2 = \frac{I_b}{\exp\frac{-\kappa q(V_1 - V_2)}{k_B T} + 1} \approx I_b \tanh\left(\frac{\kappa q}{2k_B T}(V_1 - V_2)\right) = I_b \tanh\left(\frac{\kappa}{2V_T}(V_1 - V_2)\right)$$

  - Since the decisive variable is the ratio $\frac{(V_1 - V_2)}{2V_T}$, there is only a small modulation range.
  - Even with an argument of 2, the tangent hyperbolic provides a function value of approximately 0.96. This means that at a differential voltage of $4 \cdot V_T$, almost all of the current supplied by the current source flows through one transistor and the other transistor is therefore almost currentless.
  - In the linear region, the *tanh* function can be simplified further:
  - So we have therewith a tunable conductance and that is would we needfor a neuron circuit.

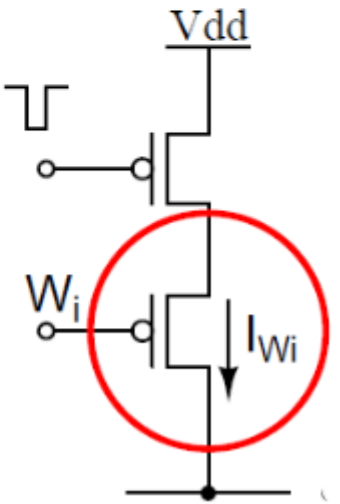
$$I_2 = I_{out} \approx \frac{I_b \kappa q}{2k_B T}(V_1 - V_2) = g_m(V_1 - V_2)$$

- I hope most of you already know this from the basic lectures:
  - What we have is an operational amplifier or also named as integrated-circuit amplifier , which is usually represented by the symbol on the right.
  - typically the internal structure of an operational amplifier is a bit more complex (load resistances, negative feedback, ...), but we want to leave it at this level and go to silicon neurons.



3.5 Silicon Neuron Models

- But before maybe a few reasons why this integrated-circuit amplifier is so important for a subthreshold IC design:
  - Differential amplifiers apply gain not to one input signal but to the difference between two input signals. This means that a differential amplifier naturally eliminates noise or interference that is present in both input signals.
  - Differential amplification also suppresses common-mode signals—in other words, a DC offset that is present in both input signals will be removed, and the gain will be applied only to the signal of interest (assuming, of course, that the signal of interest is not present in both inputs).
  - Last but not least, IC amplifiers allows a straight forward realisation of differentiators, integrators, comparators and so on. Which we need to build up analogue ICs.
- So now let's build a neuron model in hardware. For this we remember again the important functions of a neuron:
  - we need a weighted input current (ion channel currents of the synapse), an integration of the input current (membrane potential) and a spike generator.
  - weighted inputs w_i can be created by using a single transistor as shown in the figure.
    - p-FETs implement excitatory synapse –increase in synaptic coupling strength.
    - n-FETs implement inhibitory synapses –decrease in synaptic coupling strength.
    - The second transistor above is for mimic a pulsed input signal. With the pulse duration we can shape the input signal…
  - Thus, with that simple circuit a synaptic weight can be set by changing the W_i bias voltage.



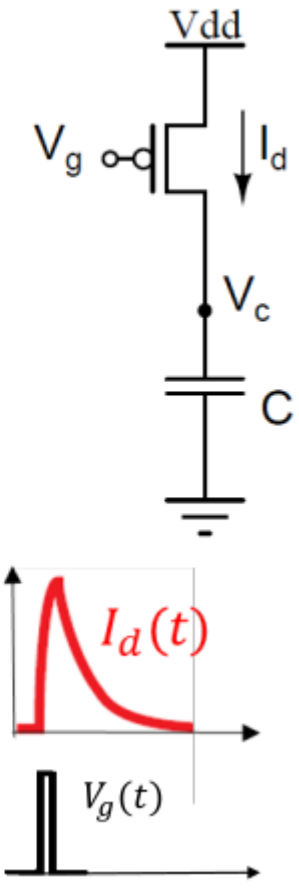
3.5 Silicon Neuron Models

- A simple and common implementation of a voltage to current conversion, is shown in the circuit on the right. The trick here is to use currents and not voltages, since voltages are much more difficult to handle in a electronic circuit.
  - Thus, the circuitry is used to mimic the ionic currents through the ion channels and serve as the input current for the voltage integration in the neuron.
  - Using the subthreshold current we can write:

