CS257 Introduction to Nanocomputing

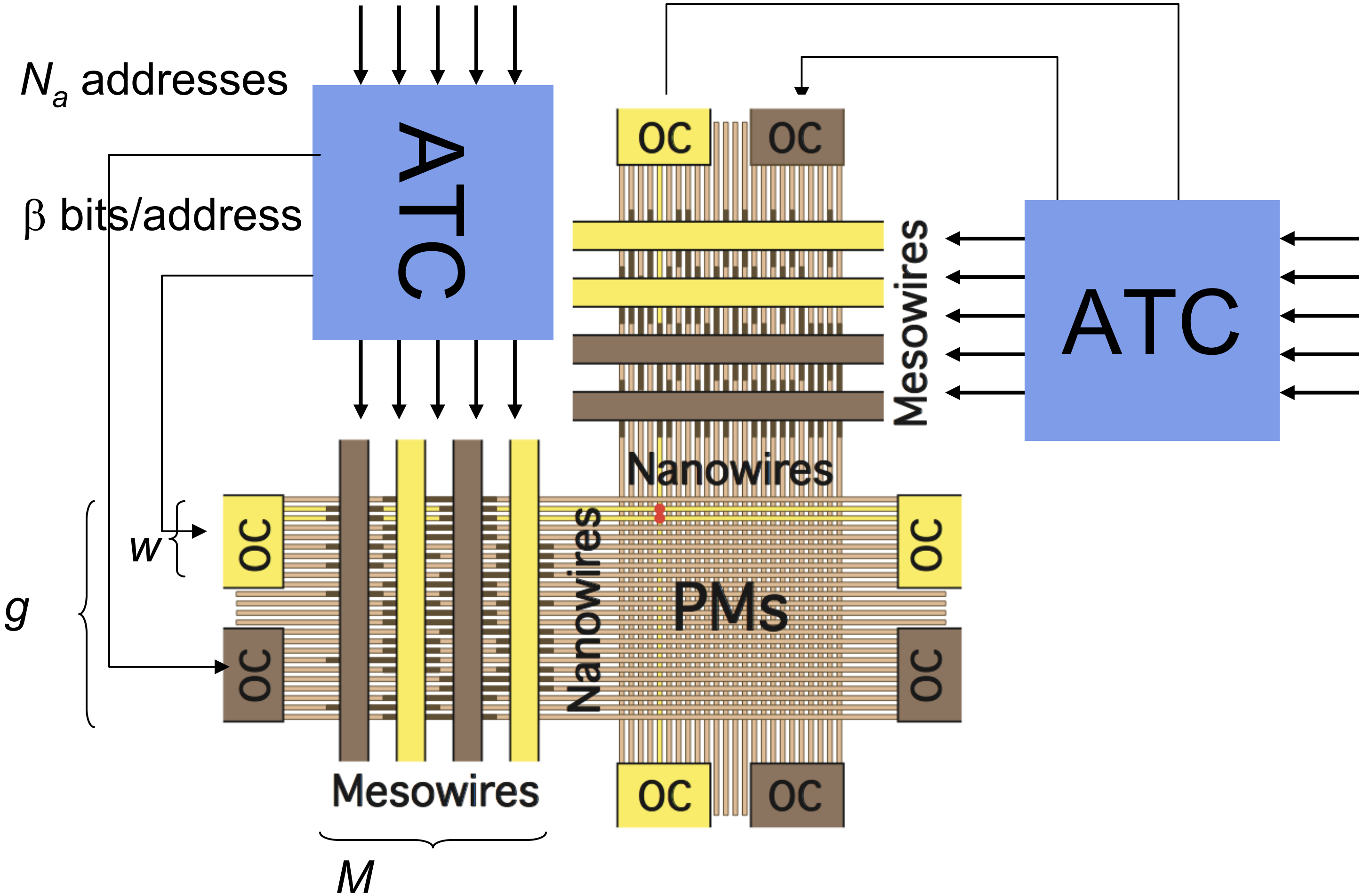
Undifferentiated NW Decoders

John E Savage

Lecture Outline

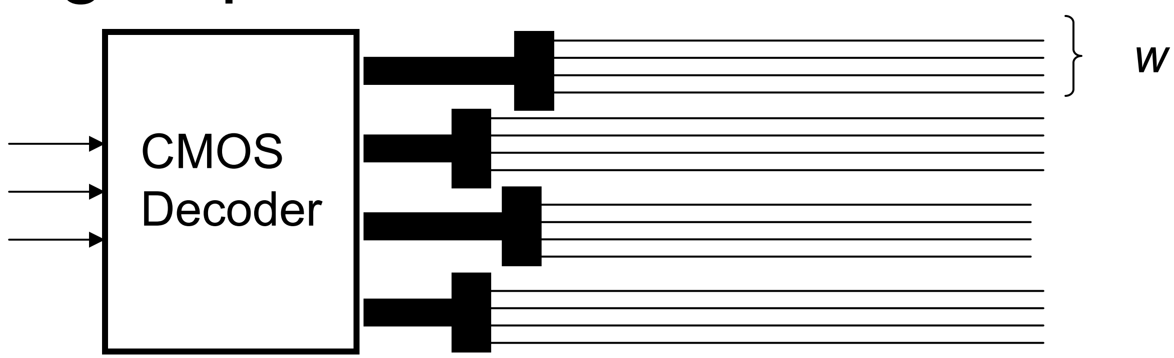
- Two undifferentiated NW Decoders
  - Randomized-contact decoder
  - Randomized mask-based decoder
- Analysis of “Take What You Get”

