Lecture 02 – CSE 40547/60547 – Computing at the Nanoscale

PN Junction Notes:

Let's start with a (very) short review of semi-conducting materials:

- N-type material:
  - Obtained by adding impurity with 5 valence elements to a valence 4 semiconductor
  - This will increase the amount of negative charge carriers

- P-type material:
  - Obtained by adding impurity to increase the number of positive charge carriers
  - Accepts only weakly bonded, outer electrons from other semi-conductor atoms
  - Semi-conductor atoms that lose electrons are holes
  - Purpose of p-type doping is to create an abundance of holes
    - Boron or aluminum substituted into a silicon crystal lattice
    - Result:
      - 1 electron missing from 1 of 4 covalent bonds
      - Dopant can accept electron from neighboring atom to complete the 4^{th}

- Bringing P-N type material together creates a PN junction and diode

  Picture: **(1)**, **(2)** two diode symbols, **(3)** diode cross-section, **(4)** Lewis structure

- Bringing P, N-type material together causes a large concentration at the gradient boundary

  - There's a higher electron concentration in N material, lower electron concentration in P material
  - Gradient causes:
    - Electrons to diffuse from N → P
    - Holes to diffuse from P → N

  - When a hole leaves the P material, it leaves behind an immobile acceptor ion (with a negative charge)
    - Therefore, a *negative* charge will exist in the P material near the vicinity of the PN-boundary
  - Similarly, *positive* charge builds up on the N side – as a diffusing electron leaves a positive charge behind

  Picture: PN-junction with depletion region

- At the depletion region, there is an electric field across the boundary directed from N → P

  Picture: Depletion region; field will counteract diffusion, and eventually there is 0 net current flow

  Therefore, under 0 bias, a voltage ϕ_0 exists which serves as a built-in potential.

- If we raise the potential of the P region with respect to the N region, we *lower* the potential barrier
  - I.e. given a forward voltage V_d applied to the junction…
  - The flow of mobile barriers across the junction increases as diffusion current is dominated by drift current
  - The net result is current flowing in the diode from P → N in this *forward biased mode*

- In *reverse bias mode*, the depletion region increases, and current flow is shut off

Diodes for Boolean Logic:

Questions to consider:

(1) How can we make logic gates from diodes?
(2) Are diodes a "good" switch for making logic gates? (Why or why not?)

Answer to Question (1):

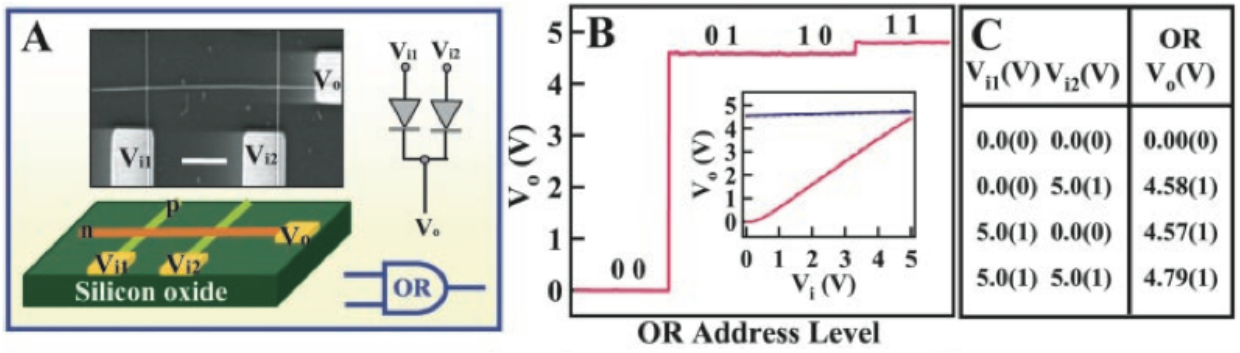
Picture: (a) diode based OR gate, (b) diode based AND gate

- OR gate operation:
  - If 1 input (A or B) is high, current flows through the associated, (forward biased) diode and brings the output node up.

