Analogue building blocks for neural-inspired circuits

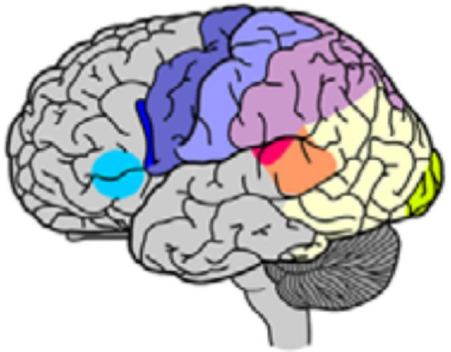
Steve Hall
University of Liverpool, s.hall@liv.ac.uk

Liam McDaid
University of Ulster, lj.mcdaid@ulster.ac.uk

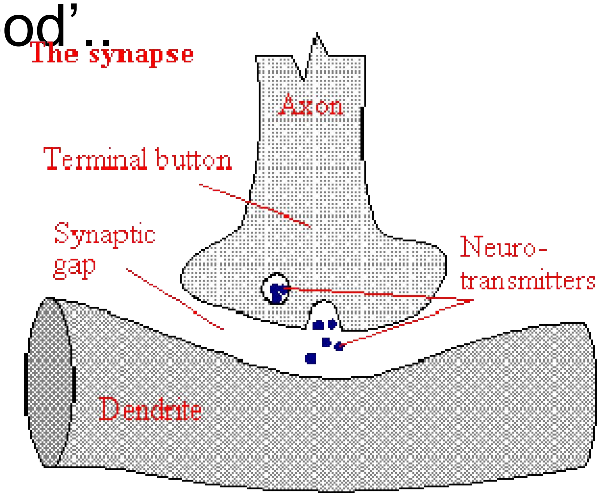
UNIVERSITY OF LIVERPOOL

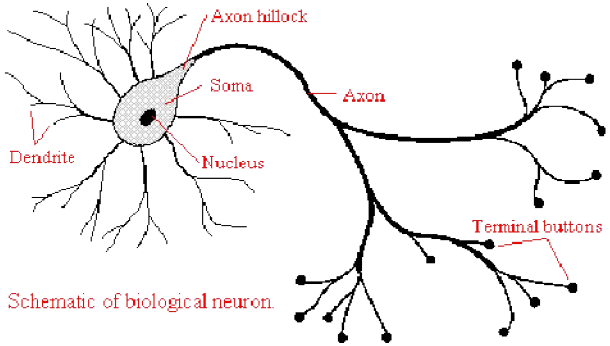
Some facts about the brain as a PC...

- The brain has ~100 billion neurons (10^{11}) – about 30µm large
  - Neuron Fan-in ~ 10^3 – 10^4 (logic gates 2-4!)
  - complex dynamics - includes several time constants,
  - maintains a more complex internal state
  - output is a time-series of action potentials or ‘spikes - no information in amplitude!



- Massively parallel in nature
  - Typical 10^{15} interconnections
  - Total computation rate of about 10^{16} complex operations /sec (cf 10 P-FLOPs)
- Millisecond time frame of ‘events’
- Low level function: ‘reasonably well understood’..
- High level function.....................???????





Schematic of biological neuron

Some other brains

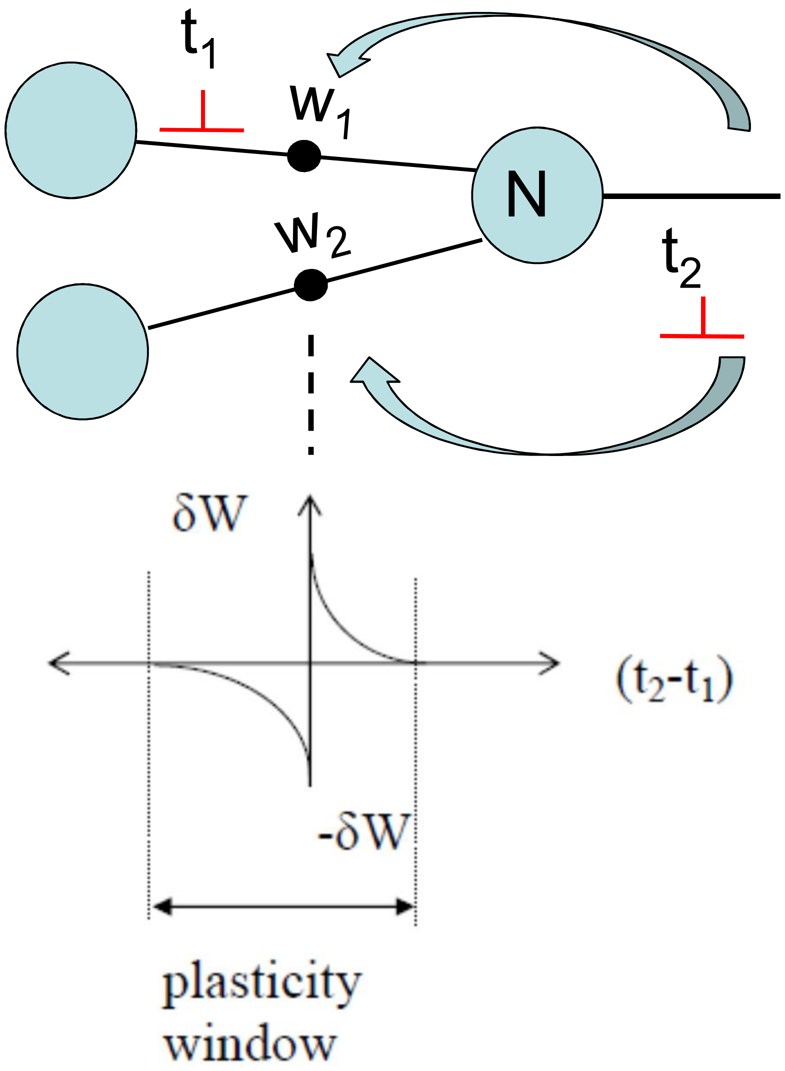
- A fly (1 grain of sugar a day to feed it!): 250 k neurons
- Honeybee (fantastic navigator!): 1 million neurons
- Rat (pretty smart animal): 55million neurons

- But how do the following work:
  - the arithmetic
  - Fault-tolerance
  - The parallelism (beat Moore's Law hands down)

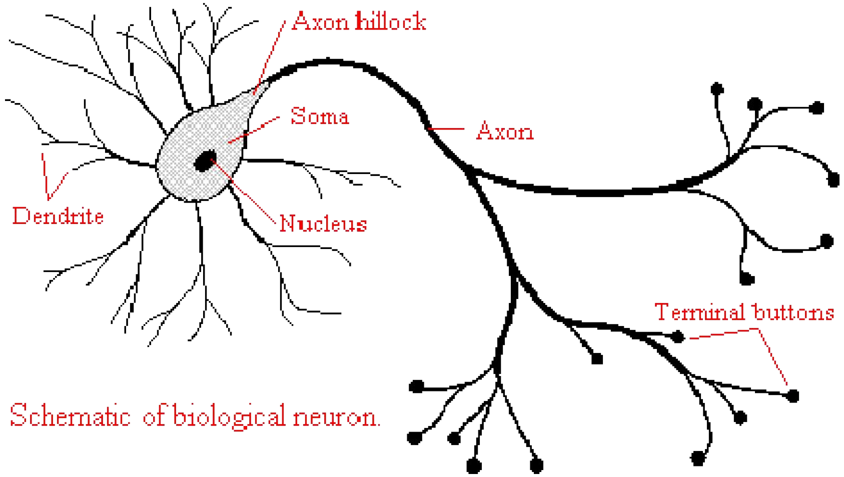
This is the inspiration!
But must find a simpler, scaleable, low power approach

Synapses and neurons

Spike-timing dependent plasticity



STDP learning rule



Schematic of biological neuron.

If spike, t_1 causes neuron, N to fire ($t_2 – t_1$ small)..

Weight W_1 may be increased

and W_2 etc decreased

Motivation

Create building blocks that can emulate biological functionality

Implement in mixed signal CMOS (cheap!)

Assess layout / scalability / systems functionality

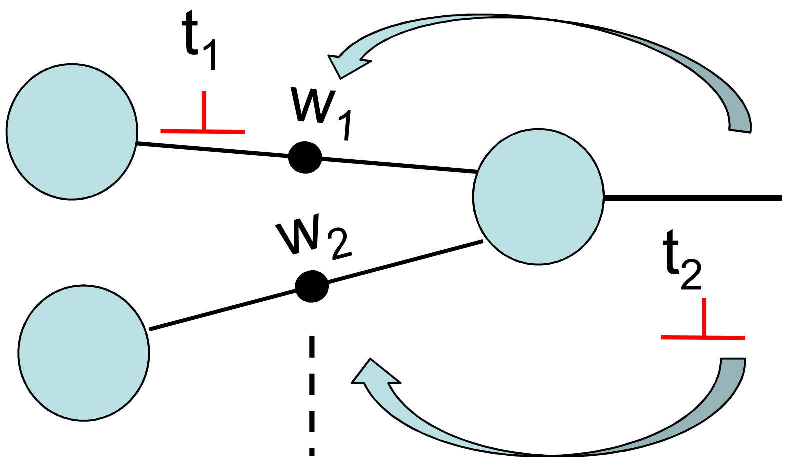
Circuits that can learn!

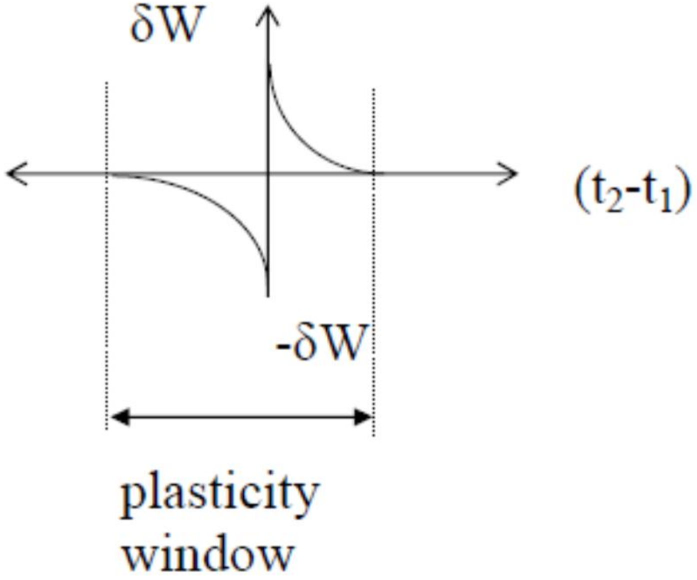
- Plasticity / decision circuits (STDP) / FG weight storage

→ Build large, useful electronic systems
learn more about ‘brain computation’ .....