$$I_d(t) = I_0 \exp \frac{-\kappa q\left(V_{dd} - V_g(t)\right)}{k_B T}$$

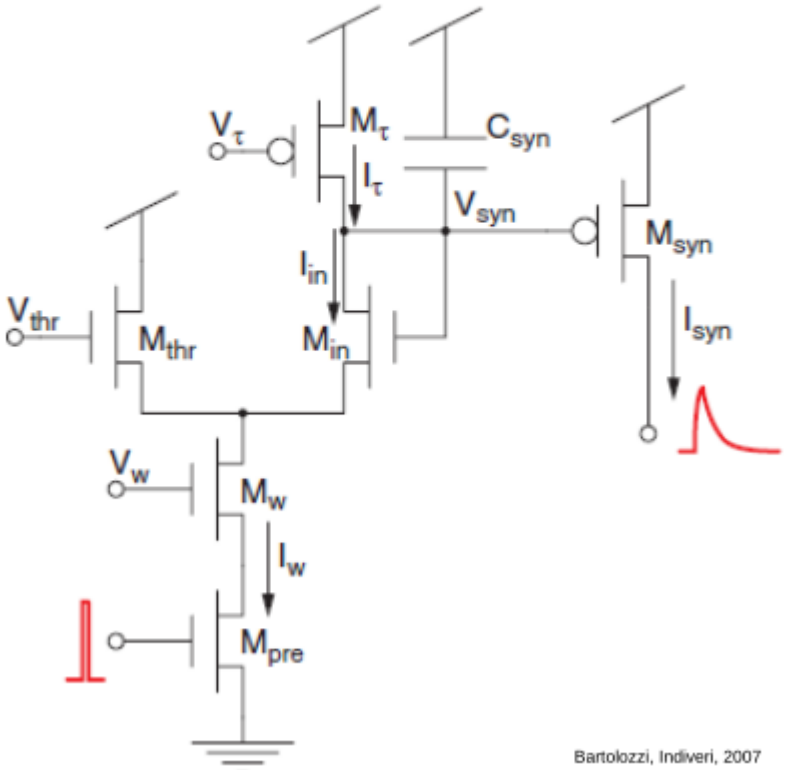
  - Thus, with the subthreshold current we get a linear low-pass filter with a impulse response that decays exponentially which contains the nonlinearity and leakage behaviour observed for post-synaptic potential in a synapse.
  - But let see this in detail: a given input pulse $V_g(t)$ applied via the gate terminal of the p-FET the capacitor leads to an integration of the input pulse according to

$$I_d(t) = C\frac{d}{dt}V_C \qquad \Rightarrow V_C = \frac{1}{C}\int I_d(t)dt$$



3.5 Silicon Neuron Models

- Thus, the part of the input current shaping is a relatively important part in the neuron circuits, because we do not realize all possible non-linearities through a non-linear voltage function as with the models before, but rather through the shape of the input current.
  - Therefore, the modelling of the input current is very important as it defines the non-linearity for the voltage integration, which we now is important for e realistic emulation of a neuron (remember on the last chapters).
  - Hence, there are many efforts made to optimize the input current by increasing the circuit complexity. In particular, the differential-pair integrator (DPI) providing a very wide range of functionalities. We want to take a closer look at this in the following.
  - The circuit scheme of a DPI is shown in the figure.
  - That looks a little scary at first glance, but it's not that bad. In the following we take a look at the most important parts of the circuit.

