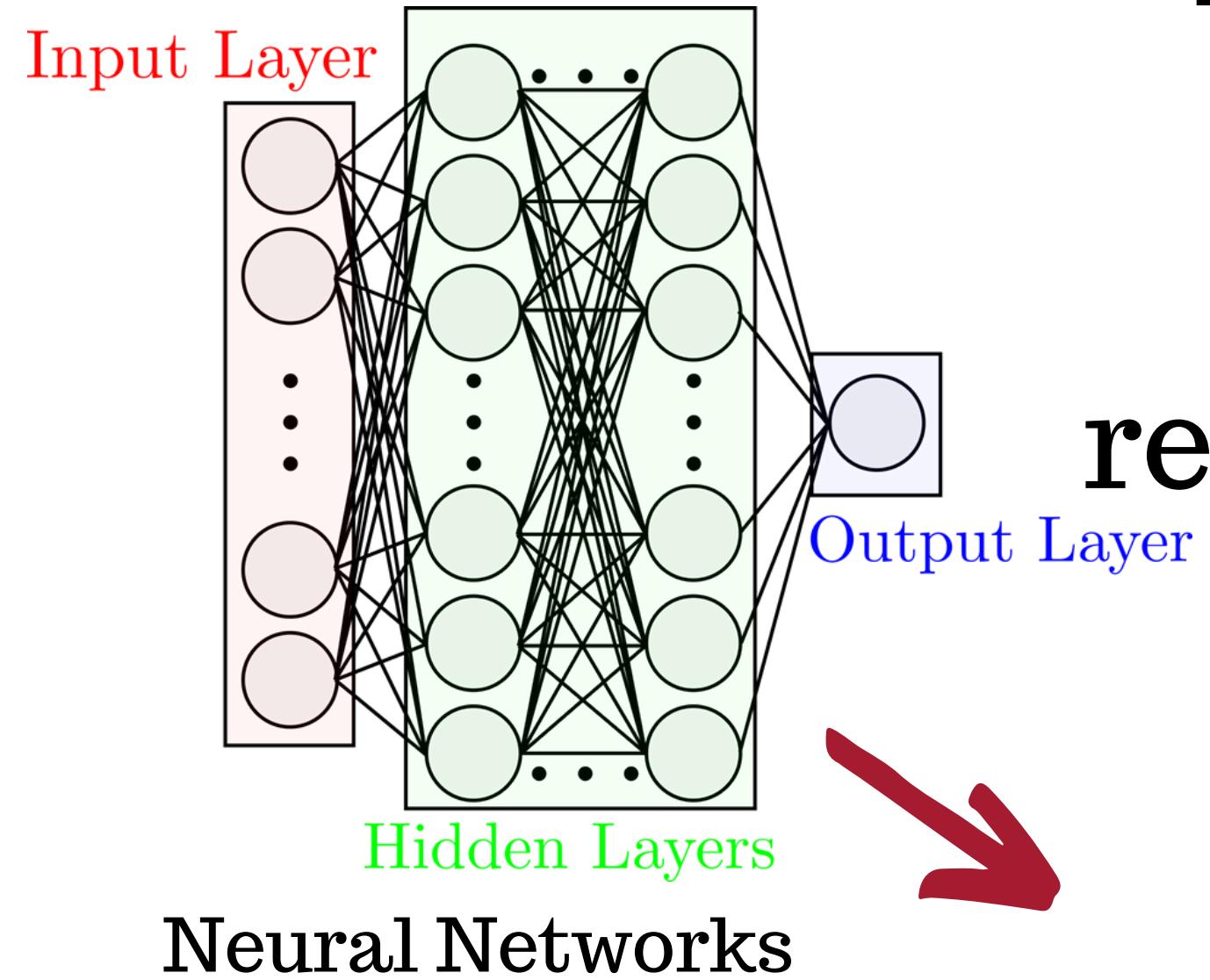


# Matrix-to-Vector Multiplication Leveraging RRAM Crossbar

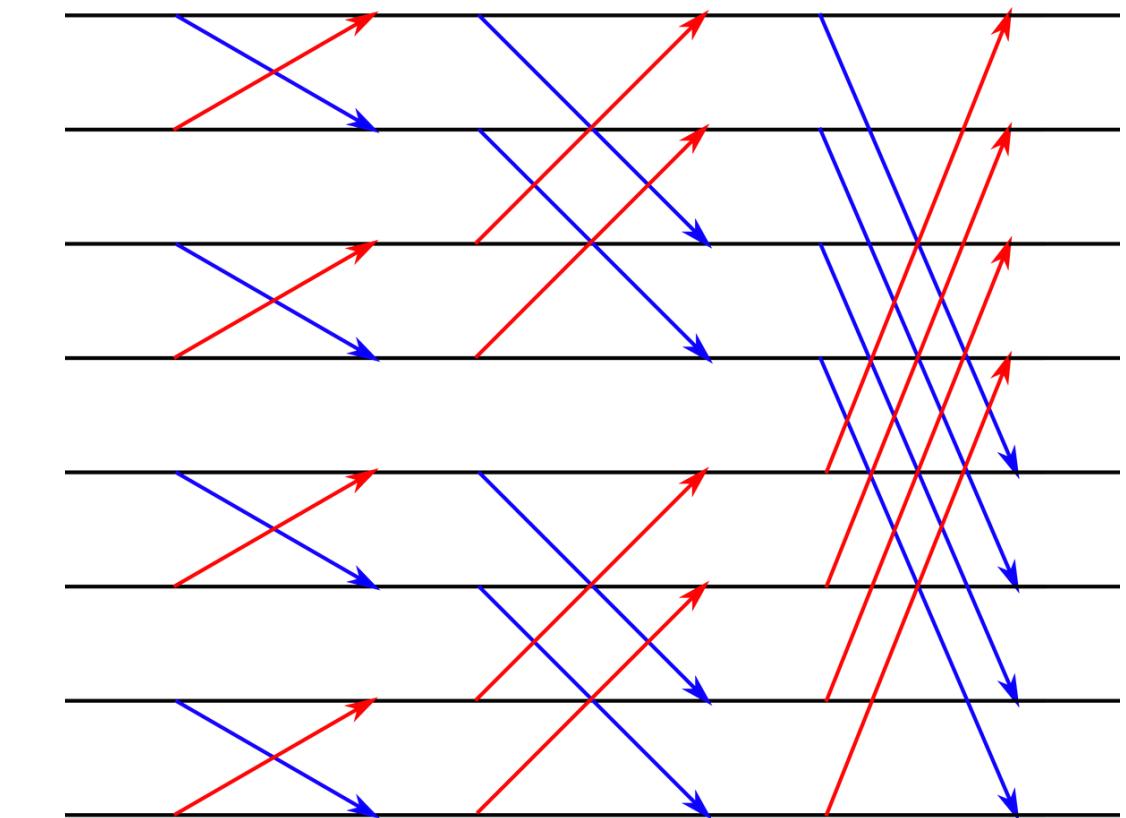
05/11/2022

# Motivation



Everything  
resumes to MMM!