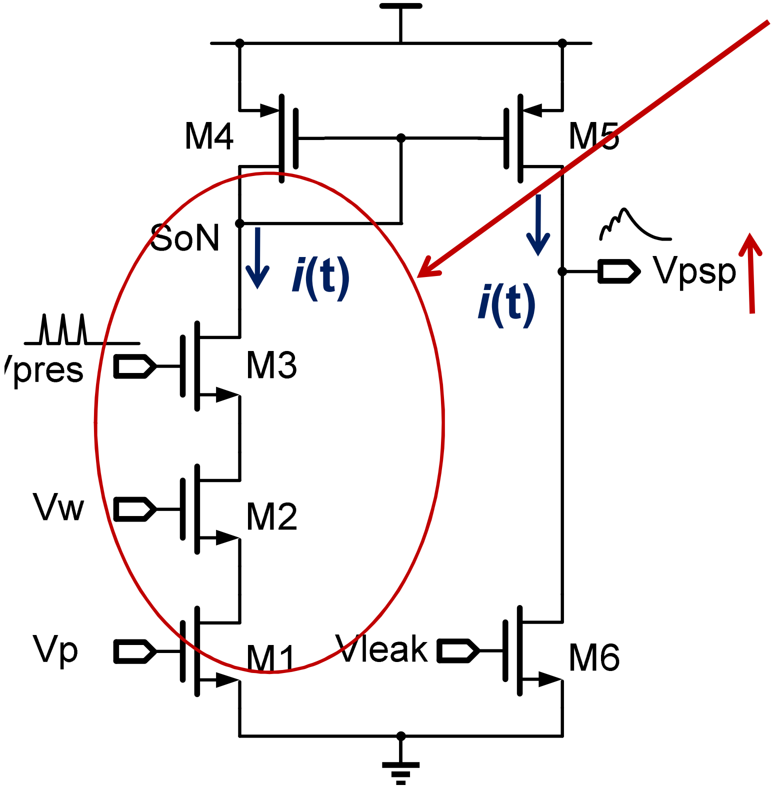
Circuit Challenges

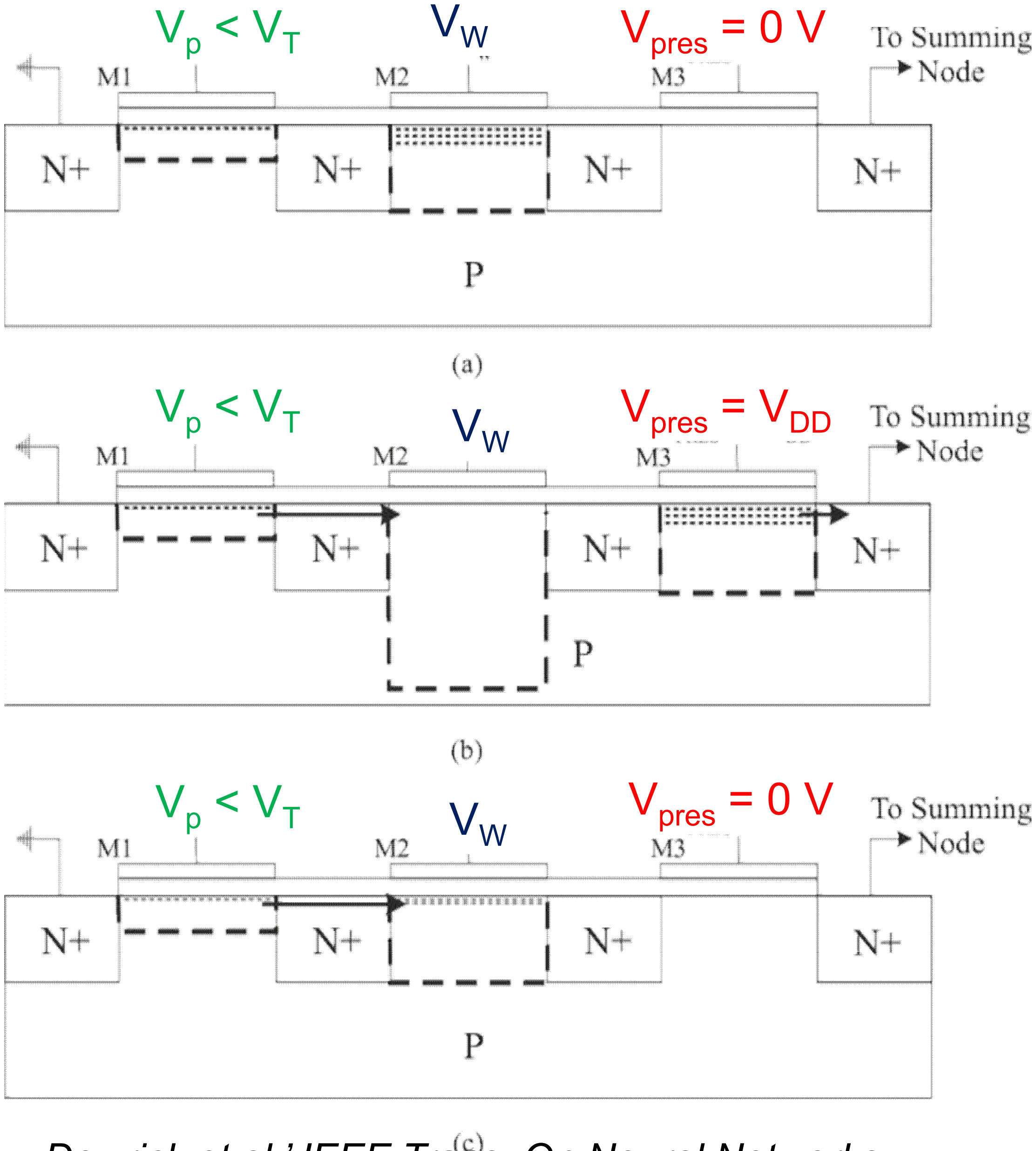
- Store and update weights
- Detect timing ($t_2 - t_1$)
- Axonal delay
- Low power operation
- Scale to VLSI
- Learn!





Dynamic synapse

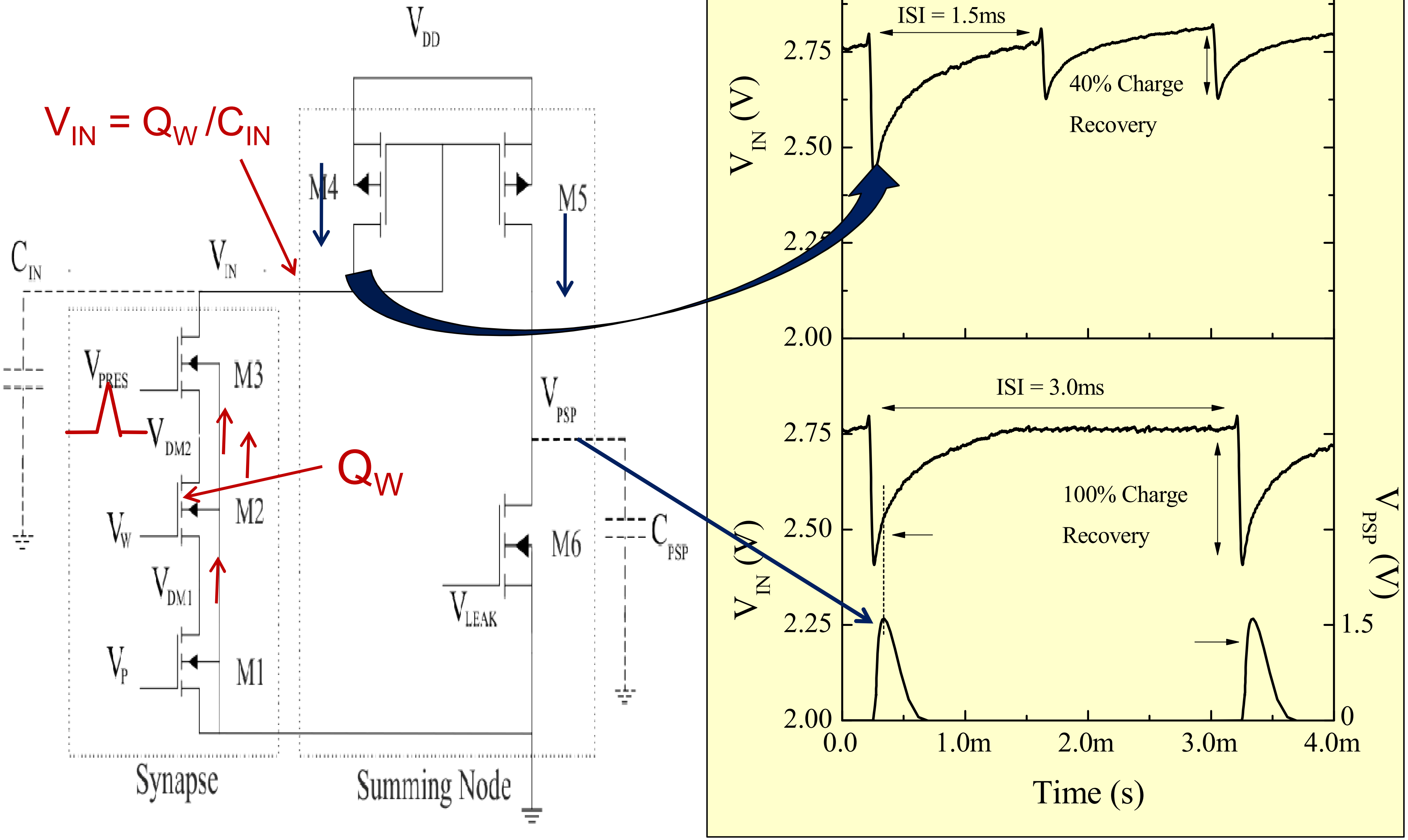




Vpres → on
Charge sharing
S of M2 increase
→ M2 clamped ‘off’
Transient *i*, mirrored in M5

Dowrick et al.’ IEEE Trans. On Neural Networks and Learning Systems, 23(10), p.1513 (2012)