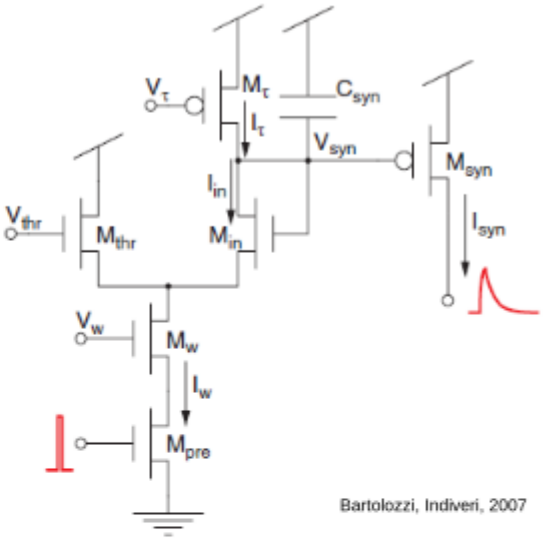


Bartolozzi, Indiveri, 2007

3.5 Silicon Neuron Models

- The circuits include a capacitor C_{syn} for synaptic integration, which is charged by the following current:

$$I_C(t) = C_{syn}\frac{d}{dt}\left(V_{dd} - V_{syn}(t)\right)$$



Bartolozzi, Indiveri, 2007

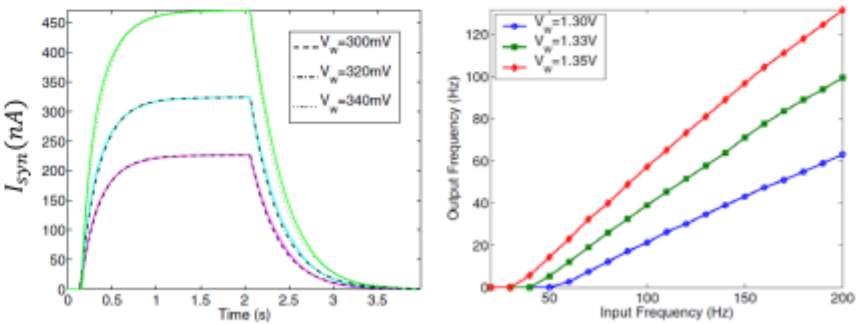
- V_{syn} is defined by the input current I_{in} which is given by I_τ (V_τ input pulse) and modulated via the differential pair by the weight cuurent I_w:

$$\left.\begin{aligned} I_w &= I_0 \exp\frac{\kappa q(V_w)}{k_B T} \\ I_\tau &= I_0 \exp\frac{\kappa q(V_{dd} - V_\tau)}{k_B T} \end{aligned}\right\} \Rightarrow I_{in} = I_w \frac{\exp\frac{\kappa q(V_{syn})}{k_B T}}{\exp\frac{\kappa q(V_{syn})}{k_B T} + \exp\frac{\kappa q(V_{thr})}{k_B T}}$$

- This results in the following relation for the synapse current:

$$I_{syn} = I_0 \exp\frac{\kappa q(V_{dd} - V_{syn})}{k_B T} \qquad \Rightarrow \frac{d}{dt} I_{syn} = -\frac{\kappa q}{k_B T} I_{syn} \frac{d}{dt} V_{syn}$$

  - this leads to a different current profile depending on the weighting through V_w as shown in the figure, which models the output frequency of the current pulse (see figure).



G. Indiveri, ICANN09

3.5 Silicon Neuron Models

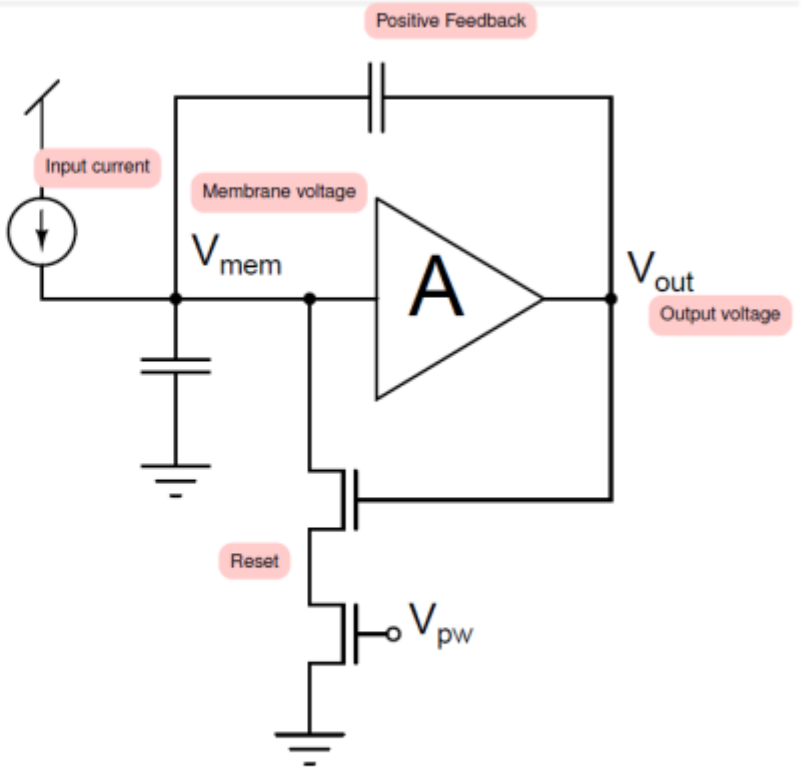
- Using suitable external signals, this circuit allows a variety of dynamic properties of the neurons to be simulated, such as adaptation and time-dependent synaptic learning, which we will consider in the next chapter.