The Crossbar Memory



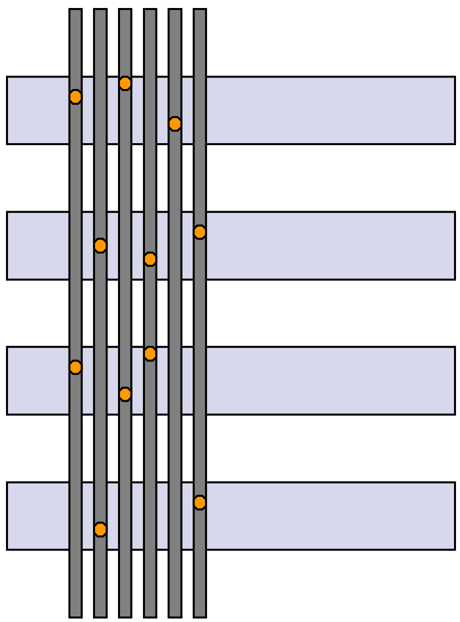
Reducing the Area of the ATC

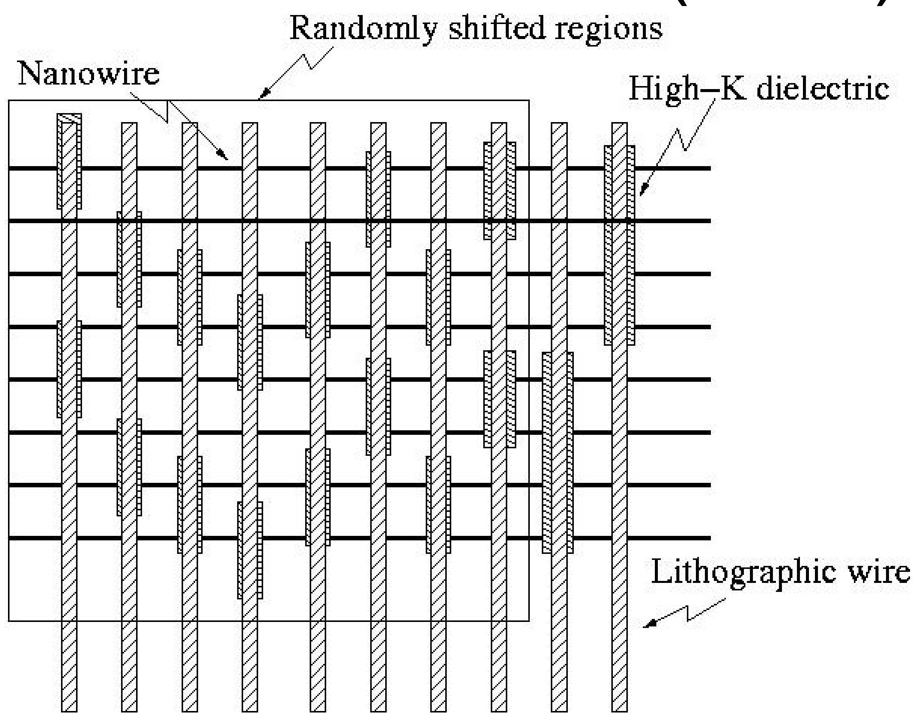
- The ATC has one word for each of the N_a addressable NWs.
- Area of the ATC can be reduced by storing inputs to a CMOS decoder, not one bit per contact group.