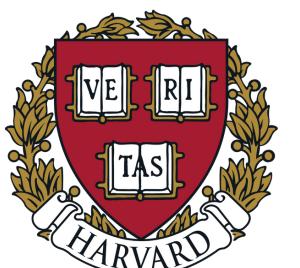
Neural Networks



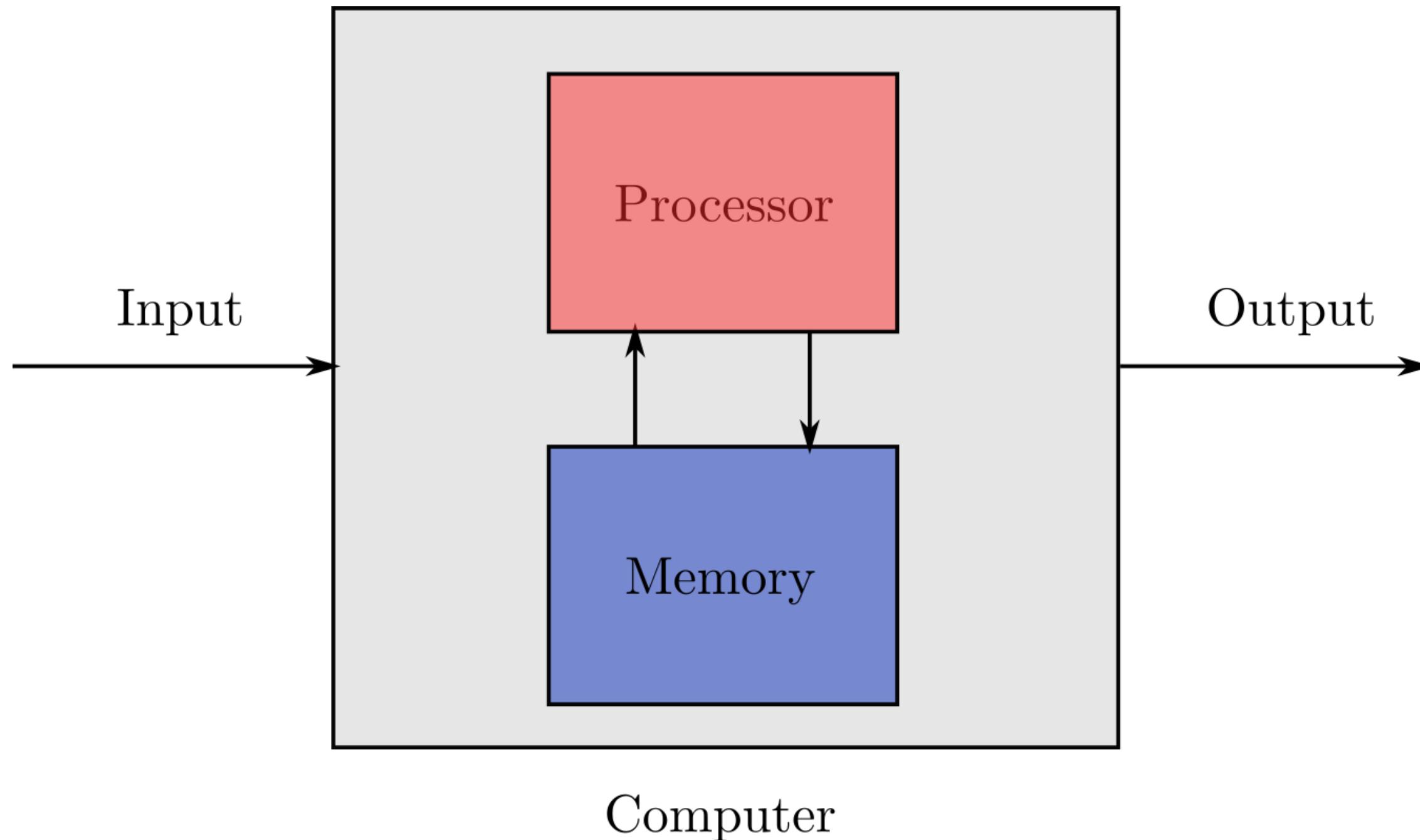
Fast Fourier Transform

$$A \times B = C$$

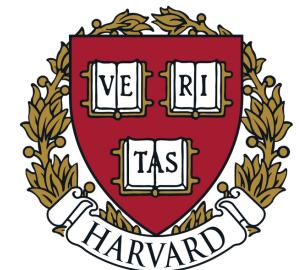
Matrix-to-Matrix Multiplication



# Motivation

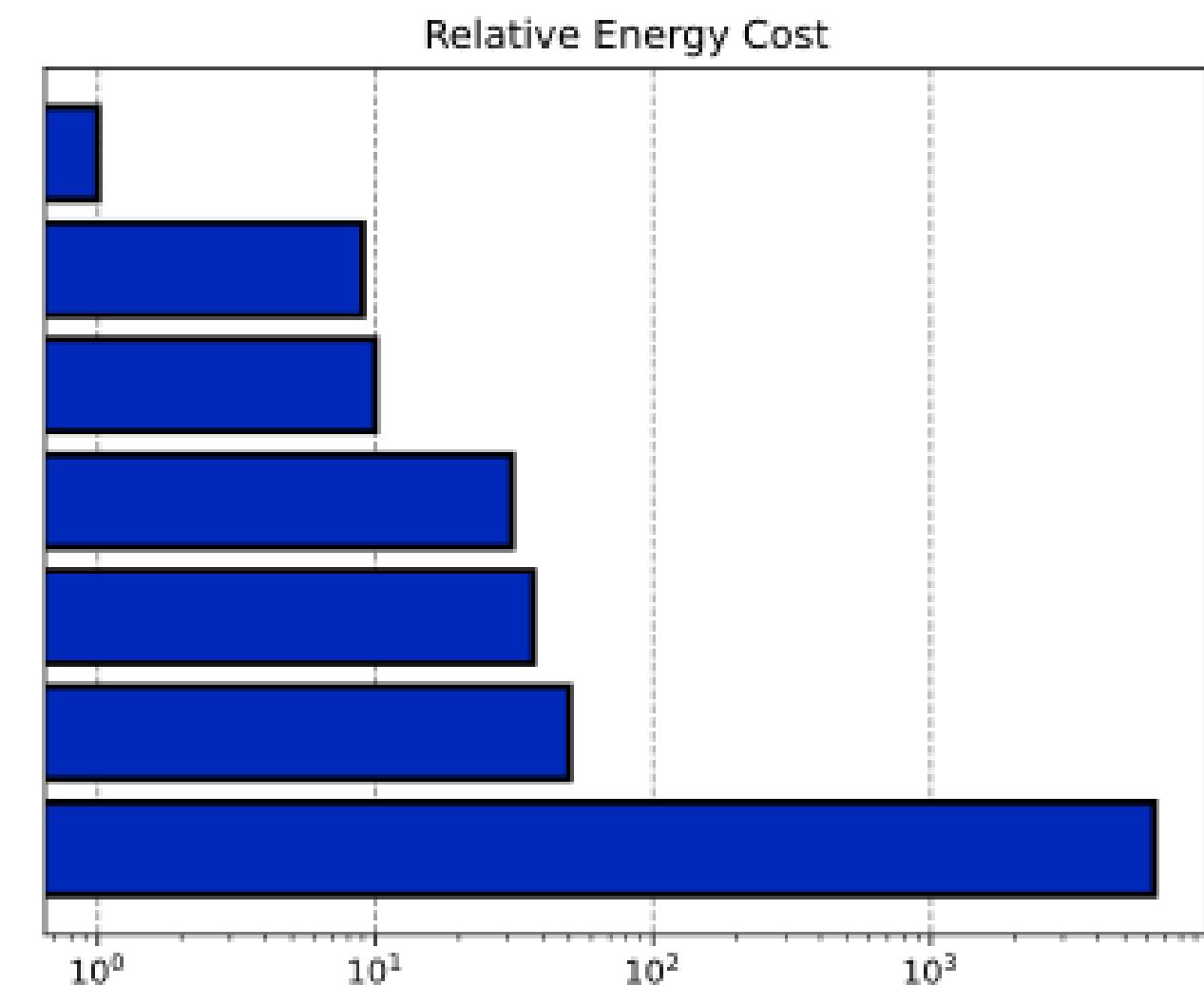


Von Neumann's architecture

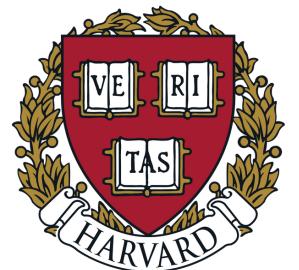