- The next stage that we have to build in order to have a functional neuron in hardware is the current integration and the spike generation.
- Therefore, we can use our gained knowledge about phenomenological neuron models:
- What we need to put in hardware is a integrate and fire neuron (I&F neuron):
  - An I&F neuron integrates weighted charge inputs triggered by presynaptic action potentials.
  - If the integrated voltage reaches a threshold, the neuron fires a short output pulse and the integrator is reset.
  - An elegant implementation of an integrate and fire neuron of has been proposed by Carver Mead and is also known as Axon-Hillock circuit.
  - Let's take a look on that circuit…

3.5 Silicon Neuron Models

- The complete circuit of the I&F neuron is shown in the figure.
  - It consists of a integrating capacitor (representing the membrane capacitance of a neuron),
  - a simple high gain amplifier (we will see it later..) that switches as the integrated input current exceeds its threshold voltage
  - and a feedback capacitor that adds extra charge to the membrane capacitor to stabilize the firing state and to avoid oscillations around the switching point.
  - The reset is achieved through a leakage that is turned on while the output is high. That leakage current determines the pulse length of the output pulse. As the switching threshold is reached a new during the reset, the feedback capacitance kicks in again and stabilizes the idle state of the neuron.

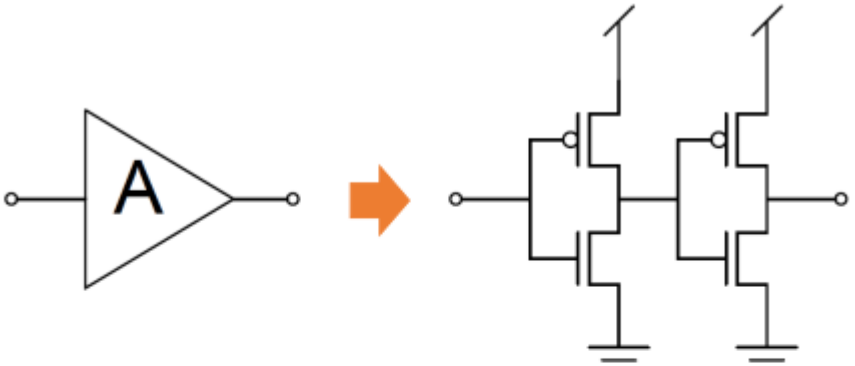
- The circuit is not that easy to understand and needs a little more explanation.
  - For this we want to take a look to the different components of that circuit in the following...



DOI: 10.3389/fnins.2011.00073

3.5 Silicon Neuron Models

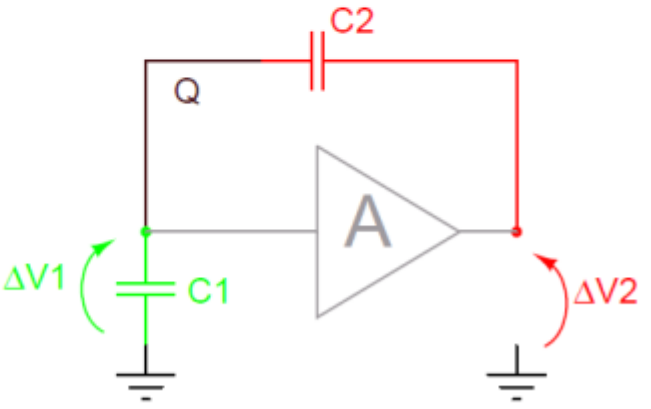
- The voltage amplification is realized in the circuit by 2 inverting CMOS amplifiers as shown in the picture.
  - This offers a simple and efficient way to simulate threshold behaviour and spike generation.
  - However, this type of implementation has a major disadvantage. The power consumption is relatively large.