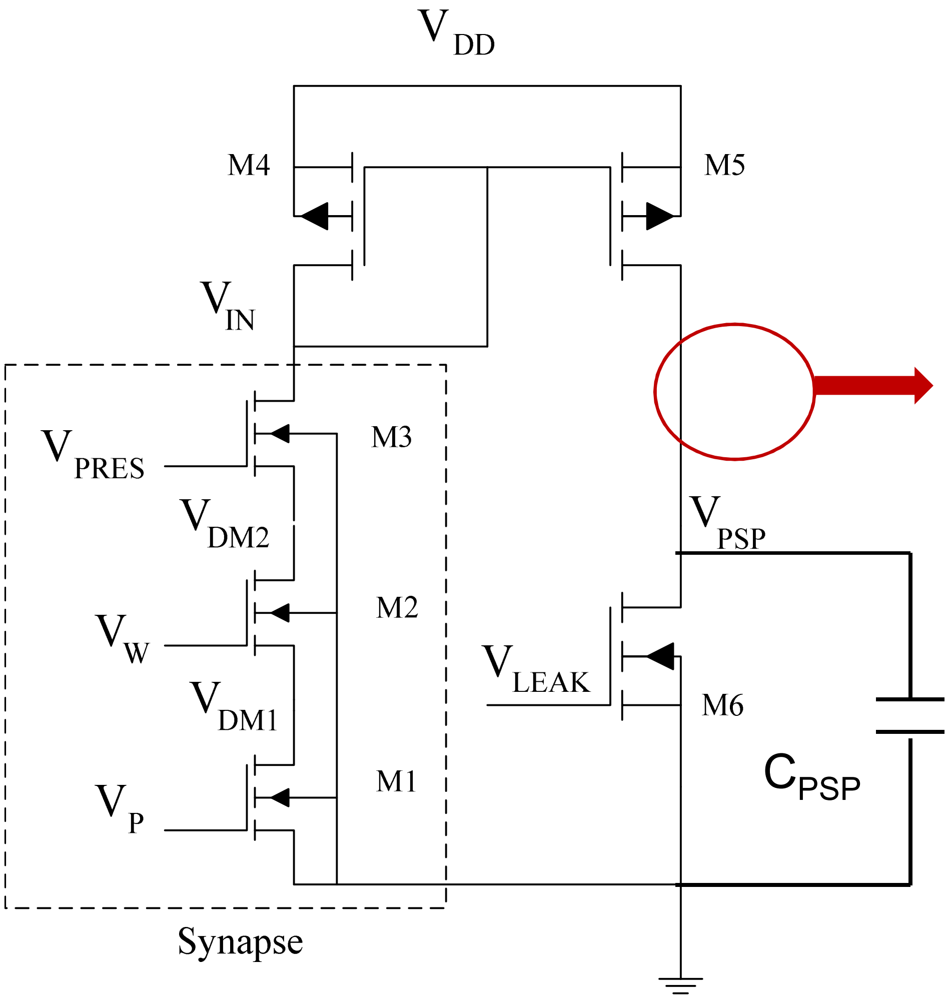
How it works



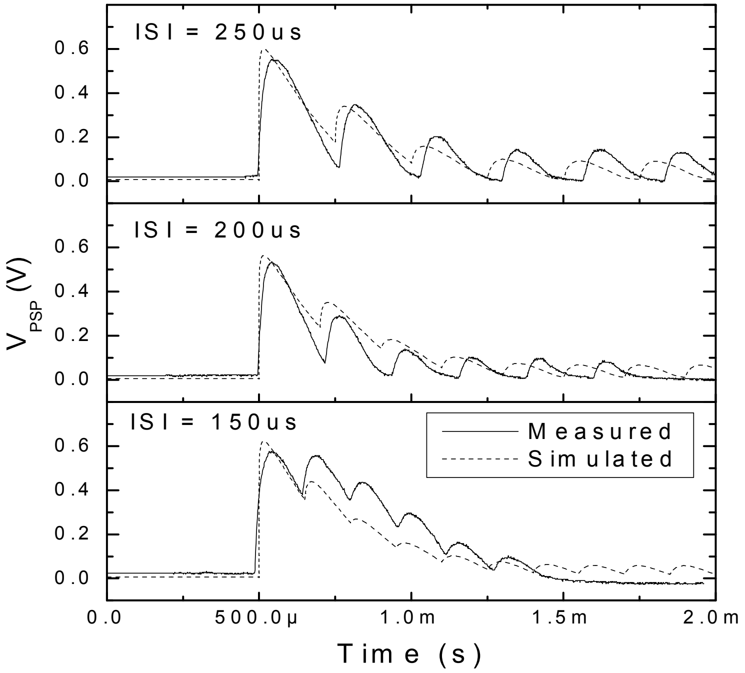
$V_{IN} = Q_W / C_{IN}$

Q_W

Post-synaptic potentials



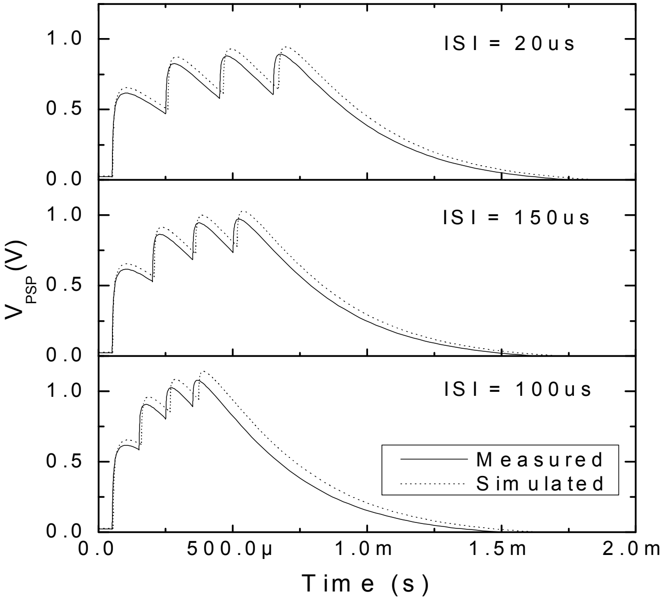
ISI



250us

200us

150us



20us

150us

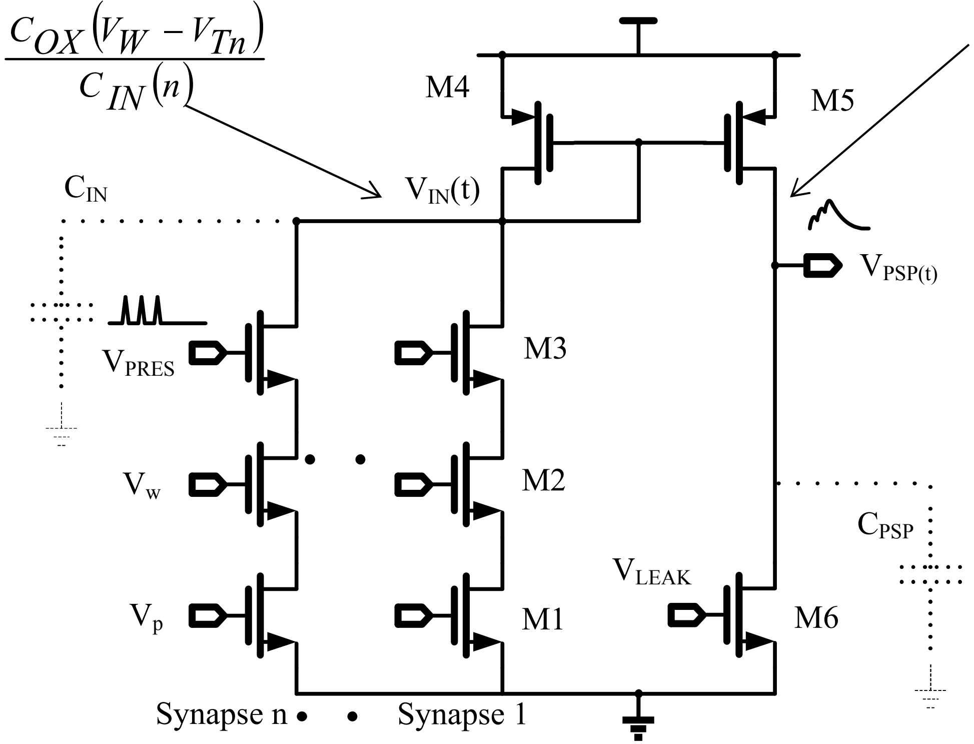
100us

Fan-in: theory

Consider transients of capacitive nodes

$$\Delta V_{IN}(t) = \frac{Q_W}{C_{IN}(n)} = \frac{C_{OX}(V_W - V_{Tn})}{C_{IN}(n)}$$

$$C_{PSP}\frac{dV_{PSP}(t)}{dt} = I_{Op}\exp\left(\frac{V_{IN}'(t)}{m_p V_t}\right) - I_{M6}$$



Rise time

$$\tau_R = \left[\frac{I_{Op}}{m_p V_t C_{IN}}\right]^{-1}\left\{\frac{I_{On}}{I_{Op}}\exp\left(-\frac{V_{LEAK}}{m_n V_t}\right) - \exp\left(-\frac{V_{IN}'(0)}{m_p V_t}\right)\right\}$$

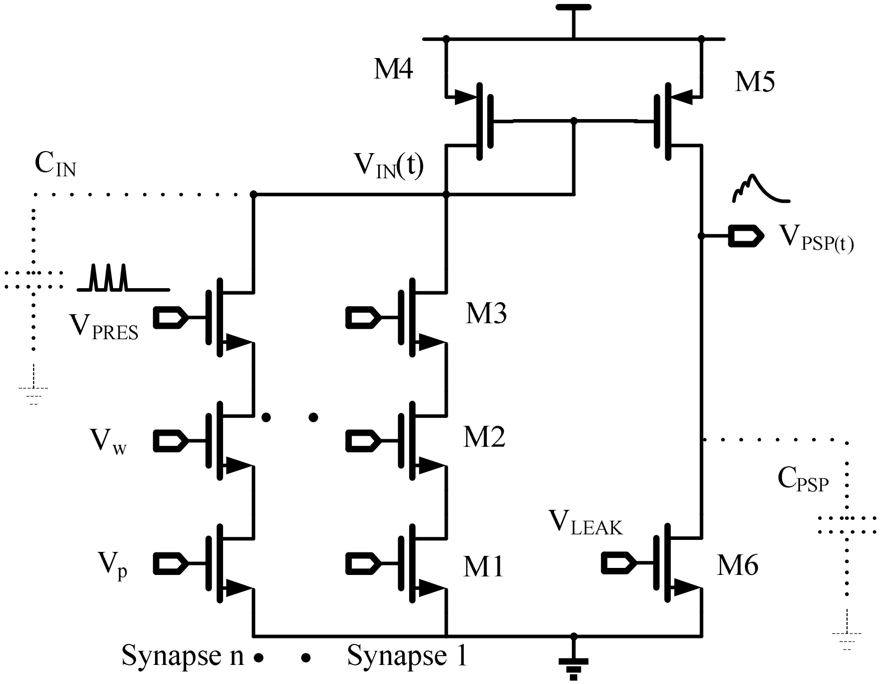
Post-synaptic potential

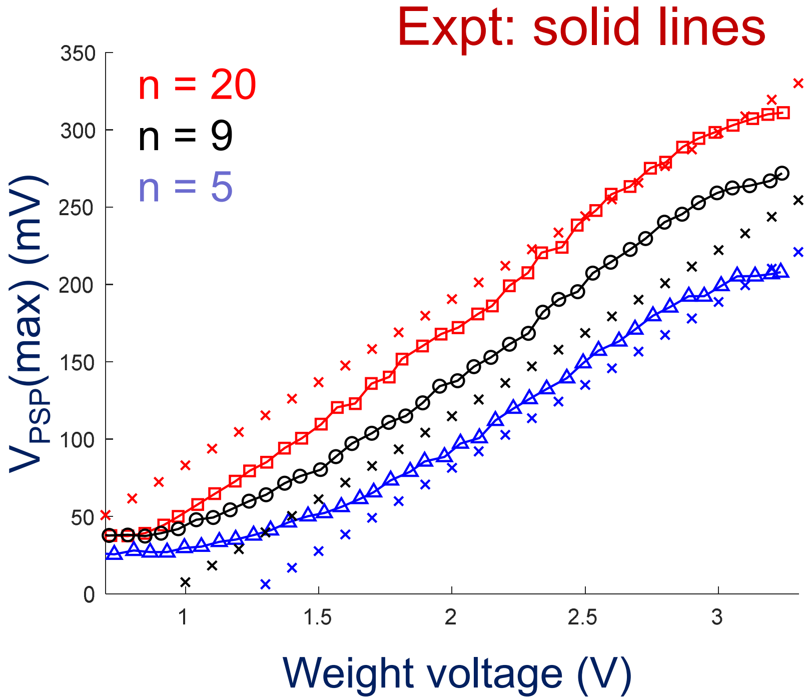
$$V_{PSP}(t) = m_p V_t \frac{C_{IN}}{C_{PSP}} \ell n\left[1 + \frac{I_{Op}}{m_p V_t C_{IN}}\exp\left(\frac{V_{IN}'(0)}{m_p V_t}\right)t\right] - \frac{I_{M6}}{C_{PSP}}t$$

Dowrick et al, Neurocom., vol.314, pp.78-85, https://doi.org/10.1016/j.neucom.2018.06.065 (2018)

Fan-in

$$V_{PSPMAX} = m_p V_t \frac{C_{IN}(n)}{C_{PSP}} \ell n \left[1 + \frac{I_{Op}}{m_p V_t C_{IN}(n)} exp\left(\frac{V_{IN}{}'(0)}{m_p V_t} \right) \tau_R \right] - \frac{I_{M6}}{C_{PSP}} \tau_R$$





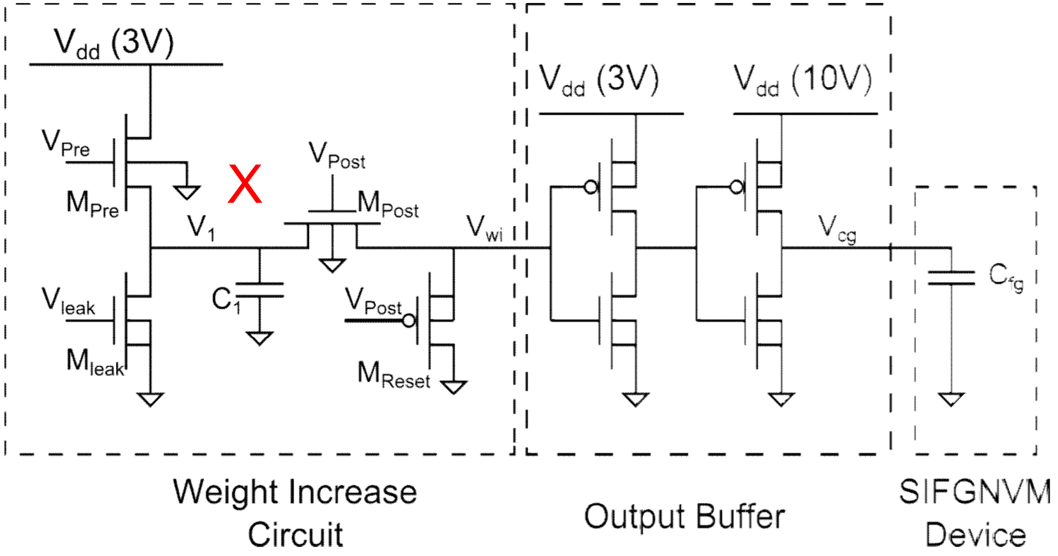
Conclusion: Fan-in intrinsic limit > 10^5 !