# Motivation

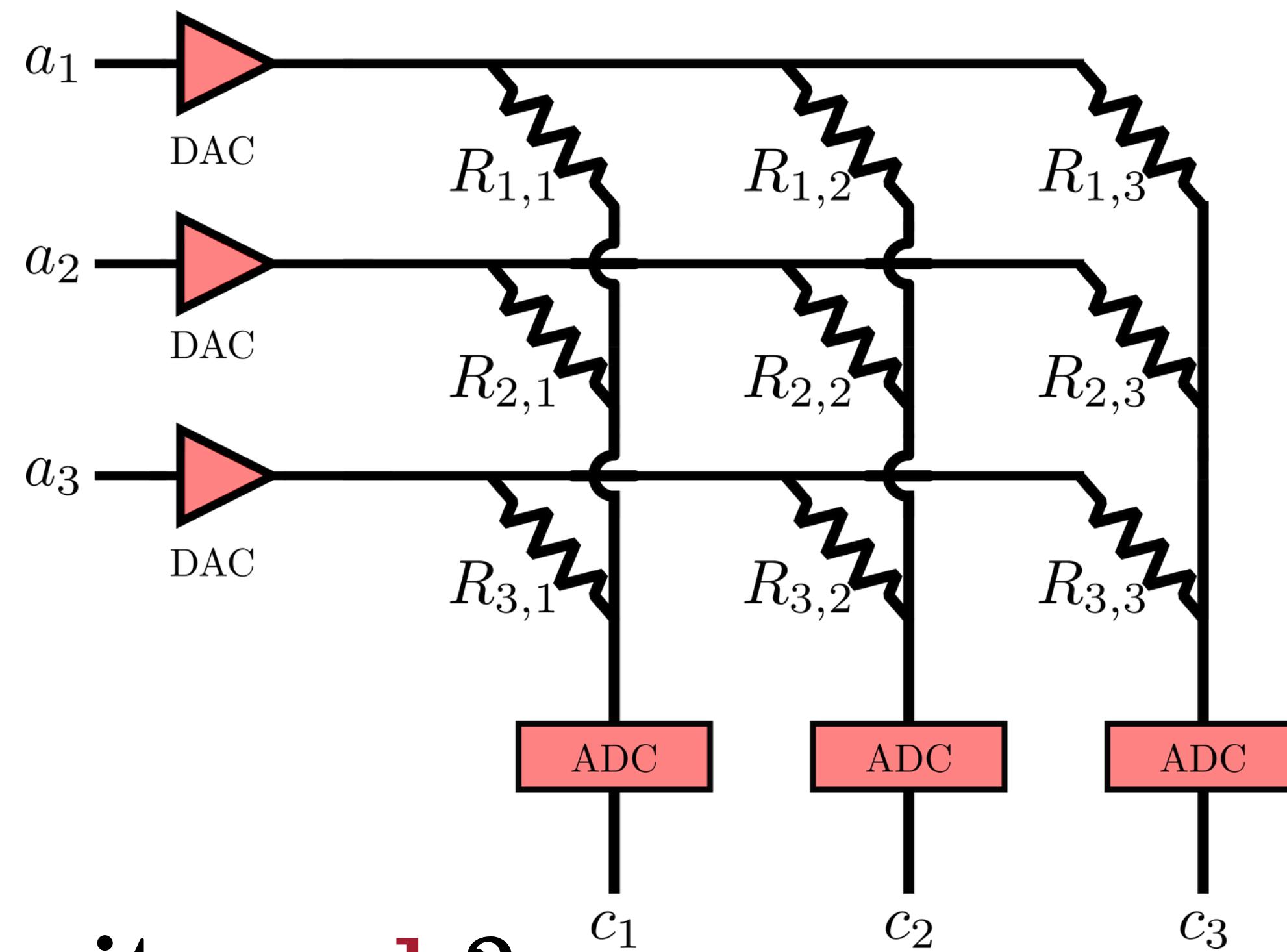
Operation (32 bits)	Energy [pJ]	Relative Cost
int ADD	0.1	1
float ADD	0.9	9
Register File	1	10
int MULT	3.1	31
float MULT	3.7	37
SRAM Cache Access	5	50
<b>DRAM Memory Access</b>	<b>640</b>	<b>6400</b>



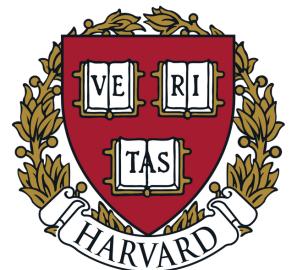
Memory accesses are a **huge** bottleneck!