How to Differentiate NWs?

- Randomized-contact decoder (RCD)
- Randomized mask-based decoder (RMD)





Codewords Assigned During Decoder Assembly

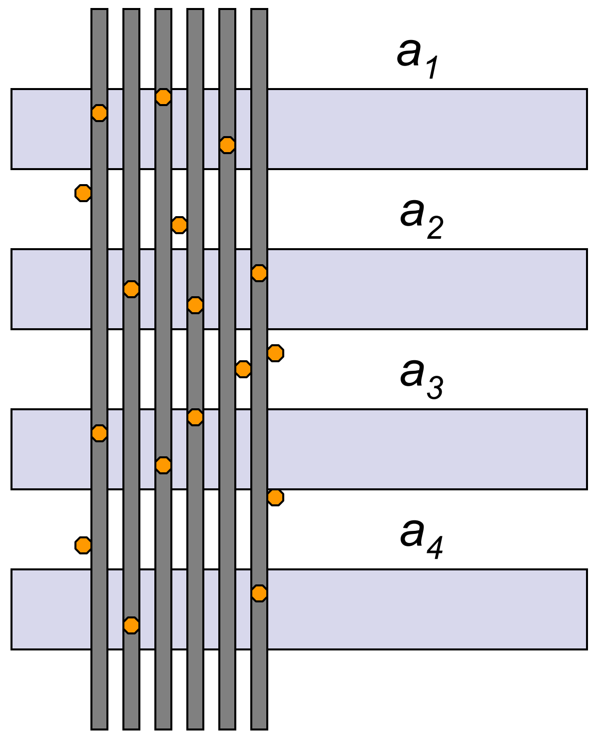
- RCD and RMD both assign codewords stochastically.
  - In RCD codeword bits are uncorrelated
  - In RMB codeword bits are correlated.

- What effect does correlation among codeword bits have on the number of MWs needed to ensure that all codewords are individually addressable?

Randomized-Contact Decoder

Randomized-Contact Decoder

- Contacts made at random between NWs and MWs.
- If contact made, MW controls NW, i.e. NW resistance is increased.
- Control of NW may not be complete, source of error.



Issues in Assembling NW Decoders

- NW decoders are assembled stochastically.
- Can't predict which NW addresses will occur.
- Some NWs cannot be controlled.
- Under what conditions can many NWs be addressed?
  - What's the probability that a decoder has N_a addressable NWs?
  - How do N_a and probability depend on addressing strategy?

Ideal and Non-Ideal Decoder Models

- If NW is controlled, uncontrolled, ambiguous by j^{th} MW,

$$c_j = 1, 0, e$$

- NW codeword $\boldsymbol{c} = (c_1, c_2, \ldots, c_M)$

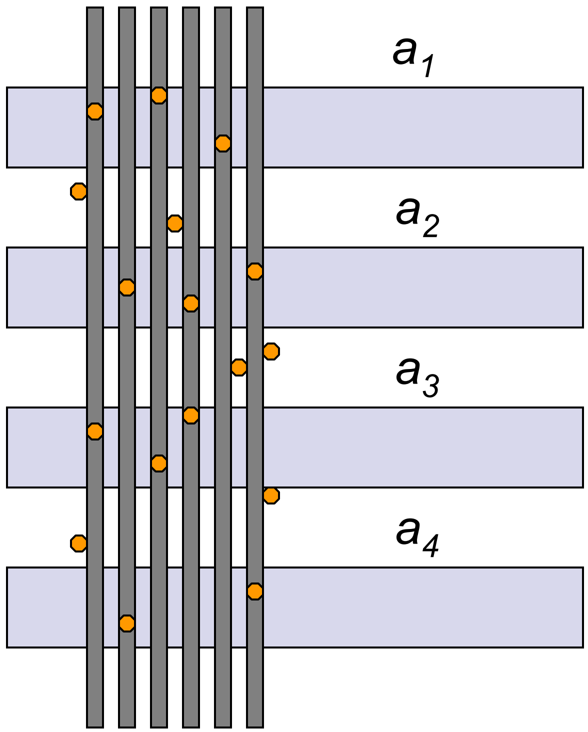
- Ideal (non-ideal) resistive model
  - $c_j = 1$ if resistance $= \infty$ ($> r_{high}$) when j^{th} MW active
  - $c_j = 0$ if resistance $= 0$ ($< r_{low}$) when j^{th} MW active
  - $c_j = e$ (error) otherwise.