Practical limit is set by layout / interconnect

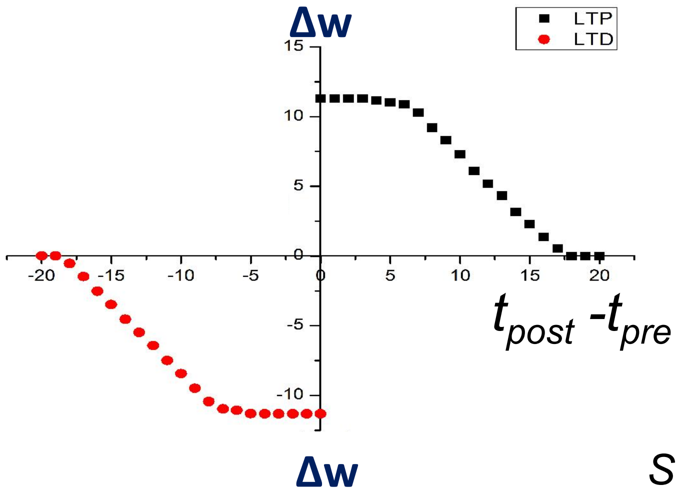
Dowrick et al, Neurocom., vol.314, pp.78-85, https://doi.org/10.1016/j.neucom.2018.06.065 (2018)

Compact decision circuits (STDP)

Weight Increase, WI, Circuit Block, Output Buffers and SIFGNVM Device



Similar circuit for weight decrease



Pass transistors gated by V_{pre}, V_{post} charge node X

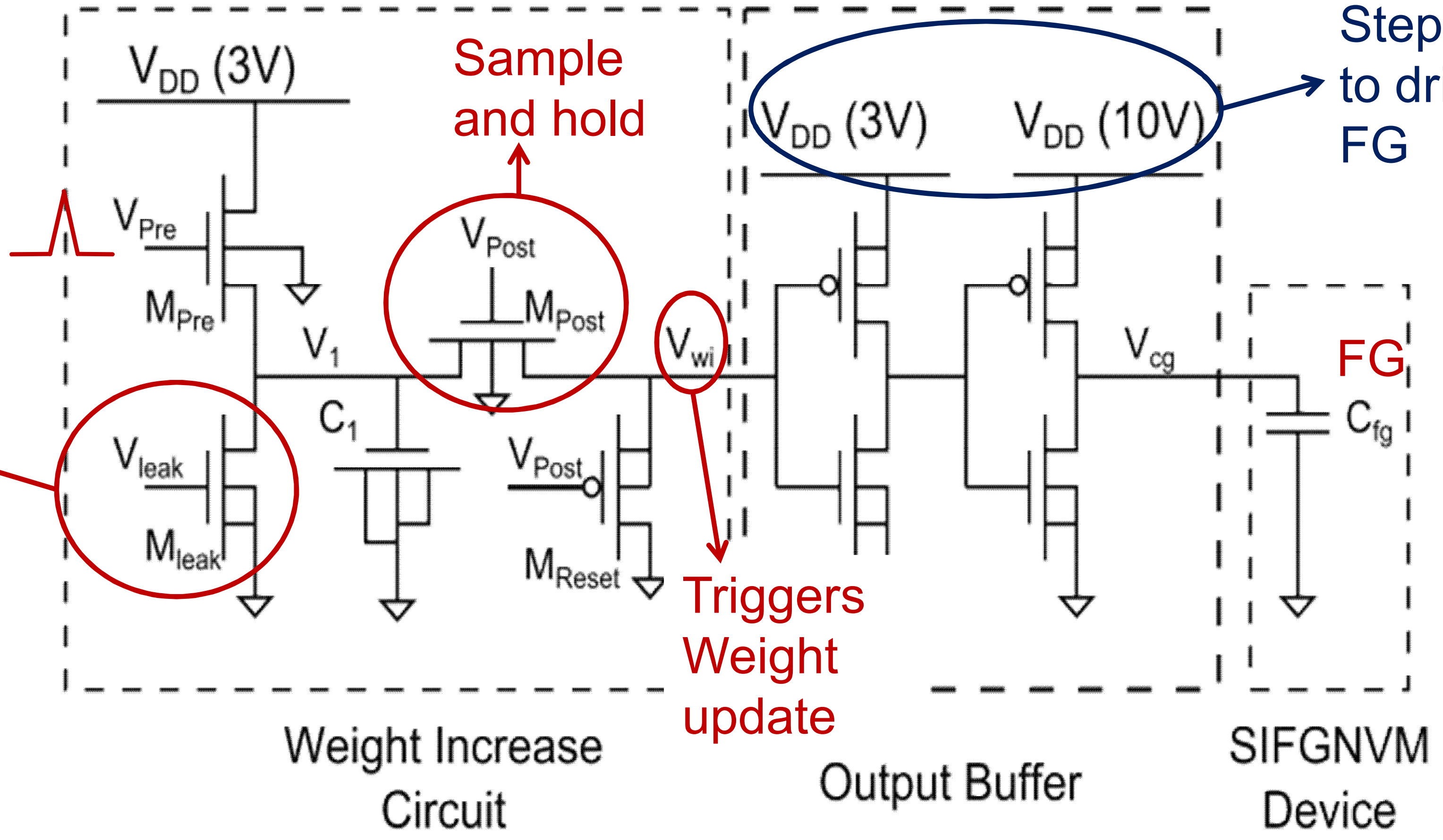
Decay via sub-Vth MOST

Sets plasticity ‘window’

Smith et al, Neurocomputing, v124, p 210 (2014)

How it works: WI Block Operation Pre-Post Spiking Event

- When a presynaptic spike occurs (V_{Pre})
  - V_1 is pulled up to 3V- $V_{TMpre}(V_1)$, C_1 charges via M_{pre}
  - C_1 Slowly discharges via sub-threshold M_{leak}
  - V_{post} triggers the sample/hold as some time, t after V_{pre}



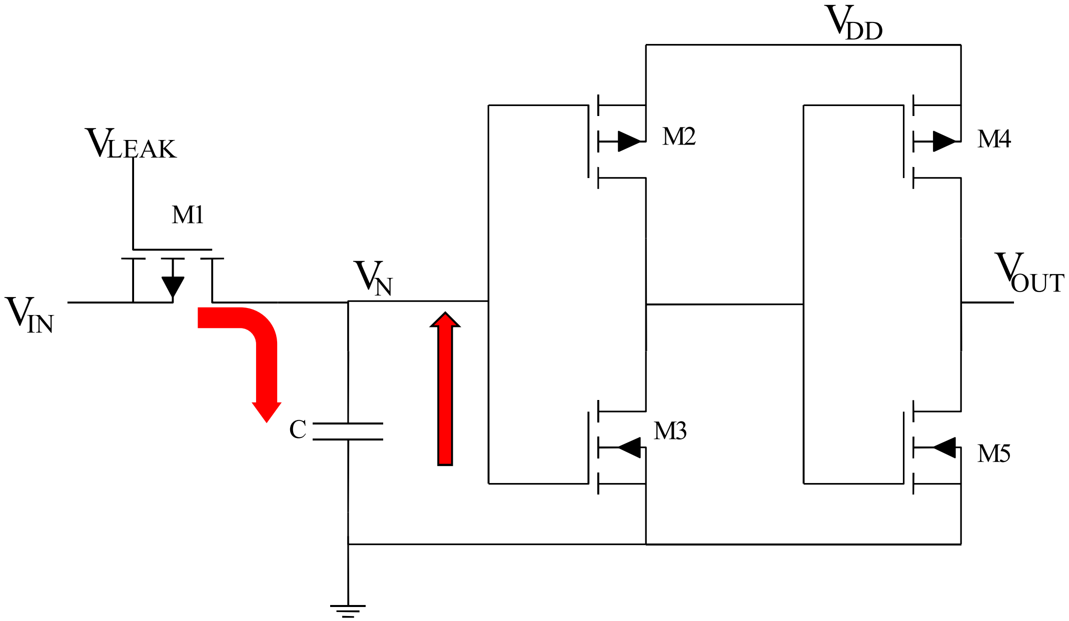
Sets timing window

Sample and hold

Triggers Weight update

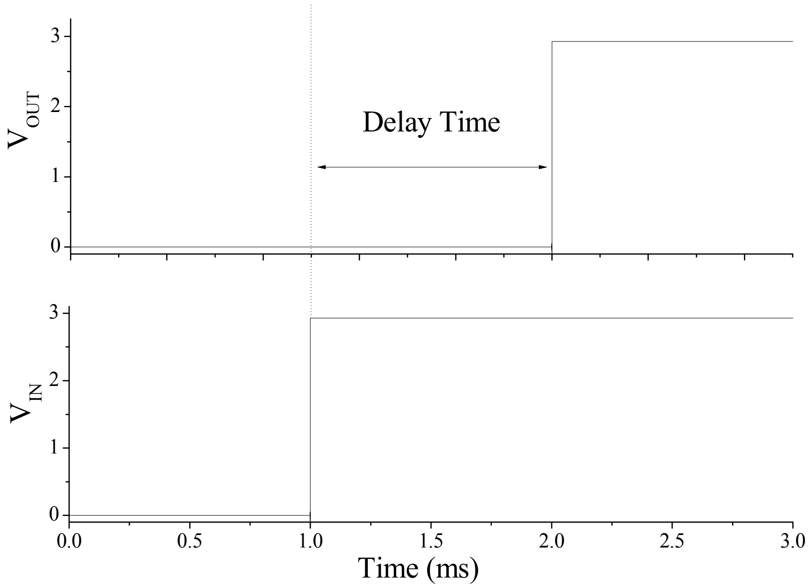
Step up V to drive FG

Axonal delay



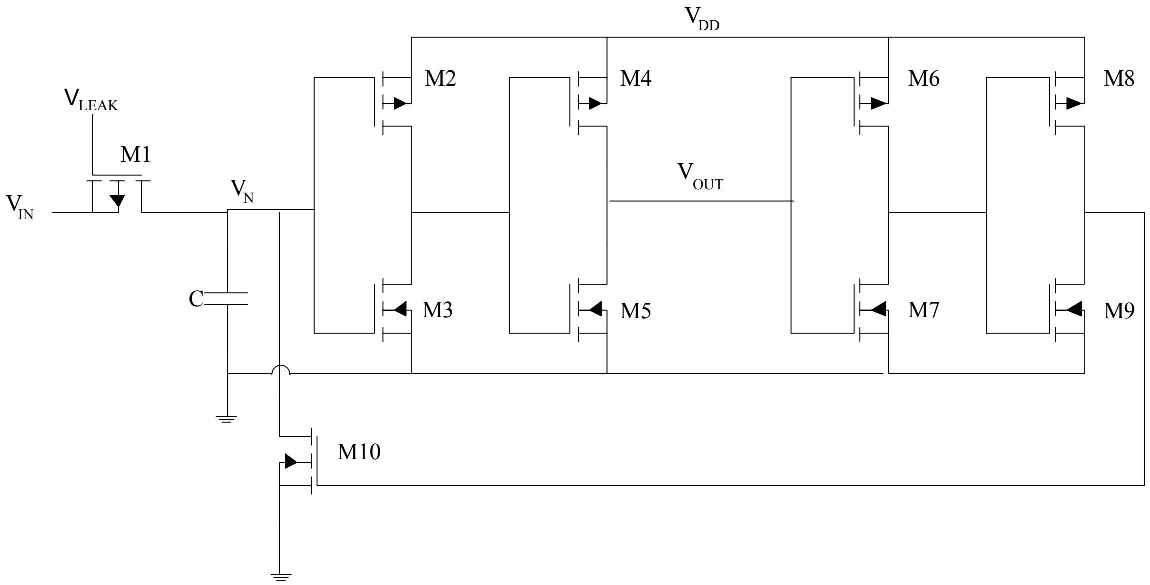
- M_1 operates subthreshold
- Slow charging of C
- V_N rises and inverters turn on
- Tune delay with V_{LEAK}

$$t_d = \frac{C}{I_0 \exp\left(\frac{V_{GSM1}}{mV_{th}}\right)}(V_{TRIG} - V_{N0})$$



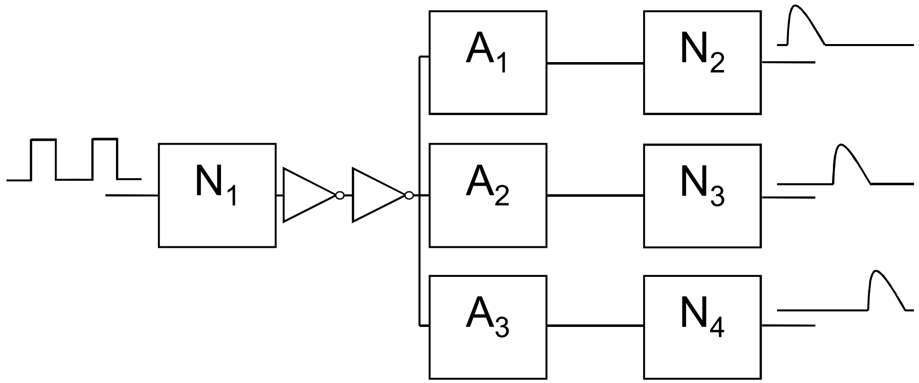
Dowrick et al, Neurocomputing, 2012
http://dx.doi.org/10.1016/j.neucom.2012.12.004,