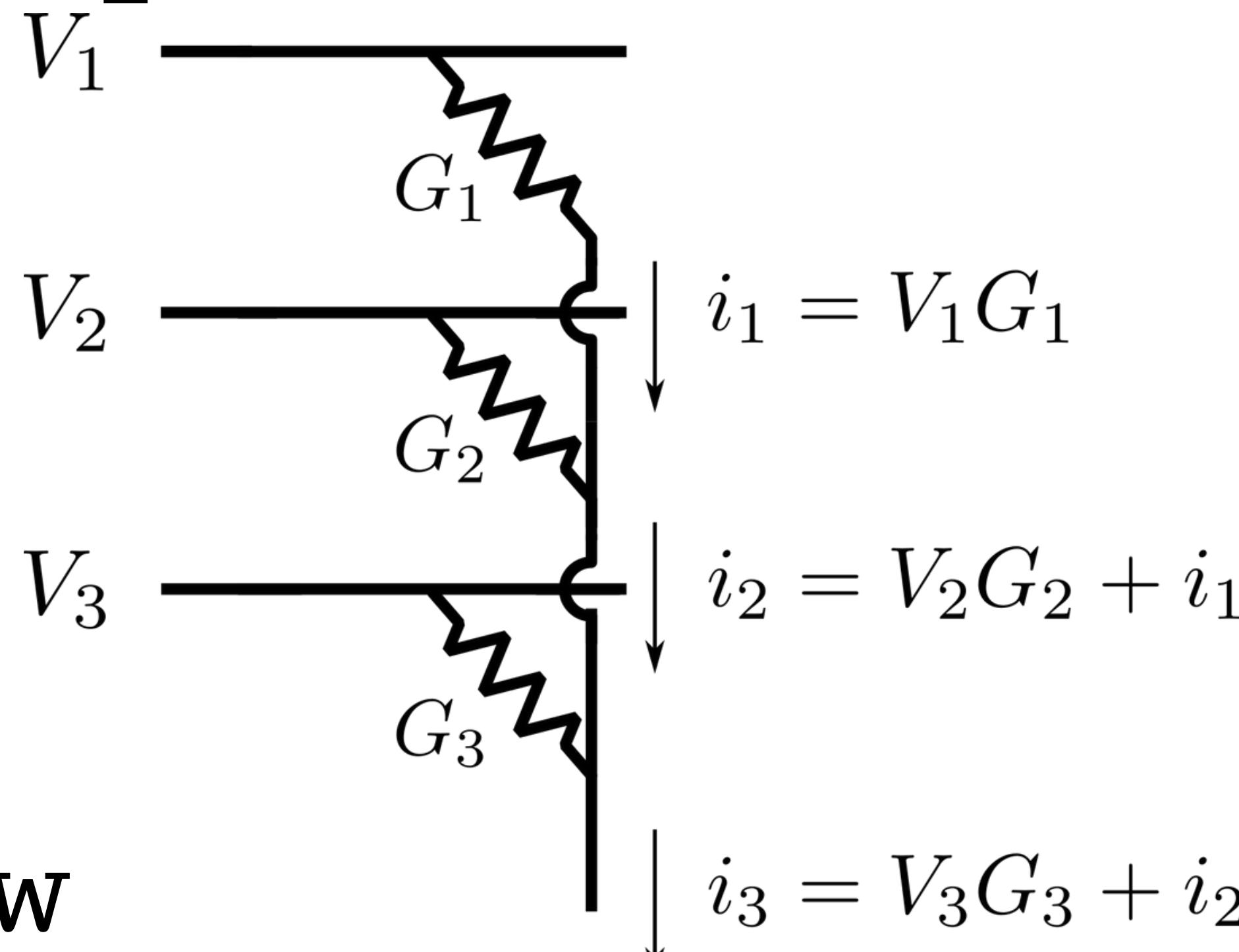
# Proposed Architecture



But...  
How does it work?

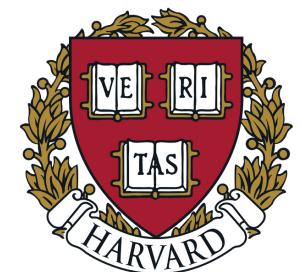


# Proposed Architecture



Ohm's law

Kirchhoff's law



# What can be **inside** each cell?

1R



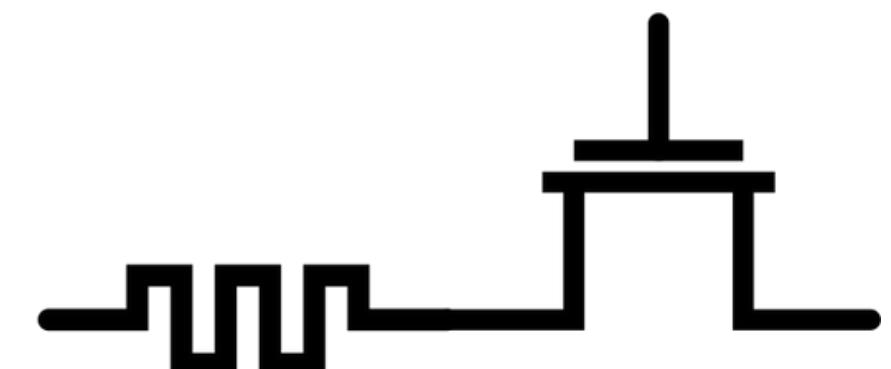
Fixed matrix (**boring**)

1T1R

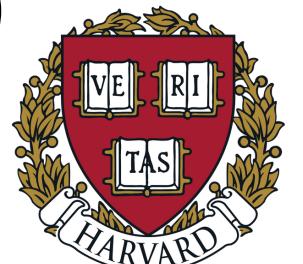


We can define **digital resistance** (at least we change the matrix)

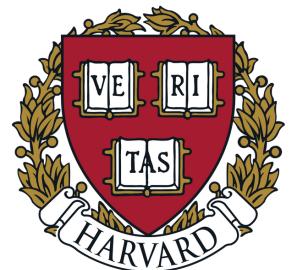
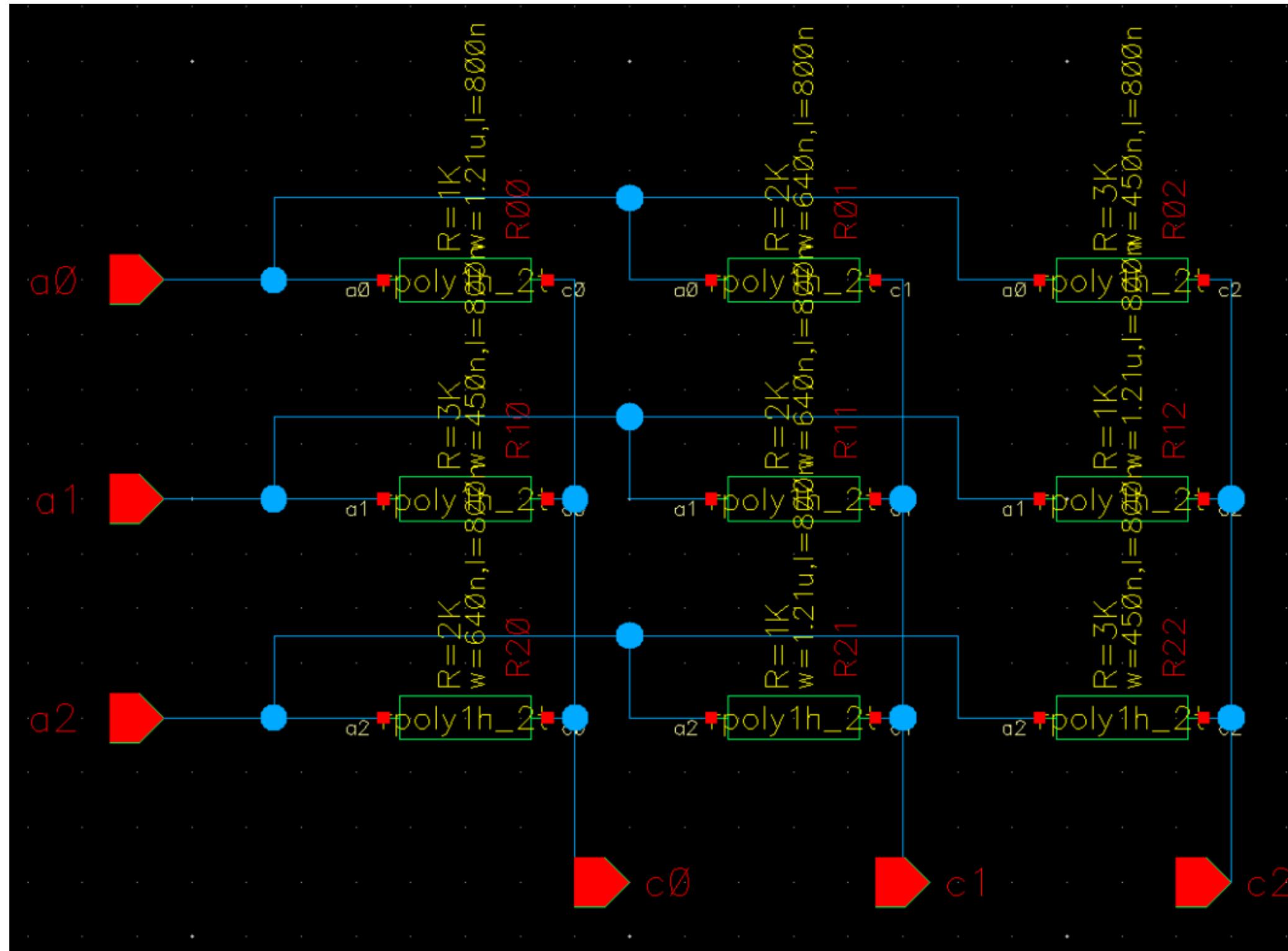
RRAM