- Next we look at the capacitive divider that comes from the capacity for voltage integration and from that from the positive feedback.
  - The question is the connection between ΔV_1 and ΔV_2:

$$Q = C_1 V_1 + C_2(V_1 - V_2) = \text{const.}$$

$$C_1 \Delta V_1 + C_2(\Delta V_1 - \Delta V_2) = 0 \qquad \Rightarrow \Delta V_1 = \frac{C_2}{C_1 + C_2} \Delta V_2$$

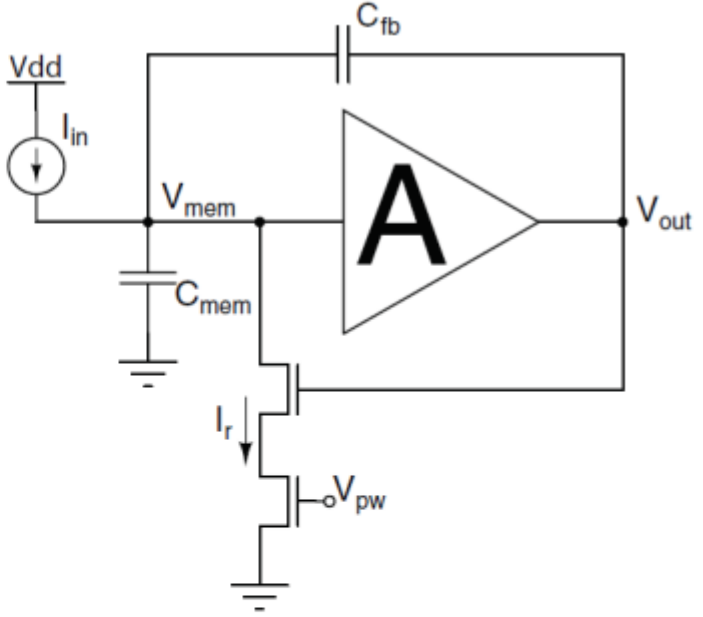


DOI: 10.3389/fnins.2011.00073

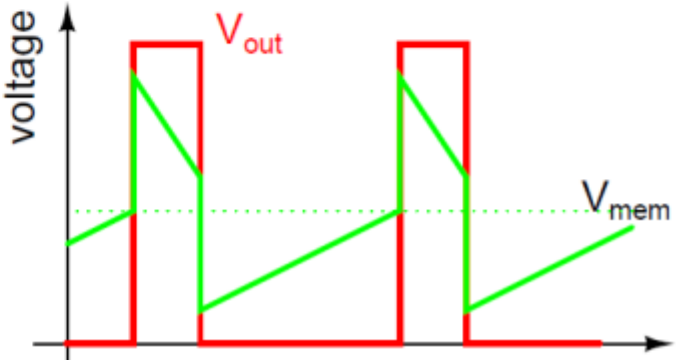
3.5 Silicon Neuron Models

- Now let's take a look to the function:
  - Input currents I_{in} (corresponds to I_{syn}) are integrated on the membrane input capacitance C_{mem}, and the analogue voltage V_{mem} increases linearly
  - until it reaches the amplifier switching threshold (see Figure). At this point V_{out} quickly changes from 0 to V_{dd}, switching on the reset transistor and activating a positive feedback through the capacitor divider implemented by C_{mem} and the feedback capacitor C_{fb}.

$$\Rightarrow \Delta V_{mem} = \frac{C_{fb}}{C_{mem} + C_{fb}} V_{dd}$$



  - If the reset current set by V_{pw} is larger then the input current, the membrane capacitor is discharged, until it reaches the amplifier's switching threshold again.
  - At this point V_{out} swings back to 0 and the cycle repeats.