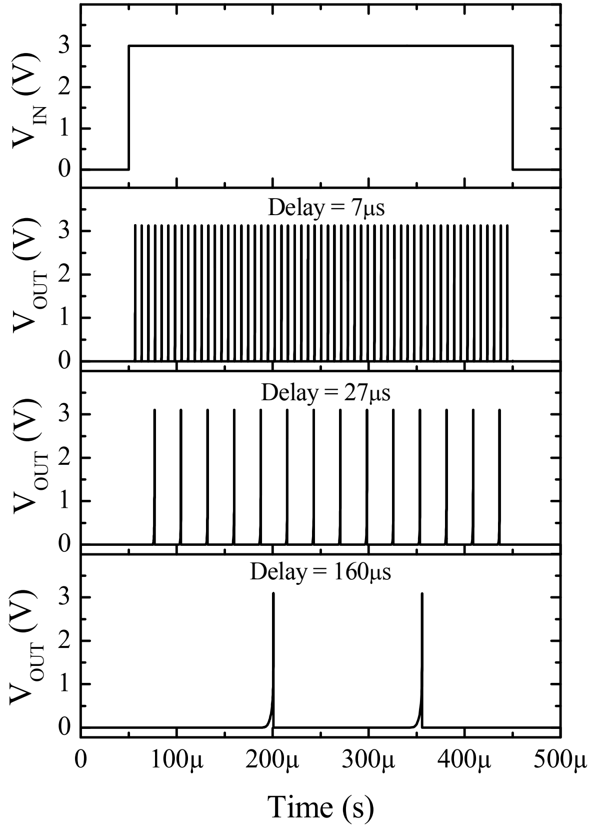
Pulse burst creation



Add feedback (M_{10})
Define pulse trains

Integrate axon delays (A) into paths

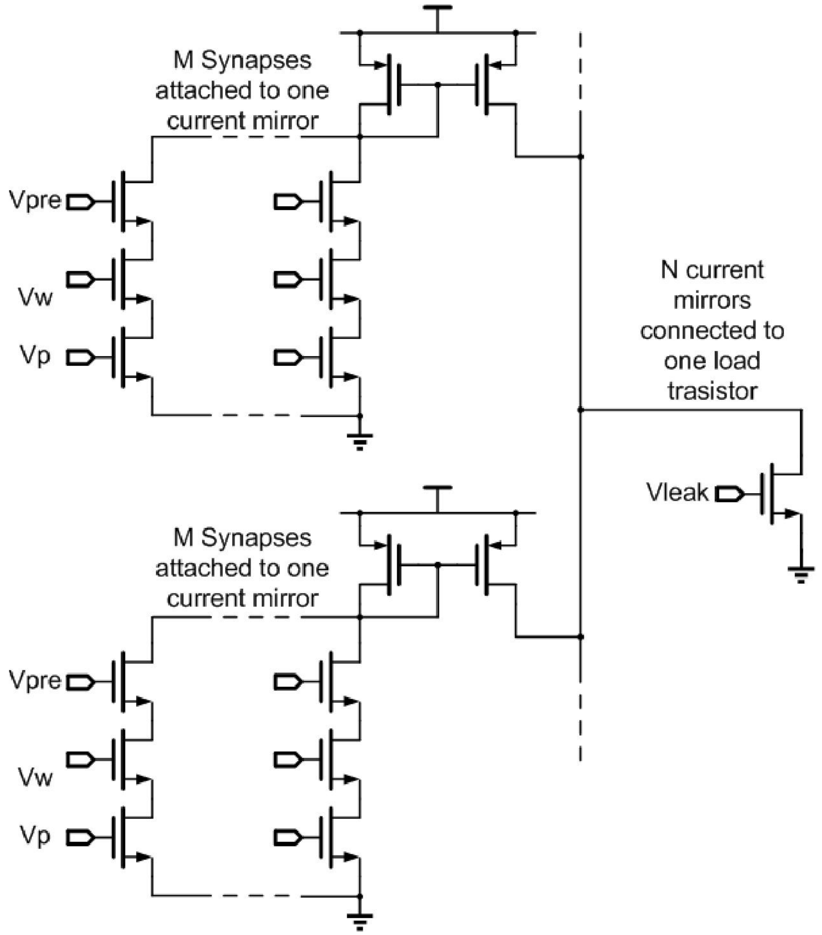


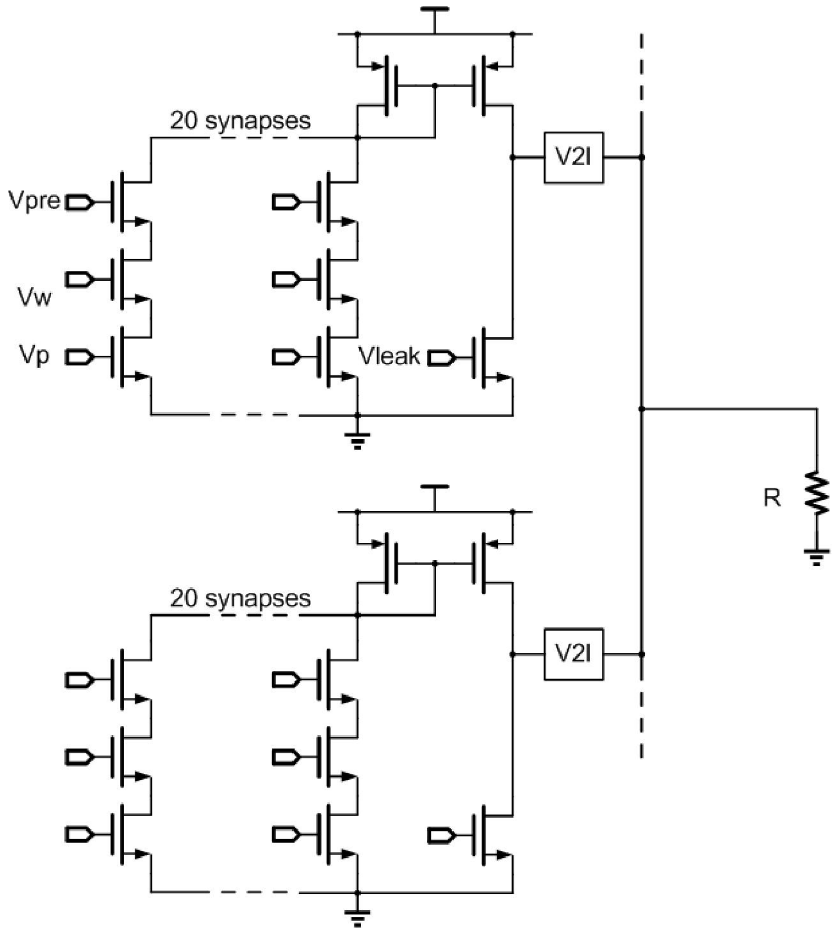


Dowrick et al, Neurocomputing, 2012
http://dx.doi.org/10.1016/j.neucom.2012.12.004,

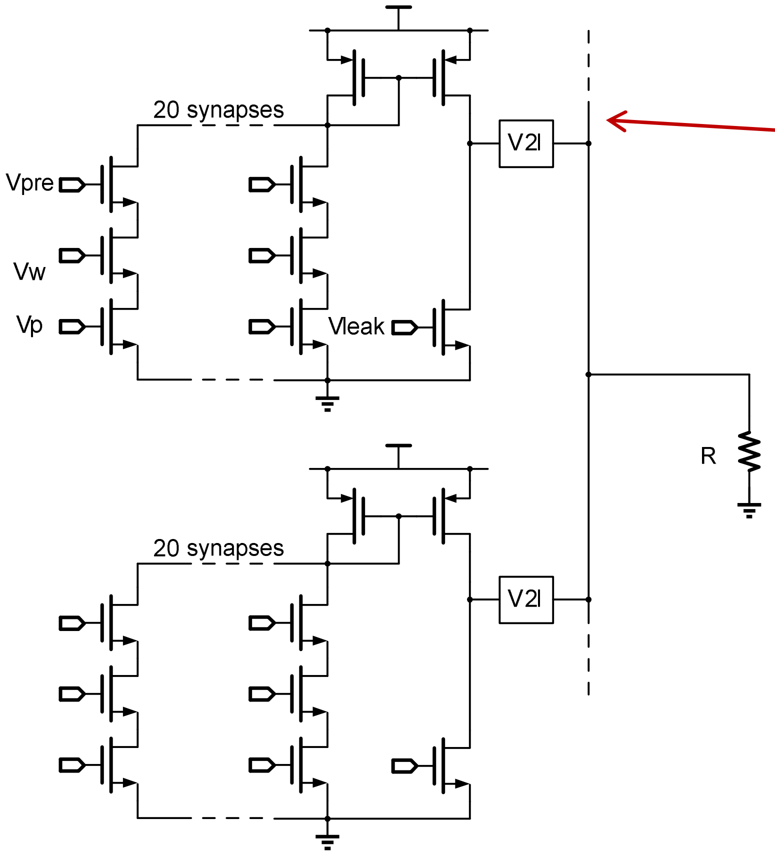
Scaling

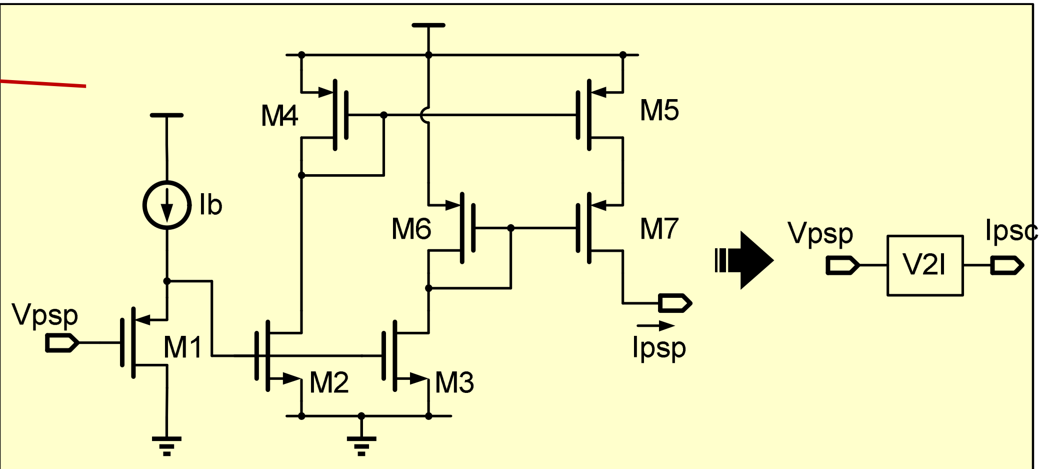
- Two solutions: sum voltages or sum currents


M Synapses attached to one current mirror
Vpre
Vw
Vp
M Synapses attached to one current mirror
Vpre
Vw
Vp
N current mirrors connected to one load trasistor
Vleak



Scaleability: easier to sum currents





linear voltage-to-current converter (V2I)

Transmit voltage steps and re-create spikes for long interconnect

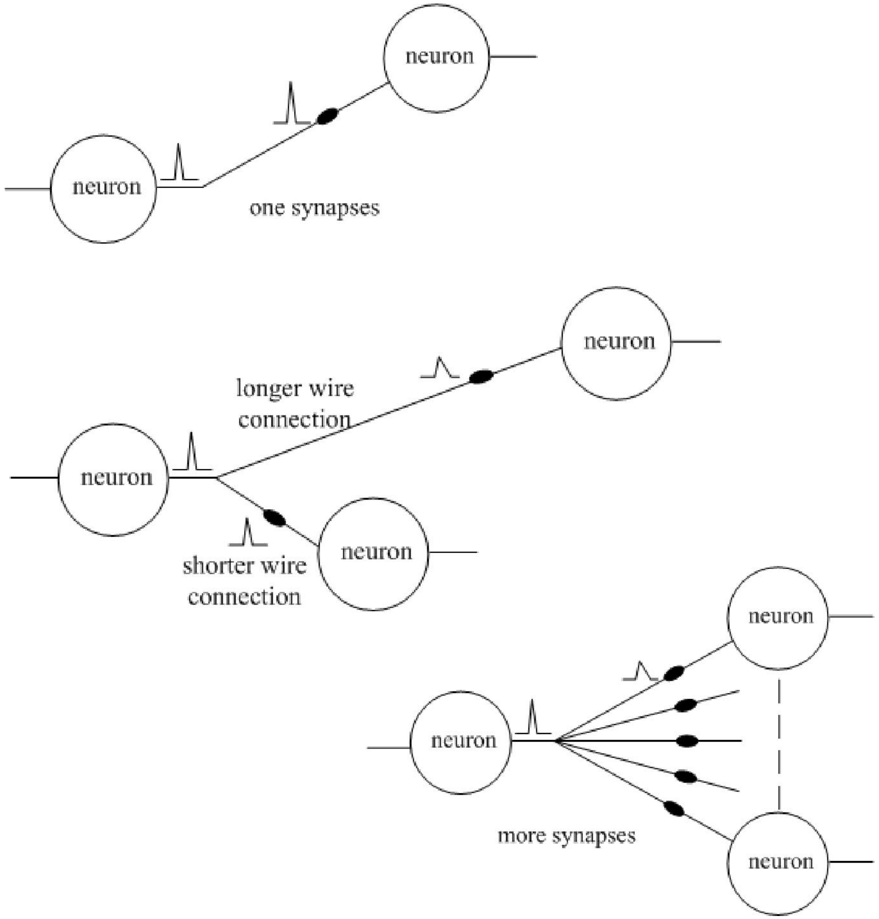
But added complexity!