NW Addressability

- MW address
  - $\boldsymbol{a} = (a_1, a_2, \ldots, a_M)$ where $a_j = 1$ if j^{th} MW is active
  - NW is “on” if its resistance is “low.”
  - A set of NWs is “off” if cumulative resistance is “high.”
- A NW is **individually addressable** (i.a.) if for some address $\boldsymbol{a}$ it is “on” & all others are “off.”
- In ideal model, codeword $\boldsymbol{c}$ activated by address $\boldsymbol{a} = \overline{\boldsymbol{c}}$ (Boolean complement).

History of the Randomized-Contact Decoder (RCD)

- Kuekes and Williams 2001 patent.
- Hoggs, *et al* (IEEE Trans. Nano, March 2006)
  - Analyzed idea using simulation & empirical analysis
- Our contributions
  - Tight probabilistic analysis of RCD.
  - Application to decoder with errors.
  - Identification of a very good addressing strategy.



Nanowire Addressing with Randomized-Contact Decoders1, Eric Rachlin, John E. Savage, Procs. IEEE/ACM Int. Conf. on Computer-Aided Design (ICCAD), pp. 735-742, 2006.

RCD Model

- g contact groups, w NWs/group, $N = gw$ NWs
- N_a = number of i.a. NWs
- Calculate probability that N_a NWs are i.a.
  - p = probability a MW controls a NW
  - q = probability a MW doesn't controls a NW
  - $r = 1\text{-}p\text{-}q$ = probability error in MW controlling NW
  - An error occurs if MW control is uncertain.

Three Decoder Addressing Strategies

- **All Wires Addressable (AWA)**
  - In every contact group all wires are i.a.
- **All Wires Almost Always Addressable (AWA^3)**
  - Most contact groups satisfy AWA.
- **Take What You Get**
  - Use all i.a. NWs in all contact groups.
- Determine N_a, number of i.a. NWs
- Use N_a to calculate area of ATC.

Hoeffding's Inequality

- Analyze number of different NW codewords using Hoeffding's Inequality. Let $S = n_1 + \ldots + n_t$ where $\{n_i\}$ are ind. r.v.s in $a_i \leq n_i \leq b_i$. For $d > 0$ and $c_i = b_i - a_i$.

$$P\left(E[S] - S \geq d\right) \leq e^{-2d^2/\sum c_i^2}$$

"Take What You Get" Strategy

$$P\left(E[S]-S\geq d\right)\leq e^{-2d^2/\sum c_i^2}$$

Theorem Let N_a be total no. addressable NWs in a decoder with g contact groups, w NWs per group, and $N = gw$ NWs.

$$P(N_a\leq E[N_a]-Nk)\leq e^{-2k^2Nw/(w-1)^2}=e^{-2k^2g^*}$$

for k >0 and $g^* = g(w/(w-1))^2$.

Proof Let $t = g$, $d = Nk$, $S = N_a$, $c_i = (w-1)$ and B be lower bound to $E[N_a]$. Set $\kappa N = B - kN$

Bounds on Addressable Wires Using Hoeffding's Inequality

Theorem Let N_a be the no. of addressable NWs in an RCD with g contact groups, w NWs per group, and $N = gw$ NWs in total. If $\kappa \leq 1 - \sqrt{-\ln(\varepsilon/(2g^*))} - (w-1)(1-pq)^M$,

$$P(N_a > \kappa N) \geq 1 - \varepsilon$$

Proof $k = B/N - \kappa$. Set κ such that $e^{-2k^2 g*} = \epsilon$

To bound $E[N_a]$ let $x_j = 1$ if n_i is i.a., else 0. Event $e_{k,j}$ is true if n_k is on whenever n_j is on. Thus, n_j is not i.a. if $e_{k,j}$ is true for some k not j. $E[N_a] = gw\, P(x_j = 1)$.