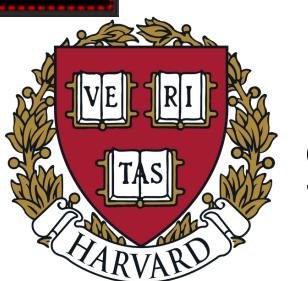
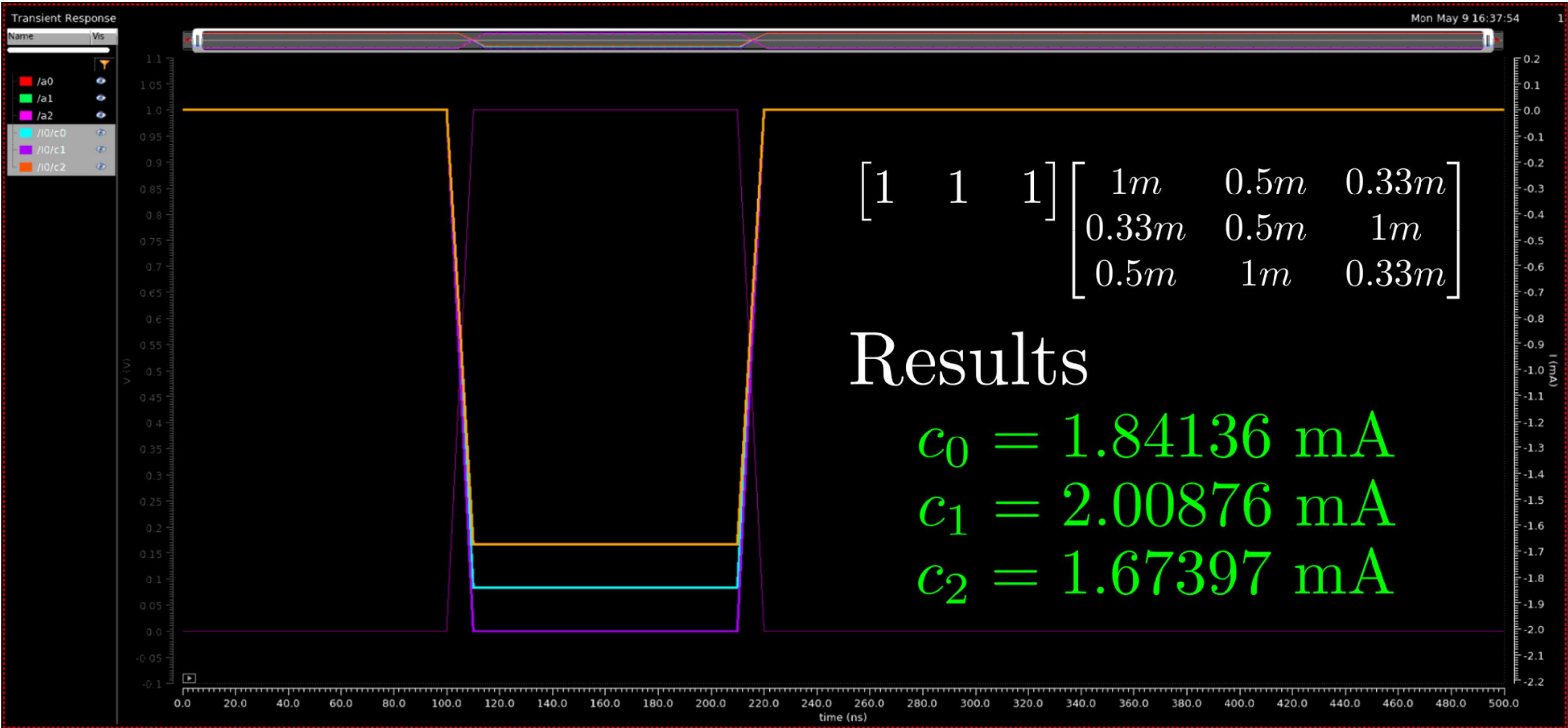
We have a **continuous** and **large** interval of resistances (**beautiful**)



# 1R case



# 1R case

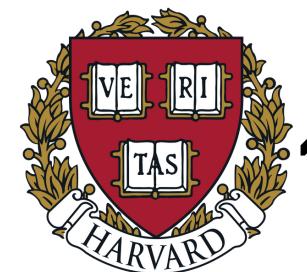
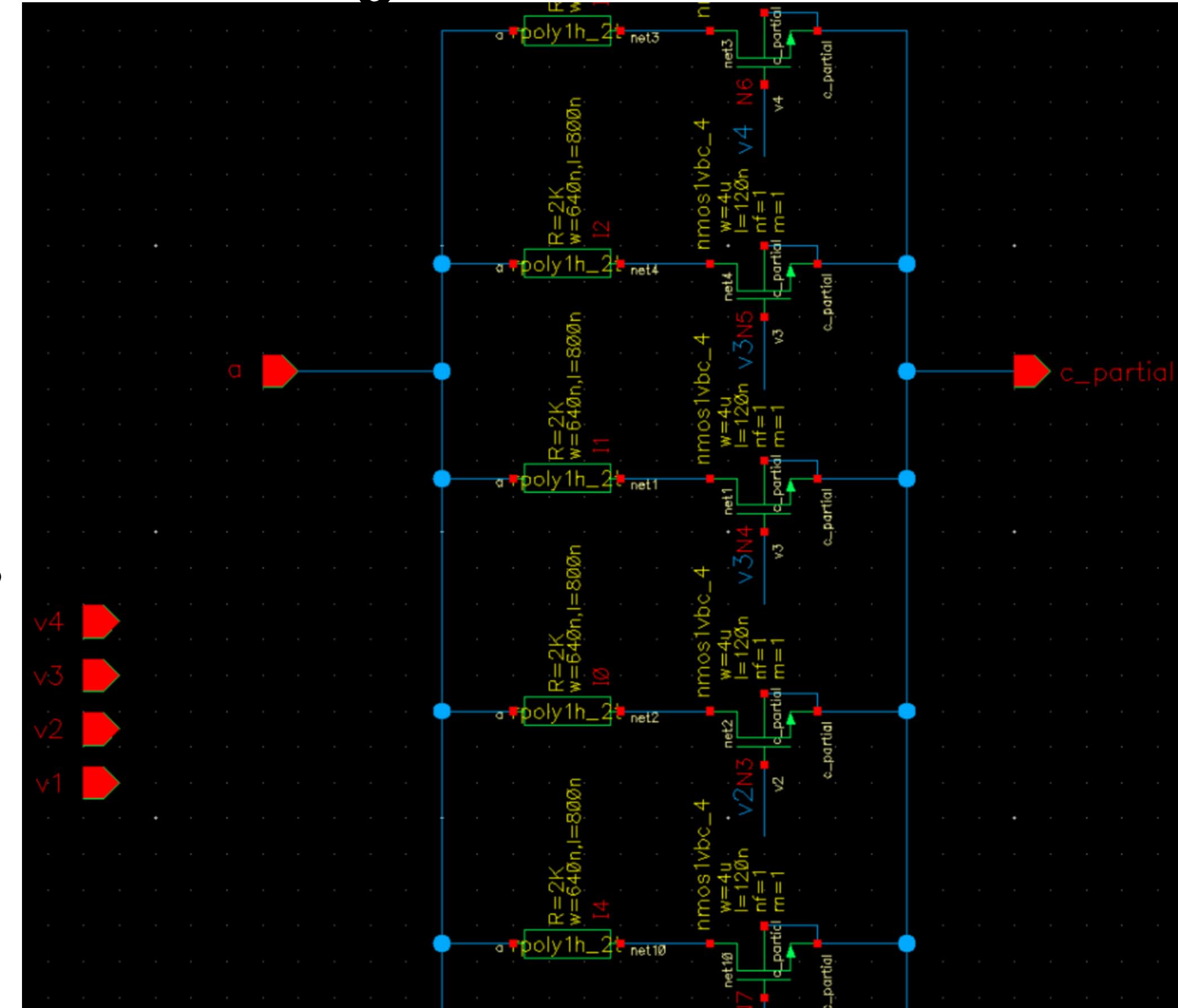