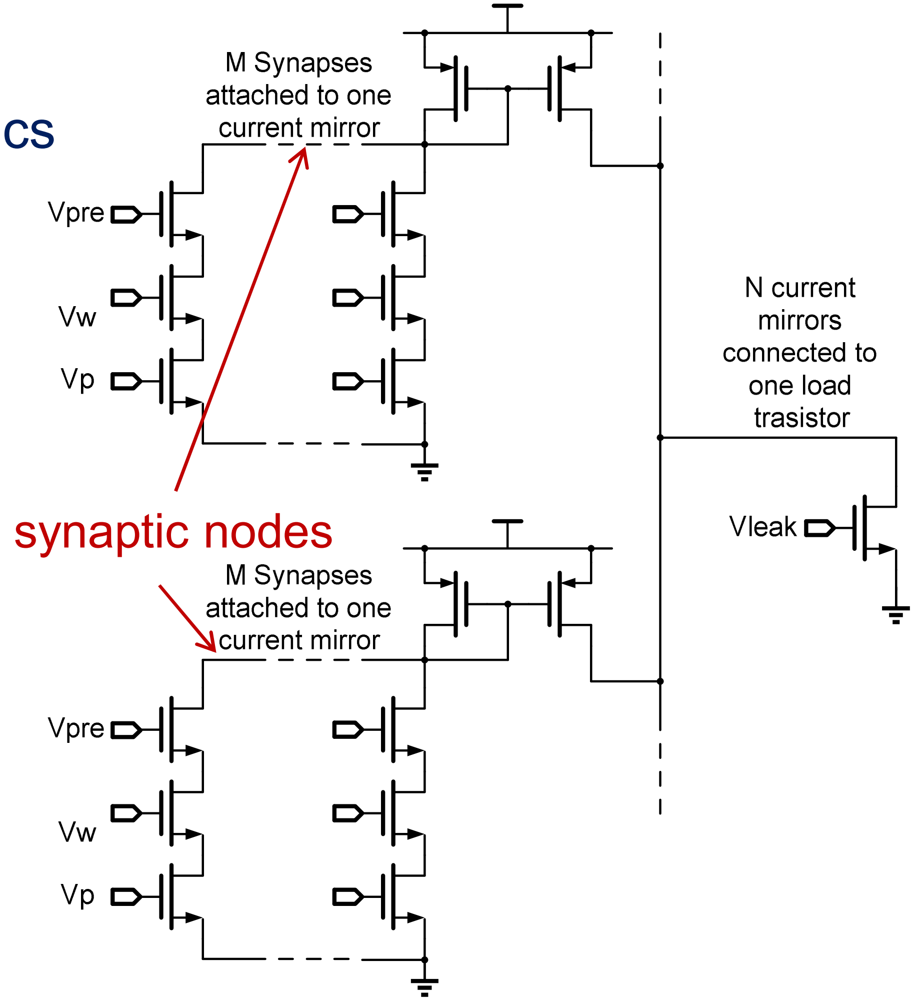
Scaling: circuit issues

Large synapse fan-out problem:

non uniform spike inputs due to parasitics

non-linearities occur in currents

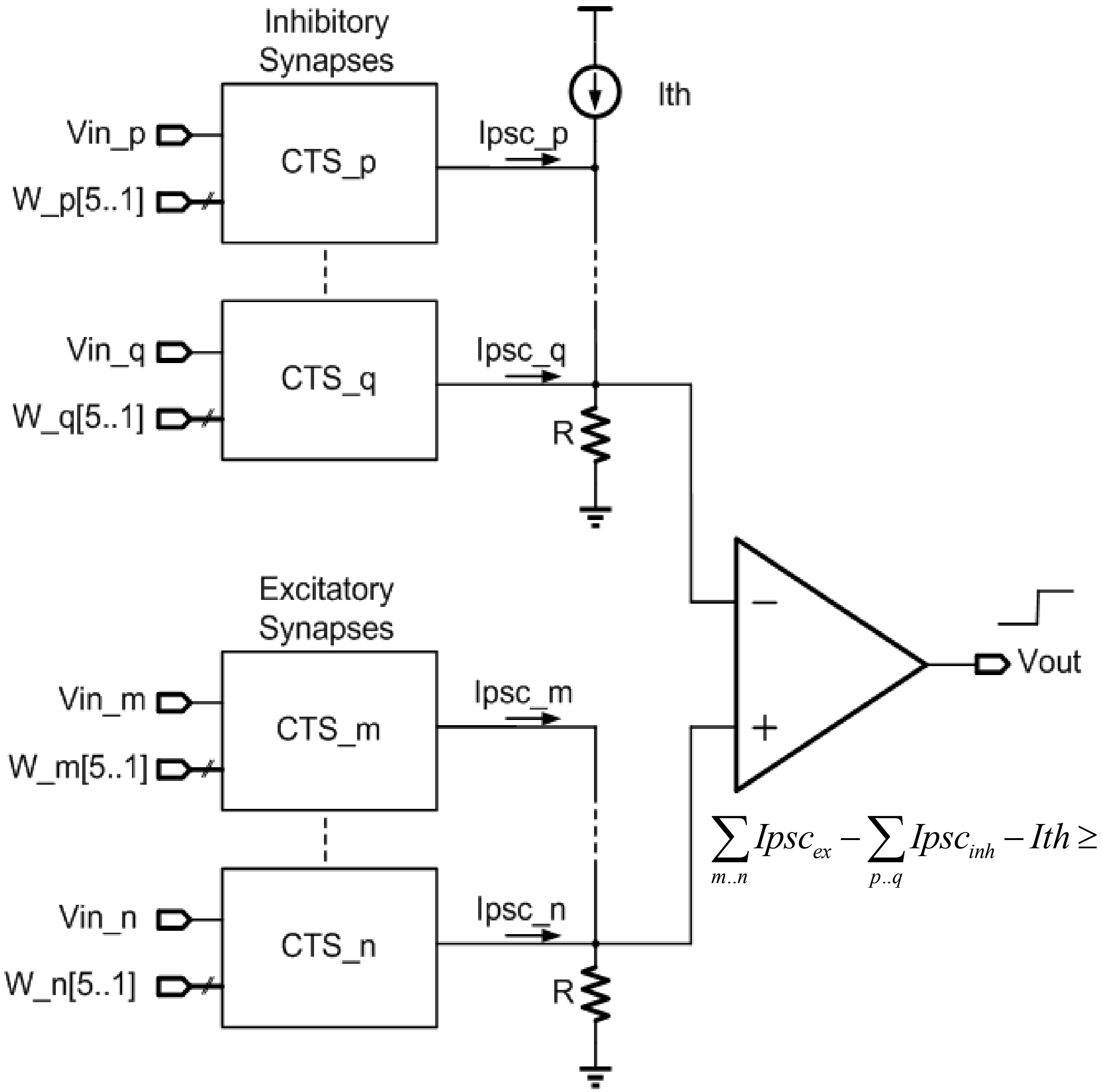



M Synapses attached to one current mirror
Vpre
Vw
Vp
N current mirrors connected to one load trasistor
Vleak
M Synapses attached to one current mirror
Vpre
Vw
Vp

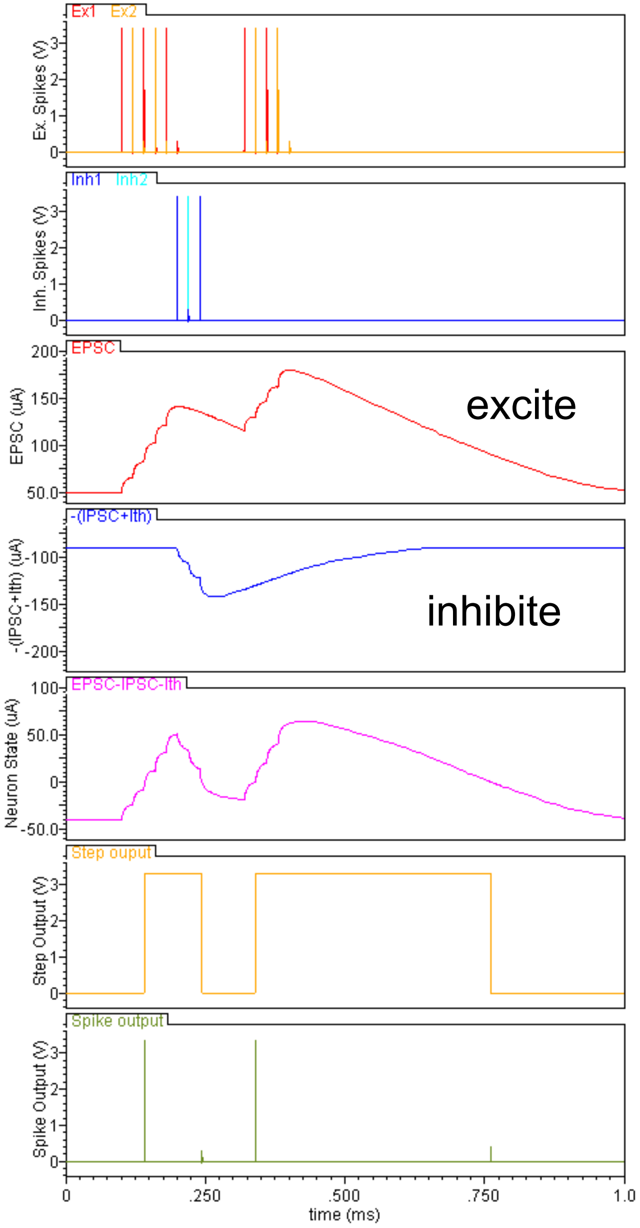
synaptic nodes

Hope it all comes out in the wash!
Nature is messy as well

Neurons with excitatory and inhibitory synapses



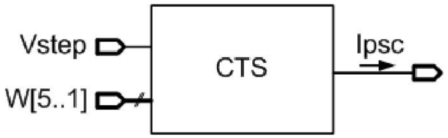
$$\sum_{m..n} Ipsc_{ex} - \sum_{p..q} Ipsc_{inh} - Ith \geq 0$$