- AND gate operation:
  - If 1 input (A or B) is grounded, current flows through the diode and output X is at a low voltage
    - (assumes that diodes in the above picture are forward biased)
  - The only way to bring / keep output X high is for both inputs to be high
    - (i.e. the diodes need to be reverse biased)

- Foreshadowing:
  - What if I told you, that you could build a diode in an area that's approximately 25 nm^2?
    - Keep in mind, that the minimum feature size (which represents just 1 part of 1 device) for the current technology node is ~ 16 nm.
  - This is an appealing thought, but diodes have their issues – i.e. how is *gain* achieved?

Picture: Experimental results of diode-based, OR gate made from nanowires



C V_{i1}(V)	V_{i2}(V)	OR V_o(V)
0.0(0)	0.0(0)	0.00(0)
0.0(0)	5.0(1)	4.58(1)
5.0(1)	0.0(0)	4.57(1)
5.0(1)	5.0(1)	4.79(1)

Example setup:

- For both AND, OR gates, let's assume V_{high} = 5V[1], V_{low} = 0V
- Also, (for now) we'll assume that diodes don't introduce any errors or loses into circuits
- In actuality, this is not really the case – a silicon-based diode would have a drop of ~0.7V while conducting
- Maybe we could overcome this problem by saying:
    - If voltage is >> 3.5V, we have a logic 1
    - If voltage is << 1.5 V, we have a logic 0

Example:

- How would we use diodes to implement the logic function: AB + CD

Picture: (a) AB+CD based on diode logic; (b) voltage divider

Example discussion:

- If all inputs are logic 0 (a low voltage) the output will be held at logic 0.
- If both inputs to one of the AND gates are a logic 1, what will the output be?
    - Diode in OR gate will become forward biased for the AND gate where both inputs are at a high voltage
    - Current flows through the AND gate resistor and the OR gate diode
    - Current flows through the OR gate resistor
- If all resistors are the same value, we have a voltage divider and equally share a (5V) supply
- After the OR gate diode inserts its loss, the output voltage may be on the order of ~2.1V – 2.2 V

Question:

- What logic state is represented by 2.1V?

[1] In reality, more like 0.9V, 1V now, but 5V is fine for illustrative purposes.

Take Aways:

- For diode logic, it's difficult to create cascades of multiple logic levels
  - (The voltage drops add up)
- This leads us to 3 terminal devices
- Going back to the nanowire example:
  - It's great to have junctions that are 10s of nm^2
  - But, as seen above, diode-based logic may be limited as an information processing technology
  - Moreover, how do you do inversion?
    - For a functionally complete logic set, need AND, OR, + NOT.
- Can engineering gain, inversion into nanowire-based circuits (as you will see later), but overhead strips away potential density and performance gains

Digression:

So, ideally, what features should a device have it is to be used to implement a digital logic system?

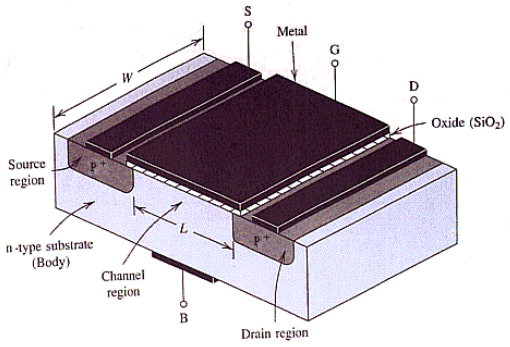
Consider "5 tenets" of digital logic:

1. a device should have non-linear response characteristics
2. a device should enable a functionally complete logic set
3. power amplification (or gain) is needed
4. the output of one device must drive another (i.e. with no state variable change)
5. dataflow directionality must be well defined

Use nanomagnet logic as an example…

Transistor Basics:

Disclaimer: *This discussion is by no means meant to replace an electronics or VLSI design course. Rather, it is just meant to briefly review/introduce material from these courses, and set the stage for a discussion of roadblocks to transistor-based scaling and computation.*



Devices at a glance:

- At the most superficial level, transistors can be thought of as a switch

- If a voltage is applied to a gate (that is greater than some threshold voltage V_t), a conducting channel is formed between drain and source

- If the voltage difference between drain and source, than current flows between them
  - The greater the voltage difference between gate and source, the smaller the resistance of the channel and the greater the current

- If the gate voltage is lower than the threshold, no channel exists and the switch is open

(Assuming an NMOS device, current is carried by electrons moving through an N-type channel between the transistor's source and drain.)