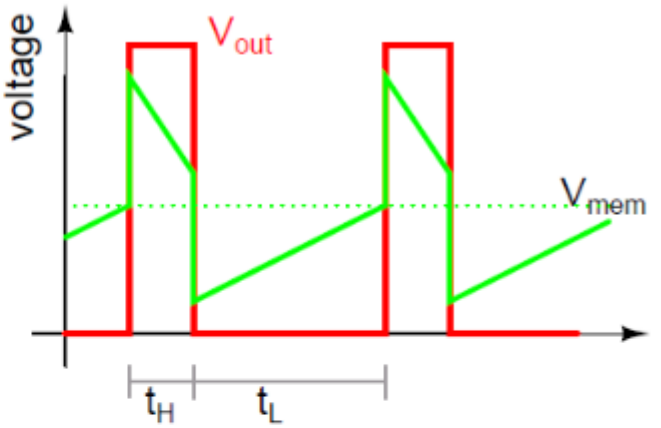
3.5 Silicon Neuron Models

- The inter-spike interval t_L is inversely proportional to the input current:

$$t_L = \frac{C_{mem} + C_{fb}}{I_{in}} \Delta V_{mem} = \frac{C_{fb}}{I_{in}} V_{dd}$$

- The pulse duration period t_H depends on both the input and reset currents:

$$t_H = \frac{C_{mem} + C_{fb}}{I_r - I_{in}} \Delta V_{mem} = \frac{C_{fb}}{I_r - I_{in}} V_{dd}$$



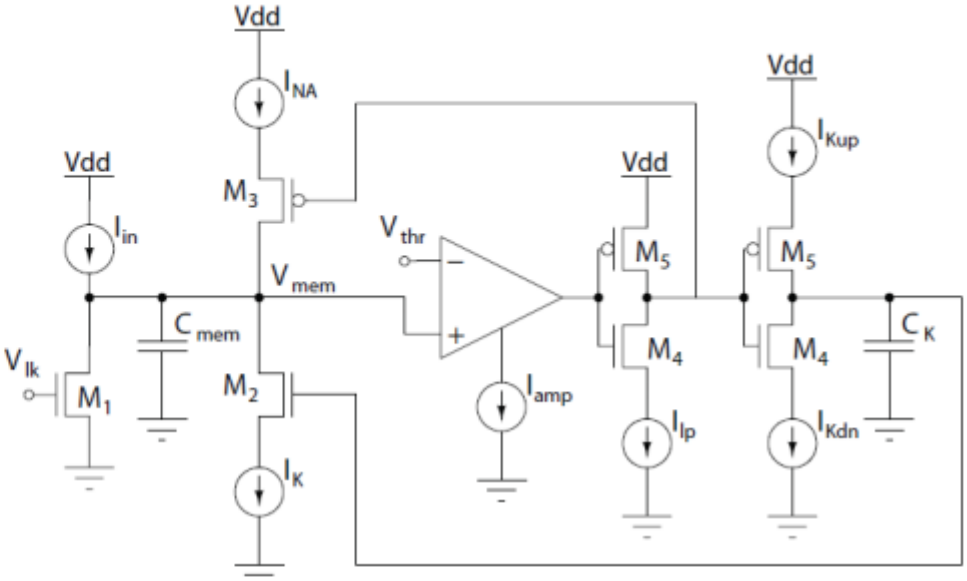
DOI: 10.3389/fnins.2011.00073

- This means that the frequency is proportional to I_{in}, while the pulse width is defined by $\frac{1}{I_r}$ for $I_r \gg I_{in}$.