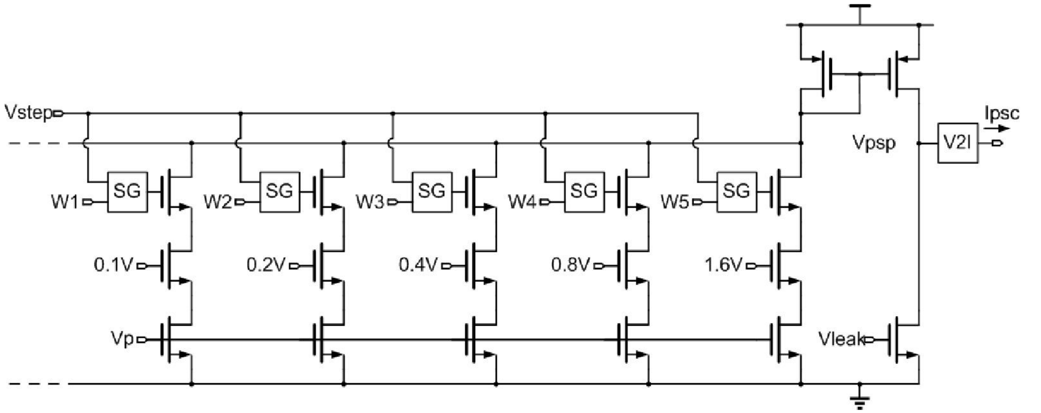


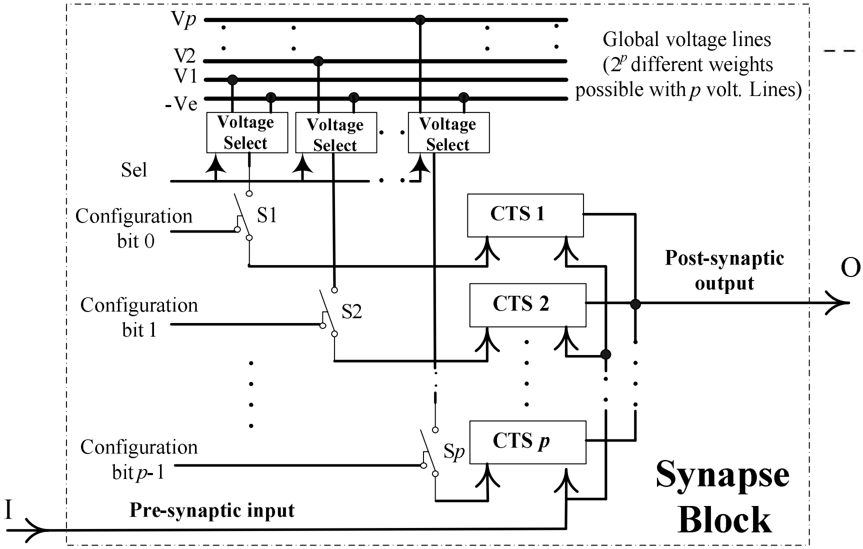
Programmable weights

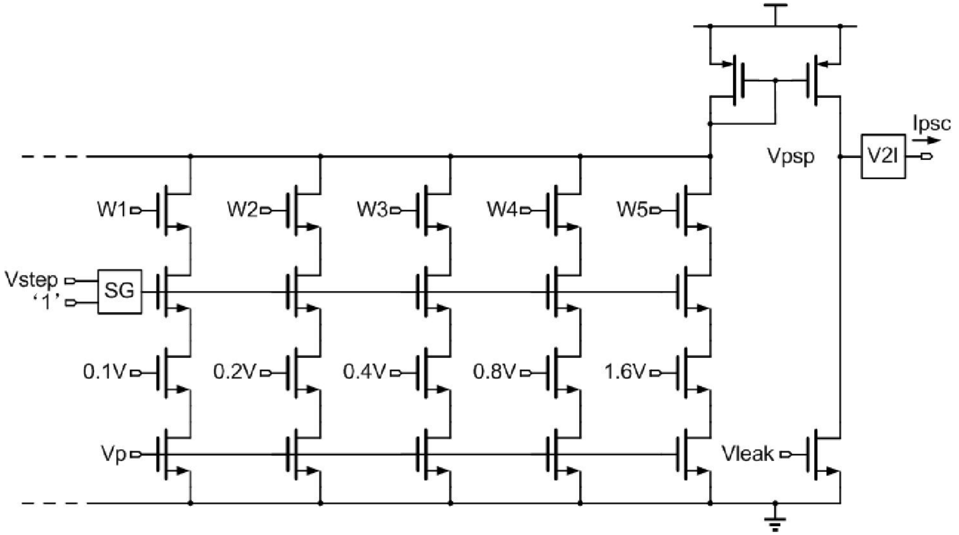
- Analog weight
  - Good: Continuous weight value, compact analog storage circuit
  - Bad: Inaccurate, require bias reference circuit and complex control circuit for high resolution, also require high voltage rail and undocumented
  - technology feature
- Digital weight
  - Good: accurate, mature digital memory technology, easy to program
  - Bad: discrete quantitative weight, require more space

Programmable weight



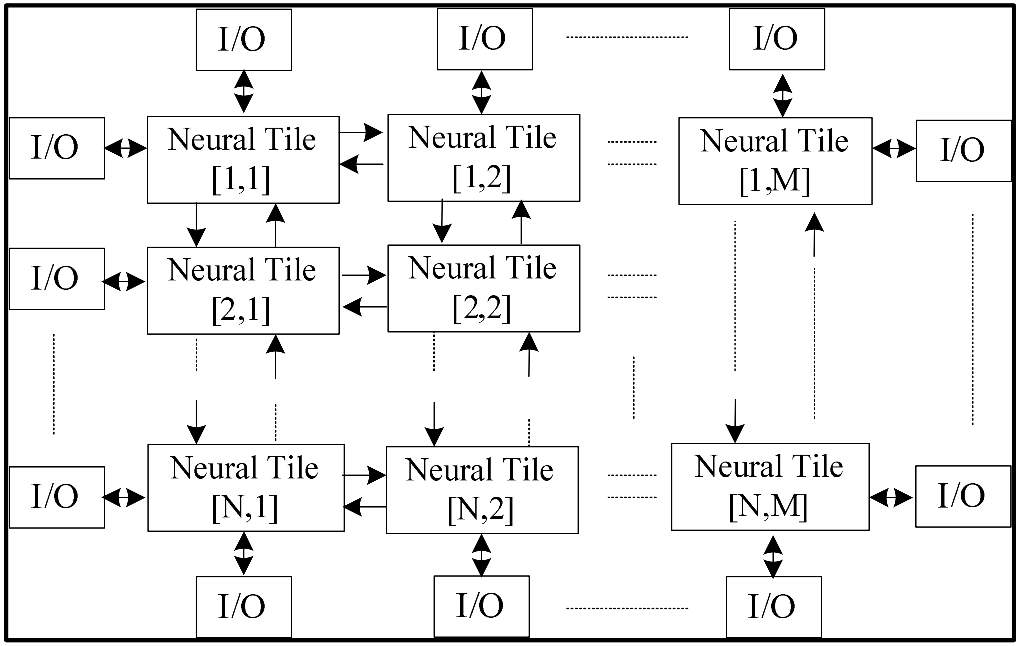


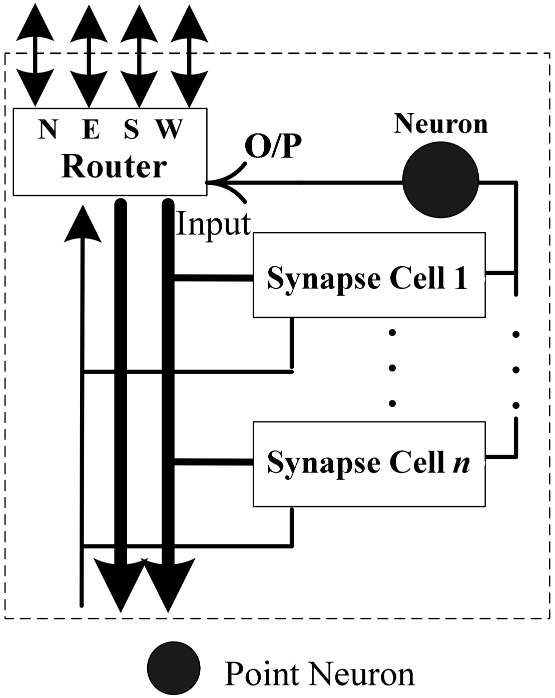




Embrace: an alternative approach

- Network-on-chip address the issues of scalability and connectivity between components.
- Low-area/power spiking neuron cells with associated training provides neural computing capability.





- 2-dimensional array of interconnected neural tiles + I/O blocks.
- Neural tiles connected in North, East, South and West.
- Tile can be programmed to realise neuron-level functions.

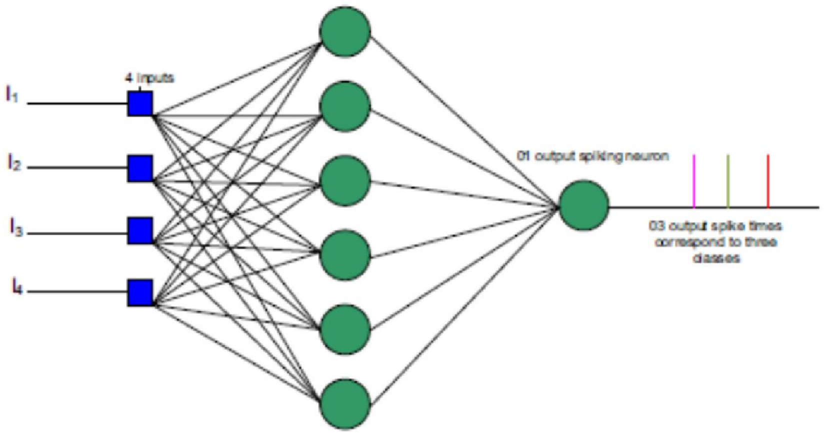
Harkin et al, Int. Jnl of Reconfigurable Computing, doi:10.1155/2009/908740 (2009)

Slide courtesy of Jim Harkin

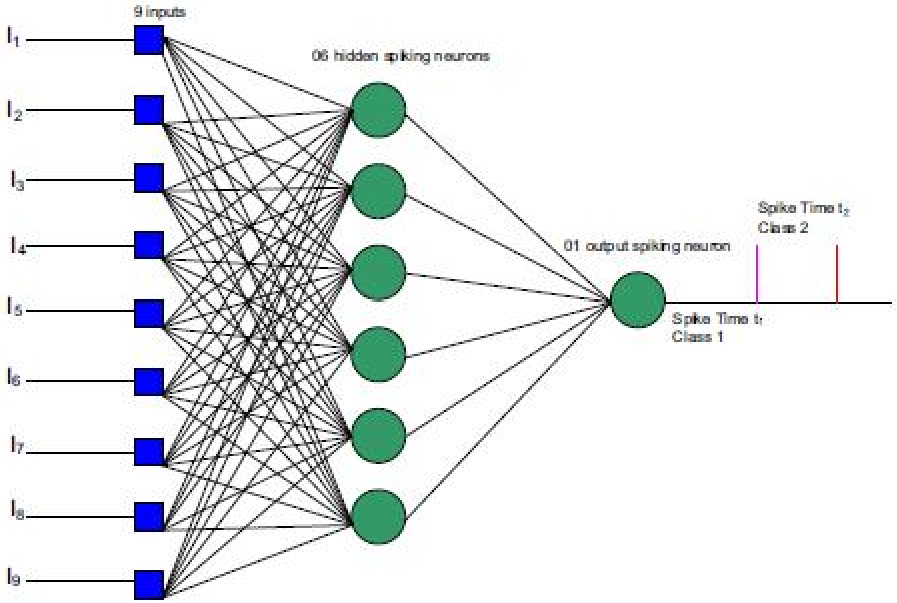
Evaluation

- Learning in software (calculate weight values)
  - Fit the experimental synapse results
- Solve benchmark problems
  - Wisconsin breast cancer (WBC) dataset
  - IRIS dataset
- Temporally encoded input values