Threshold Voltage:

- Consider $V_{gs} = 0$ → the drain, source, and bulk are all connected to ground

- The drain and source are effectively connected by back-to-back PN junctions
  - (substrate-source and substrate-drain)

- With the conditions above, both junctions have 0 bias – and the device can be considered "off"
  - (i.e. there is a high resistance between source and drain indicative of an open switch)

See Slides 6-8

- Now, assume that a positive voltage is applied to the gate (with respect to the source)
  - The gate and substrate form plates of a capacitor where the gate oxide forms the dielectric
  - A positive gate voltage causes +/- charge to accumulate on gate/substrate side respectively

See Slide 6

- When the gate voltage becomes sufficiently high, the potential at the silicon surface reaches a critical value, and the semi-conductor surface inverts to N-type (strong inversion)
  - Further increases to the gate voltage produce no more changes in the depletion layer width, but results in a thin inversion layer under the oxide
    - Drawn into inversion layer from heavily doped N source region
    - N-type channel formed between source and drain, modulated by V_{gs}
    - Value of V_{gs} where strong inversion occurs = threshold voltage V_t

5 **See Slide 6**

Resistive / Linear Operation:

- Now, assume $V_{gs} > V_t$ and a small voltage is applied between drain and source…
- Voltage difference causes current i_D to flow from drain to source
- At a point *x* along the channel, the voltage V(x) at gate-to-channel voltage at that point = $V_{gs} – V(x)$
- If voltage exceeds V_t at all points along the channel, current can be calculated

See Slides 9-13

Saturation Region:

- As the value of the drain-source voltage is further increased, the assumption that the channel voltage is greater than the threshold voltage at all points ceases to hold
  - Occurs if $V_{gs} – V(x) < V_t$
- Here, the conducting channel is *pinched off*
- Conditions must be met at the drain region…therefore $V_{gs} – V_{ds} <= V_t$
  - E.g. if $V_{gs} = 5V$, $V_{ds} = 5V$, and $V_t = 0.7V$ → $5V – 5V < 0.7V$
- Now, the transistor is in the saturation region, and current is no longer a function of V_{ds}
  - (it acts as a current source)

See Slides 14-16

How scaling impacts transistor parameters, performance:

See Slide 17

- Let's talk more about the voltage scaling factor U.
  - Assumes all voltages scale by the same factor U
  - Product compatibility might be one reason

A bit more on 3 models in the chart:

Full scaling:
- Voltages and dimensions are reduced by the same factor S; leads to:
  - Greater density
  - Better performance
    - (intrinsic delay gets better)
  - Power consumption reduced to
    - (and chip-level power density stays constant as more devices placed in same area)
- For a long time, this model held

- Consider impact on something like Power:
  - $P = I_{sat} \times V$
    - I_{sat} scales as 1/S
    - V scales as 1/S
    - Therefore P scales as $1/S^2$ → seemingly great!
  - What about t_{ox}?
    - Affects other parameters which influence delay and power
    - What if it doesn't scale in lockstep (as it must)

See Slide 18 for t_{ox} scaling trends

Fixed Voltage Scaling:
- To keep new parts compatible with existing components, V cannot be scaled arbitrarily
- Can be hard, expensive to support multiple supplies (e.g. 5V, 3V)
- What if voltage does not scale?
  - Look at example: $P_{density}$
    - $P_{density} = P / Area$
    - Scales as $1 / (1/S^2)$ → increases as S^2!

General Scaling:
- Supply voltage scales, but not at the same rate as technology
- Some fundamental limits come into play here…
  - You can't make V_t arbitrarily low – as you'll see, you couldn't turn the transistor off
  - There are intrinsic device voltages
    - Material parameters can't be changed
    - Need new materials