# 1T1R... actually more than 1T

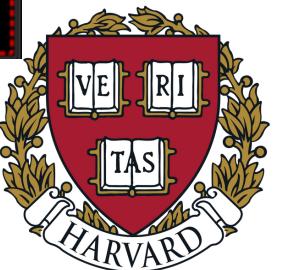
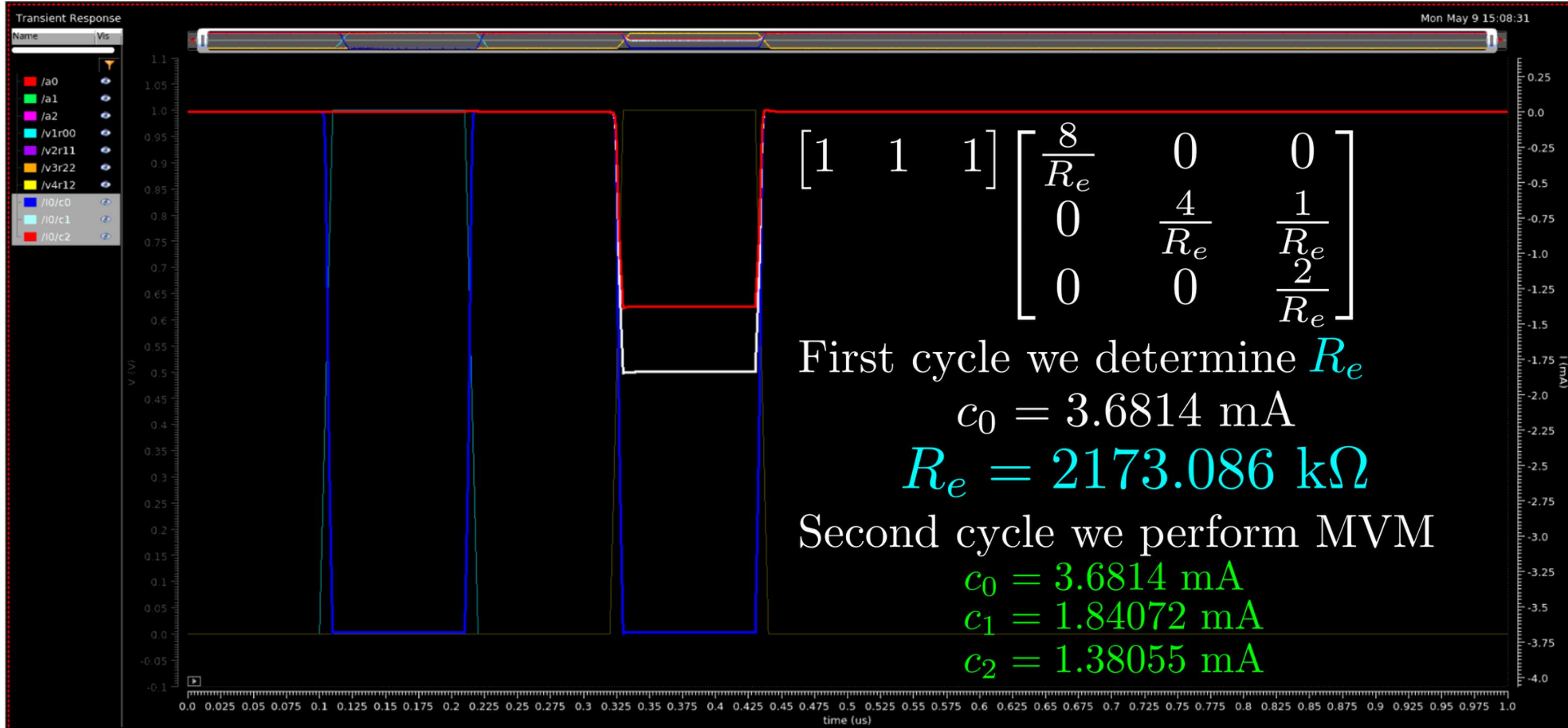
15T  
4-bit resistor

Watch out!

