Exam EITP25 VT2020, Solutions

1.

a) Information in Spiking neural networks is transmitted through voltage spikes, and energy is only consumed for the generation of these spikes. In a regular artificial network power is consumed constantly as the data representation is constant. And training is performed by many matrix multiplications requiring a large overhead. **(2p)**

b) STDP is a learning mechanism present in biology in which the relative timing between a spike from a pre-neuron and post-neuron determines the magnitude and direction of synaptic weight change. In the most common implementation, prespikes preceeding postspikes lead to weight potentiation, while depression occurs for prespikes coming after a postspike. In biology the magnitude of the weight change decays exponentially with increasing $|\Delta t|$ such that spikes with little correlation does not affect each other much. In a memristor implementation the preneuron pulse has a voltage amplitude too low to achieve a conductance change in the memristor by itself. Upon post-neuron spike the post-neuron also sends a voltage pulse back along the synapse. The combination of the pre- and post-neuron pulse the bias across the memristor will cross the threshold for conductance change and thus will be potentiated or depressed depending on the relative timing according to the STDP rule described above. **(5p)**

c) High random input
→ Initial high random output (perfect integrator). However, assuming that LTD dominates over LTP synapse weights will overall decrease in strength, as probability is larger that a spike is in the depression time window
→ very weak output due to weak weights, only output on bunches of input spikes (coincidence detector), synapses with correlated spikes will be strengthened

→ Moderate output activity, moderate synaptic strengths on relevant synapses (timing matters) **(5p)**

2.

a) P_r = 25 µC/cm^2, E_c = 1.25 MV/cm. **(2p)**

b) C has lowest R contrast, because the insulator is thin and direct tunneling should be the dominant mechanism for transport. The effect of barrier height change is small on the resistance. A comes next as it has a thicker insulator the transport is dominated by thermal emission giving a stronger effect of barrier height. B is the one with strongest contrast as this relies on tunneling but has a modification of the barrier width thanks to the introduction of a semiconductor contact which goes between depletion and accumulation depending on direction of polarization. **(5p)**

c) $E_{dep} = E_c \rightarrow |P| = \varepsilon_0 \varepsilon_r \left(\frac{C_{di}}{C_{FE}} + 1\right), C_{FE} = \frac{\varepsilon_0 \varepsilon_r}{t_{FE}} = 1.77\ \mu F/cm^2 \rightarrow |P| = 10\ \mu C/cm^2$ **(5p)**

3.

a) Assuming conductance values specified in 8 bit values 0-255, where 0 = 0 S and 255 = g_max [S] (not necessary for full score). For g_max = 255 µS (arbitrarily chosen) the output currents of each bit line are: $I_1 = 75.5\ \mu A, I_2 = 71.5\ \mu A, I_3 = 127.5\ \mu A$.

The system maps to the linear equation system: $\boldsymbol{Ax} = \boldsymbol{b}$

Assume drive voltage V_0 = 2.5 V, which matches the largest input. It can be chosen larger but must be at least as large as the largest input. Then the input vector is $\boldsymbol{x} = \frac{\begin{bmatrix}0.05\\2.5\\0.5\end{bmatrix}}{V_0} = \begin{bmatrix}0.02\\1\\0.2\end{bmatrix}$, and the output vector is $\boldsymbol{b} = \frac{\begin{bmatrix}75.5\\71.5\\127.5\end{bmatrix}}{V_0} = \begin{bmatrix}30.2\\28.6\\51\end{bmatrix}$. The A matrix is the transpose of what's in the memory array i.e. $\boldsymbol{A} = \begin{bmatrix}30 & 25 & 23\\40 & 23 & 24\\100 & 45 & 20\end{bmatrix}$ **(4p)**

b) FORMING: A strong electric field across the dielectric causes oxygen ion diffusion (negative) into the positive contact (bottom contact), leaving behind oxygen vacancies. The filament of oxygen vacancies grows towards the negative contact (top contact) until it bridges the gap leading to a strong current. The predefined current compliance level now sets the amount in which the filament grows laterally. (one can also describe the above using mobile vacancies coming from the positive contact). The device is now in a low resistive state.
RESET: With voltage bias in the other direction oxygen ions is now diffusing back and disrupting the narrow end of the filament close to the top contact. After disruption the device goes into a high resistive state.
SET: The bias is now set to positive again and at a certain point oxygen ions move back towards the top contact and reform the end of the filament once more, taking the device back to the low resistive state. (**4p**)

c) Giant Magnetoresistance due to crystalline MgO barrier that reduces tunneling probability of non-spin polarized states,

Spin-torque transfer in which the spin of the electrons interacts with the spin of the atoms causing a torque (i.e. a pull) on it, a strong enough such pull (current) leads to a flipping of the magnetization.

Perpendicular Anisotropy is a magnetic anisotropy that is perpendicular to the plane of the film. Thus it is not depending on shape anisotropy and is therefore more scalable. Instead interface anisotropy shows strong perpendicular anisotropy. (**4p**)

4.

a)

1. Voltage drop along the WL due to finite resistance of line at scaled dimensions and leakage through half selected devices. Problem is handled (not really solved) by using the V/3 biasing scheme, in which unselected WLs are biased to V/3 and unselected BLs to 2V/3. The voltage on the selected WL cannot decay below 2V/3 in this setup.

2. Sneak paths through unselected devices. Devices that are not selected may still conduct current, in particular in their low resistive states. By going through several devices a sneak path can arise. Since there are MANY more sneak paths than the actual path of interest this sneak current can mask or at least significantly contribute to the measured current. The solution is the introduction of a nonlinear selector device that strongly reduces leakage currents through un/half-selected devices. (5p)

b) For both memory techs, high nonlinearity is key. For PCM high on current is necessary, but not bidirectionality as PCM is both SET and RESET with the same polarity. For FeFET, bidirectionality is necessary but not current level as this is a voltage switched device. (4p)

c) A = $8F^2 = 8*(30E-9)^2$ = 7.20E-15 m^2 (7200 nm^2).

A_tot incl peripherals = A/50% = 1.44E-14 m^2.

Die size 50 mm2 means 50E-6/1.44E-14 = 3.47E9 bits on chip. -> 434 MB/chip. (3p)

5.

a)

1. Asymmetric potentiation/depression
   (quickly severe degradation)
2. Cycle-to-cycle pot./dep. Variation
   (can be beneficial if small, but large degrades performance a lot)
3. Dynamic range
   (need more than 10x, but that's not so much)
4. Non-linear potentiation/depression
   (only causes minor degradation even for large non-linearities, assuming symmetric!)

(**4p**)

b) 2-PCM consists of one PCM device for positive weight and one for negative and one comparator so that the actual weight is the difference between the two conductances. The weight is the potentiated by SET:ing the potentiation device and depressed by SET:ing the depression device. As both conductances only increase at some point one needs to do a RESET of both and new SET of one of them to change the weight representation to one with lower total conductance. This leads to many fewer RESET events than SET events, which saves energy. (**4p**)

c) (**4p**)