SNN architecture: IRIS dataset

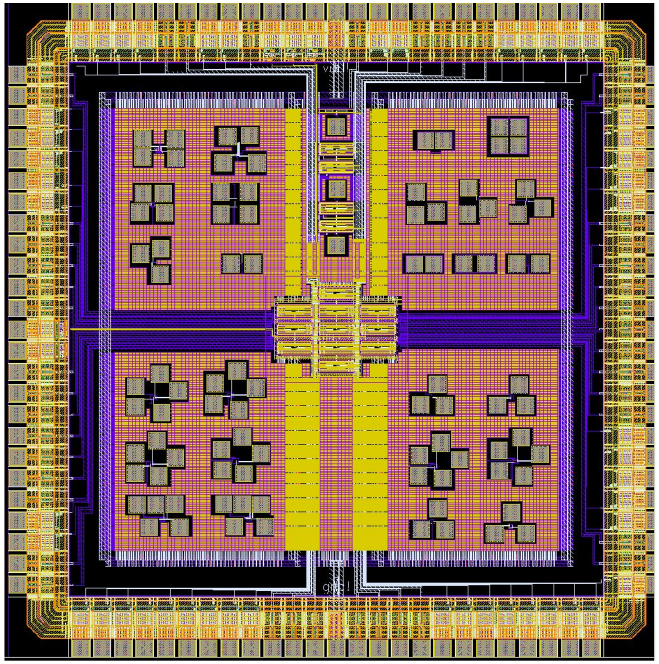


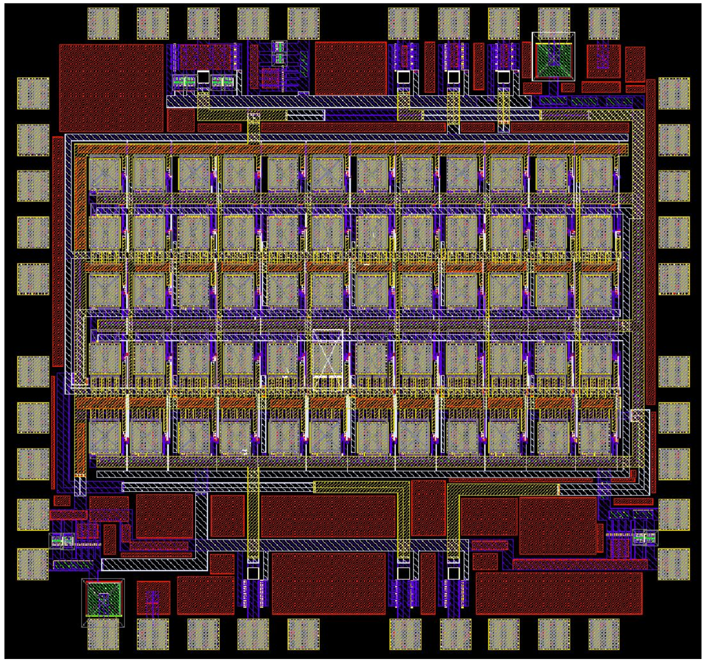
SNN architecture: WBC dataset



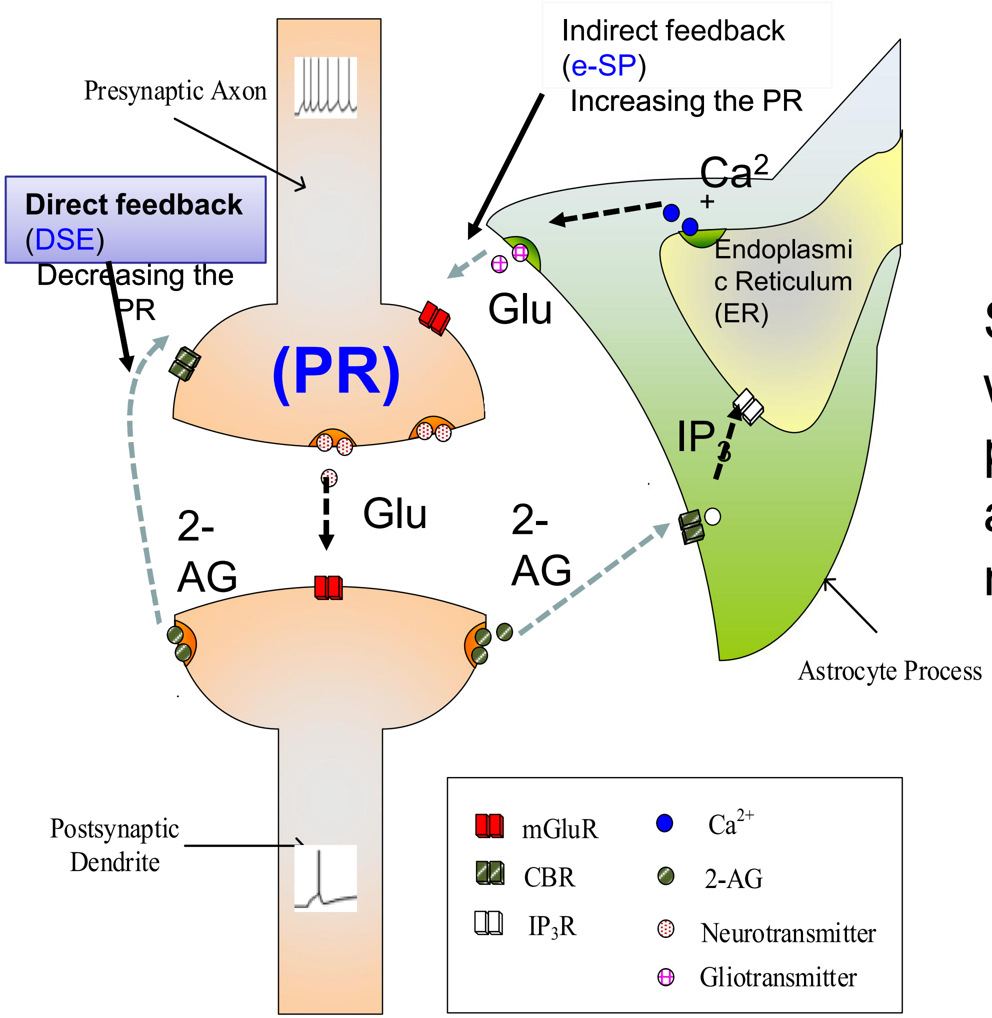
Ghani at al, Neurocomputing, 83 (2012) pp.188–197 (2011)

Circuits fabricated in AMS 0.35,mixed signal CMOS





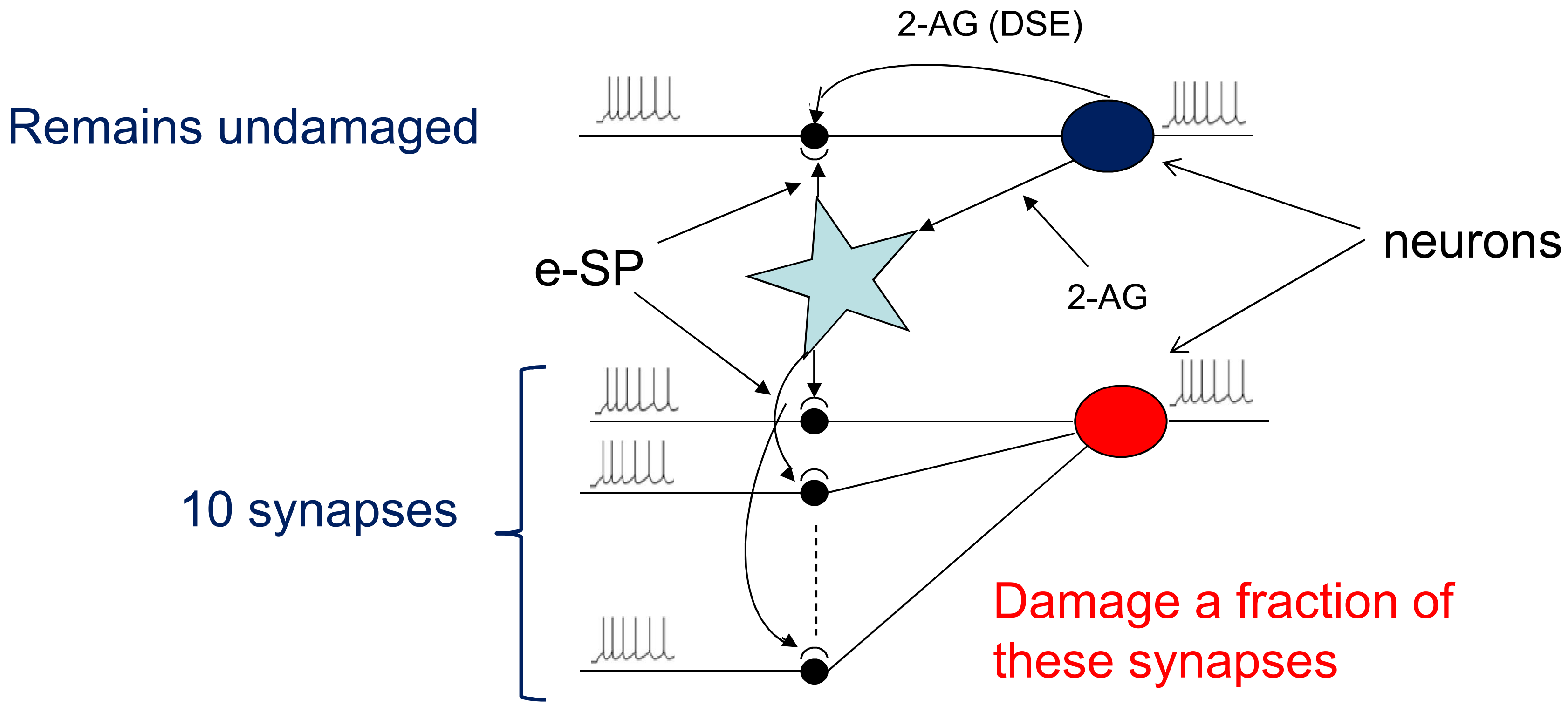
Astrocytes



Study transport within astrocyte process and between neuron/astrocyte

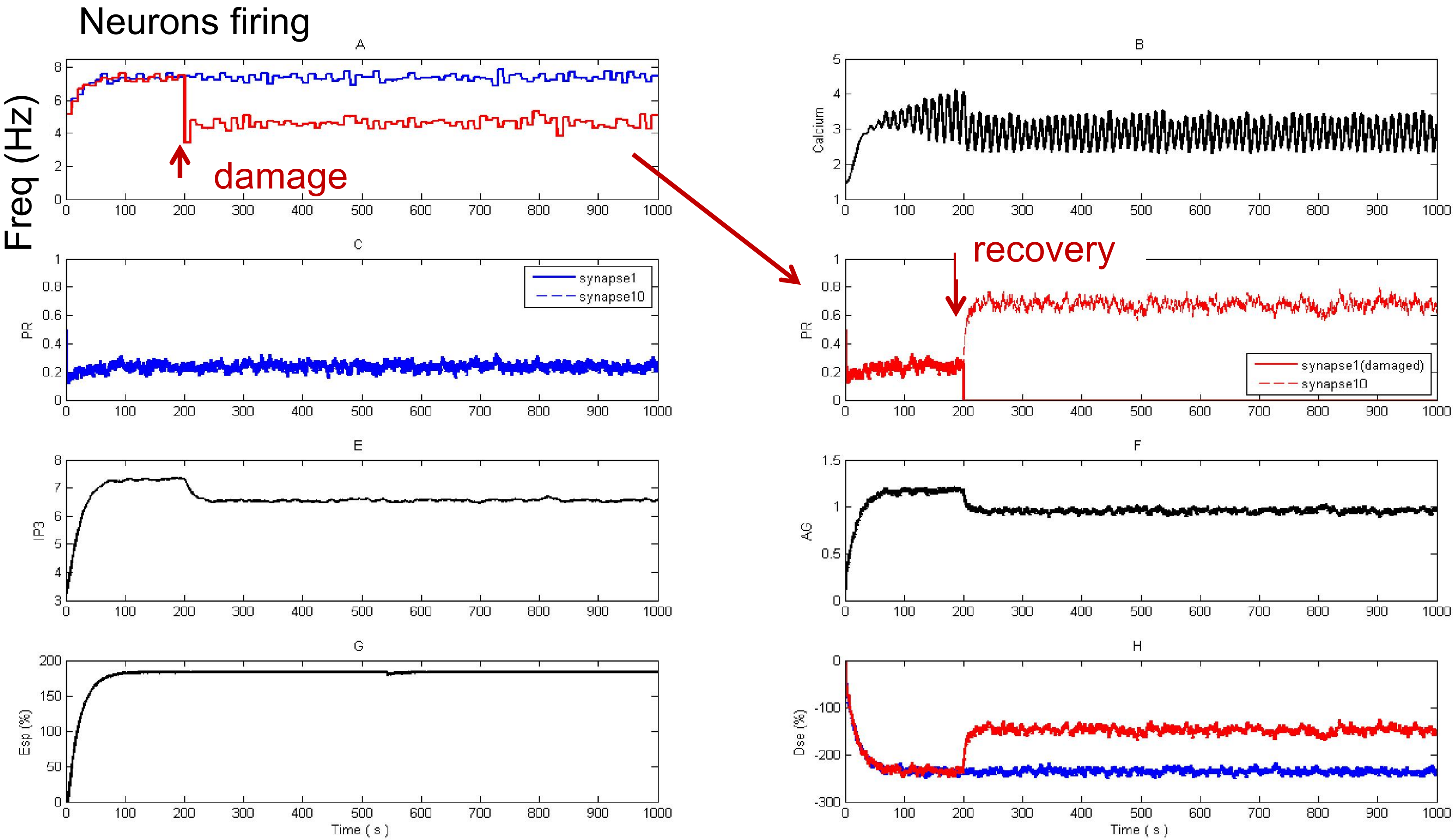
Breslin et al, PLoS Computational Biology, doi.org/10.1371/journal.pcbi.1006151, May (2018)

Endocannabinoid Mediated Self-Repair



Wade, McDaid et al, Frontiers in computational neuroscience, v6, Art 76 (2012)

Astrocytes mediate self-repair



- Astrocyte ‘forces’ synapses to ‘work harder’
- Opens up STDP window – restarts learning

Wade, McDaid et al, Frontiers in computational neuroscience, v6, Article 76 (2012)

What we learnt..

- Can build compact analogue circuits that emulate aspects of biology with a degree of success (better than in software? – potentially much faster)
- Getting them to learn is another matter..
  - Need feedback
  - Weight update
  - Starts to get very complicated…
- A lot of redundancy once the circuit has ‘learnt’
- Scaling soon results in a huge amount of interconnect

Need software/hardware combination – learning in software

Still some way to go before….

The MAN WITH THE ELECTRONIC BRAIN!
HE WAS THE MOST INTELLIGENT -- AND DANGEROUS -- MAN ON EARTH! BUT WOULD HE USE HIS POWERS TO SAVE THE WORLD -- OR DESTROY IT?
WE CAN'T LET THE MAN WITH THE ELECTRONIC BRAIN RUN WILD -- HE DESTROYS EVERY-THING HE TOUCHES!
BUT WE CAN'T DESTROY HIM, EITHER! IF WE DO THAT -- WE'LL BE DESTROYING OUR ONLY CHANCE TO SAVE THE EARTH!
MU HY

DC Comics
May 1961

Thanks to

J Harkin *(jg.harkin@ulster.ac.uk)*

T Dowrick (PhD)
A Smith (PhD)
S Chen (PhD)
S Zhang (post-doc)
A Ghani (post-doc)

Funding: EPSRC, NAP, EPSRC-DTA awards, Dorothy Hodgkin scholarship

https://www.riverpublishers.com/book_details.php?book_id=693