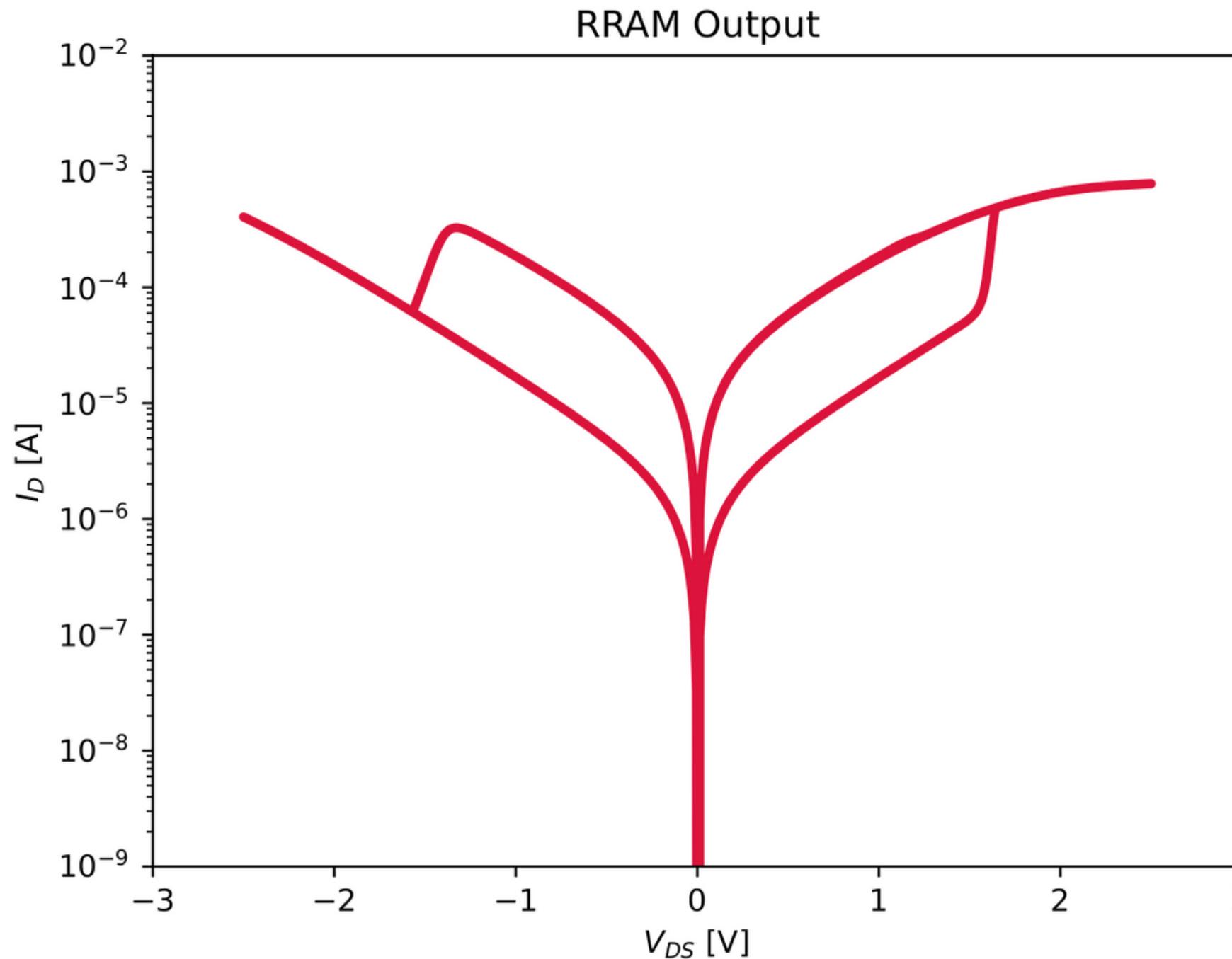
Transistors have  
internal resistance



# 1T1R... actually more than 1T



# Finally... RRAM!

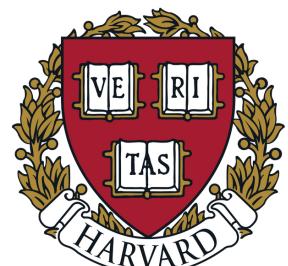


High and low  
resistance states

```
// =====
// RRAM
// -----
subckt rram0 ( TE BE )
parameters
// --- device parameters
+ L      = 5e-9          // Oxide thickness (m)
+ gap_min = 0.1e-9       // Min. gap distance (m)
+ gap_max = 0.8e-9//1.7e-9 // Max. gap distance (m)
+ gap_ini = 0.1e-9       // Initial gap distance (m)
+ a0     = 0.25e-9        // Atomic distance (m)
+ Eag    = 1.501           // Activation energy for vacancy generation (eV)
+ Ear    = 1.5              // Activation energy for vacancy recombination (eV)
// --- I-V characteristics
+ IO     = 6.14e-5
+ g0     = 2.7505e-10
+ V0     = 0.43
// --- gap dynamics
+ Velo   = 150
+ gamma0 = 16.5
+ g1     = 1e-9
+ beta   = 1.25
// --- temperature dynamics
+ T0     = 273+25         // Ambient temperature (K)
+ Cth    = 3.1825e-16      // Effective thermal capacitance (J/K)
+ Tau_th = 2.3e-10         // Effective thermal time constant (s)
// --- simulation time control
+ tstep   = 1e-9           // Max. internal timestep (s)
// --- resistance
+ Vread   = 0.1             // Read voltage (V)

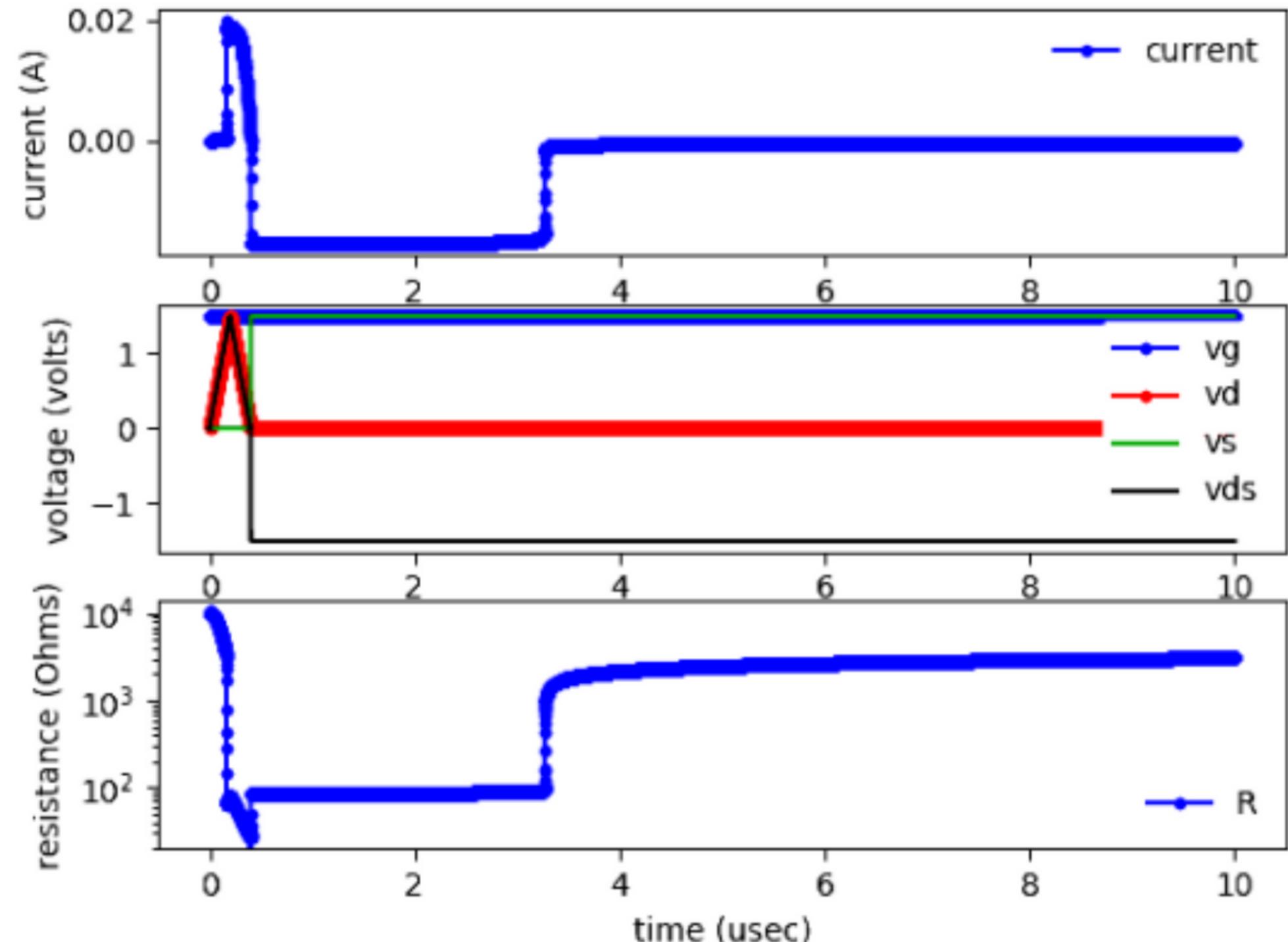
rram0_inst ( TE BE ) rram0
+ L      = L
+ gap_min = gap_min
+ gap_max = gap_max
+ gap_ini = gap_ini
+ a0     = a0
+ Eag    = Eag
+ Ear    = Ear
+ IO     = IO
+ g0     = g0
+ V0     = V0
+ Velo   = Velo
+ gamma0 = gamma0
+ g1     = g1
+ beta   = beta
+ T0     = T0
+ Cth    = Cth
+ Tau_th = Tau_th
+ tstep   = tstep
+ Vread   = Vread

ends rram0
// =====
```

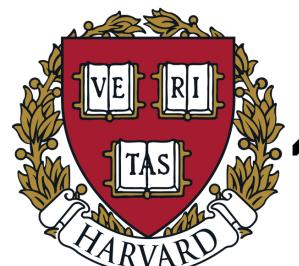


# How to program?

1. Turn **on** the wire
2. Trigger a state (high or low resistance)
3. The **length** of the pulse ( $V_D$ ) corresponds to the **resistance** you will get
4. Turn **off** the wire (and go to the next cell)