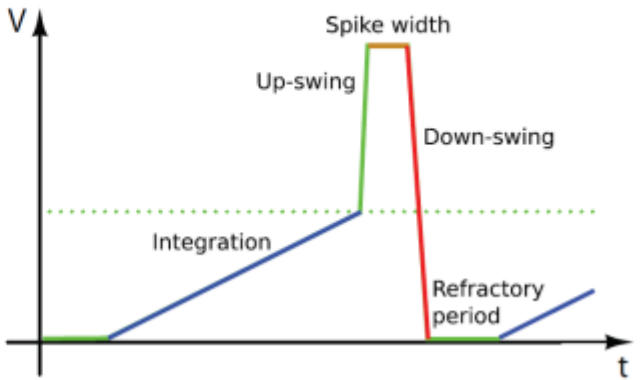
- What is missing in the model is a refractory period.
- Furthermore, a better control of the threshold voltage is required for the application.
  - Here, however, the model can be expanded, which we want to consider below.

3.5 Silicon Neuron Models

- let's consider the following circuit:
  - The capacitance C_{mem} is that of a membrane of a biological neuron, while the membrane leakage current is controlled by the gate voltage V_{lk} of an nFET.
  - Excitatory inputs (I_{in}) add charge to C_{mem}.
  - If I_{in} is larger than the leakage. than V_{mem} will increase from its resting potential.
  - The voltage V_{mem} is compared with the threshold voltage V_{thr}, using a basic transconductance amplifier.
  - If V_{mem} exceeds V_{thr}, an action potential is generated.
  - Therefore, an increased sodium conductance creates the upswing of the spike,
  - while a delayed increase of the potassium conductance creates the downswing.
  - This leads to the curve shown on the right:

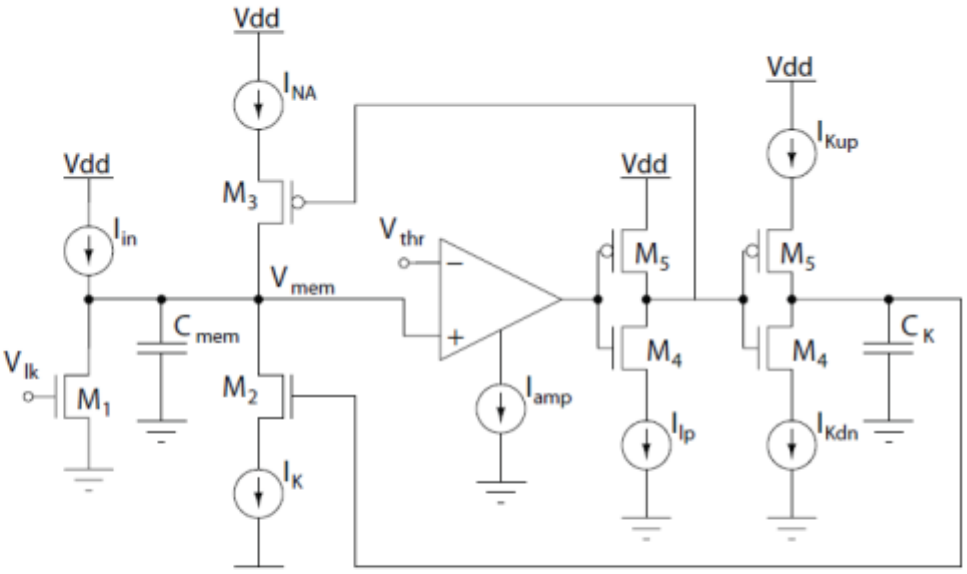


DOI: 10.3389/fnins.2011.00073



3.5 Silicon Neuron Models

- In the circuit this is implemented:
  - $V_{mem} > V_{thr}$: output voltage of comparator will rise to V_{dd} and the output of the following inverter will be low
  - This allows the *sodium current* I_{Na} to pull up the membrane potential.
  - The second inverter will allow the capacitance C_K to be charged at a speed which can be controlled by the current I_{Kup}.
  - When C_K is high, the *potassium current* I_K will be able to discharge the membrane capacitance through M_2.
  - The current I_{Kup} controls the spike width.
  - If $V_{mem} < V_{thr}$, the output of the first inverter will become high, cutting off the current I_{Na}.



DOI: 10.3389/fnins.2011.00073

  - the second inverter will then allow C_K to be discharged by the current I_{Kdn}.
  - I_{Kdn} controls the refractory period of the neuron: If I_{Kdn} is small, the voltage on C_K will decrease only slowly, and, as long as this voltage stays high enough to allow I_K to discharge the membrane, it will be impossible to stimulate the neuron