After programming, we can do the **multiplication**

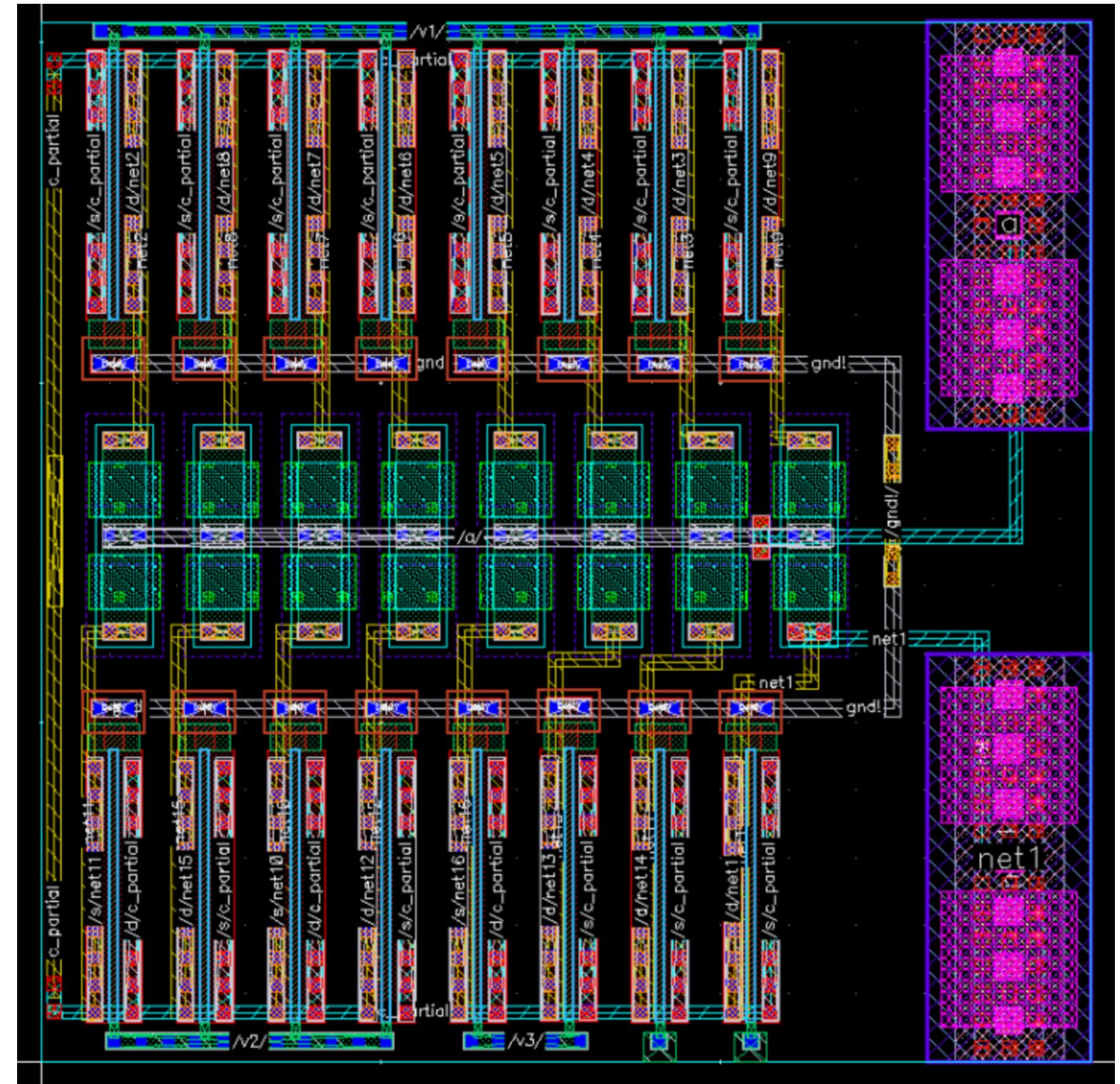


# Layout time!

15.45 um x

15.29 um

for each cell

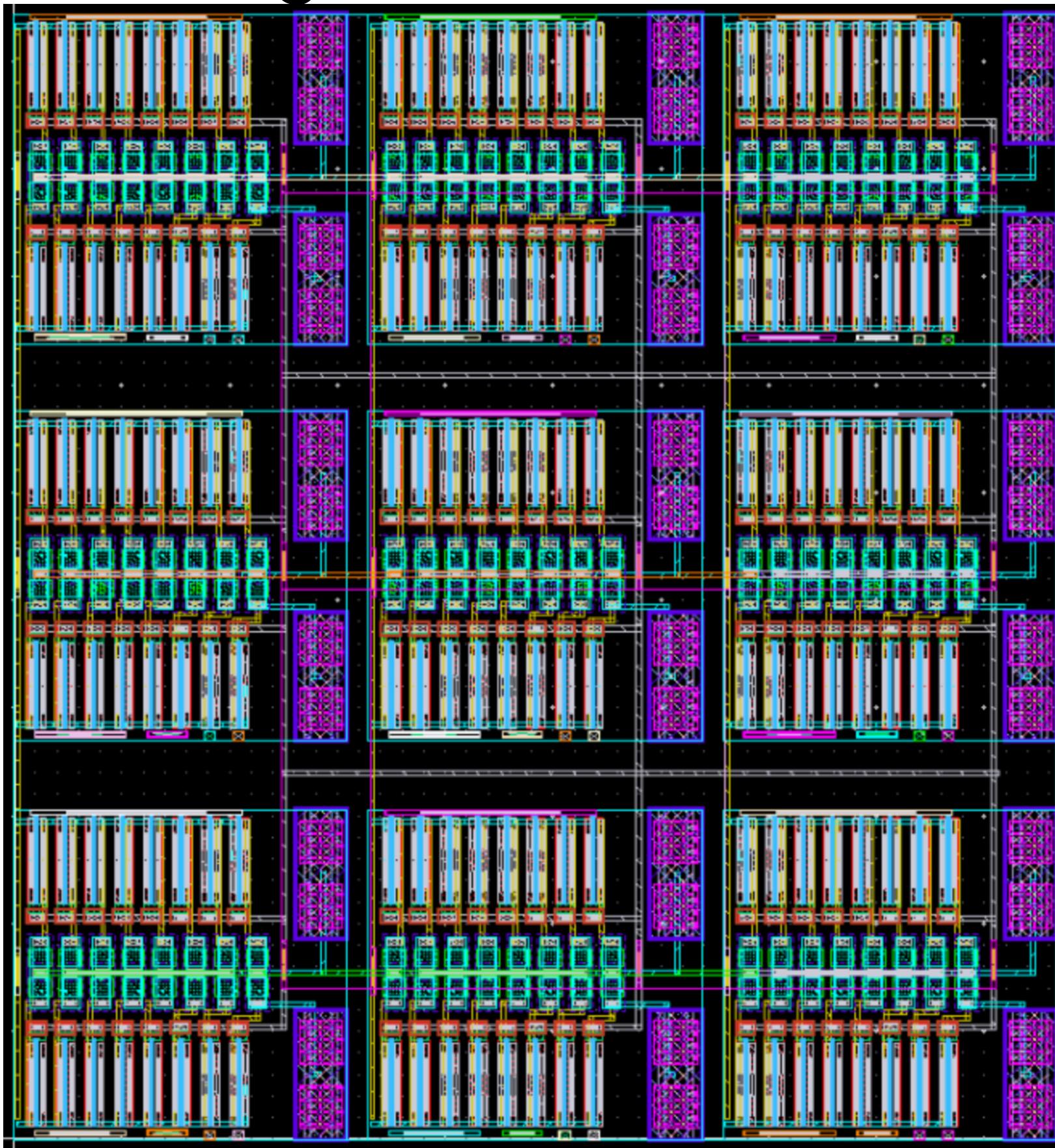


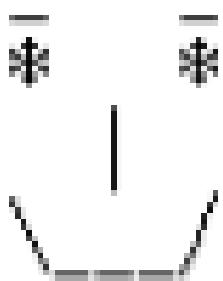
# Layout time!

48.35 um x

52.205 um

full layout  
(3x3 crossbar)





Thank you!