Bounds on Addressable Wires Using Hoeffding's Inequality

Proof (cont.) But $P(x_1 = 1) = 1 - P(x_1 = 0)$.

$P(x_1 = 0) = P(e_{2,1} \cup e_{3,1} \cup \ldots \cup e_{w,1})$

$\leq P(e_{2,1}) + \ldots + P(e_{w,1}) = (\text{w}-1)P(e_{2,1})$.

But $P(e_{2,1}) = (1-pq)^{\text{M}}$. Thus,

$P(x_1 = 1) \geq 1 - (w-1)(1-pq)^{\text{M}}$ and

$$E[N_a] \geq B = gw\,(1 - (w-1)(1-pq)^{\text{M}})$$
$$= N\,(1 - (w-1)(1-pq)^{\text{M}})$$

or $\kappa \leq 1 - \sqrt{-\ln(\varepsilon/(2g^*))} - (w-1)(1-pq)^M$.

"Take What You Get" Strategy

Note: Let p = q = ½, $w = 8$, $g = 175$, $N = 1{,}400$, $\varepsilon = .01$, and $\kappa = .733$.

Since $g^* = g\,(w/(w-1))^2$ the following condition

$$\kappa \le 1 - \sqrt{-\ln(\varepsilon/(2g^*))} - (w-1)(1-pq)^M$$

is satisfied with M = 13 and $\kappa N = 1027$.

- That is, $N_a \geq 1{,}027$ with probability $\geq .99$ when starting with 1,400 NWs, $w = 8$, $M = 13$.

Bounds for Other Strategies

See paper.

Comparison of Addressing Strategies

- Assumptions
  - Area of ATC used to make comparisons
  - Error-free comparisons (p+q = 1)
  - Goal – obtain about N_a = 1,000 addressable NWs.
- **All Wires Addressable**
  - In every contact group all wires are i.a.
- **All Wires Almost Always Addressable**
  - Only use contact groups in which all wires are i.a.
- **Take What You Get**
  - Use all i.a. NWs in all contact groups.

Comparison of Addressing Strategies

- Strategies
  - All Wires Addressable (AWA)
    - N_a = 1,024 for M = 47, g = 128, N = 1,024.
  - All Wires Almost Always Addressable (AWA^3)
    - N_a = 1,024 for M = 30, g = 133, N = 1,064.
  - Take What You Get (TWYG)
    - N_a = 1,027 for M = 13, g = 175, N = 1,400.
- Which strategy is best?
  - Second better than first. Is third better than 2^{nd}?

Area Estimates

- Area of crossbar
  - ATC – ρN_a ($M + \log_2 g$), ρ = area of a CMOS bit
  - Standard decoder – $\lambda_{meso}^2\, g \log_2 g$
  - NWs + MW area – $(M\,\lambda_{meso} + N\,\lambda_{nano})^2$
  - Assume $\lambda_{meso} = 10\,\lambda_{nano}$, $\rho = 100\,\lambda_{nano}^2$
- Area Comparisons Between AWA, AWA^3, TWYG
  - Can ignore area of standard decoders
  - ATC: AWA >> AWA^3 >> TWYG
  - NWs + MW area: AWA > AWA^3; TWYG > AWA, AWA^3
  - However, sum of areas is smallest for TWYG.

Take What You Get Strategy RCD vs Uniform NW Decoders

- RCD
  - N_a = 1,027 for M = 13, g = 175, w = 8.
- Encoded NW Decoder
  - $M/2$-hot NWs (with .8 penalty for misalignment)
    - N_a = 1,033 for M = 8, g = 180, w = 8.
  - Core-shell NWs (no misalignment penalty)
    - N_a = 1,013 for M = 12, g = 190, w = 8.
- RCD competitive (M is reasonable).

The Effect of Faults

- The effect of faults measured by $r = 1-(p+q)$.
- M set so $\kappa = 1- \sqrt{-\ln(\varepsilon/(2g^*))} -(w-1)(1-pq)^M = .733$
- Number of MWs for **Take What You Get**

$$M \geq \frac{-\ln(\kappa - \sqrt{-\ln(\epsilon/2g*)})}{-\ln(1-pq))}$$

- Errors change M by factor $\ln(3/4)/\ln(1-pq)$.

p = q	α	*M*
.5	1	13
.4	1.69	22
.3	1.82	40

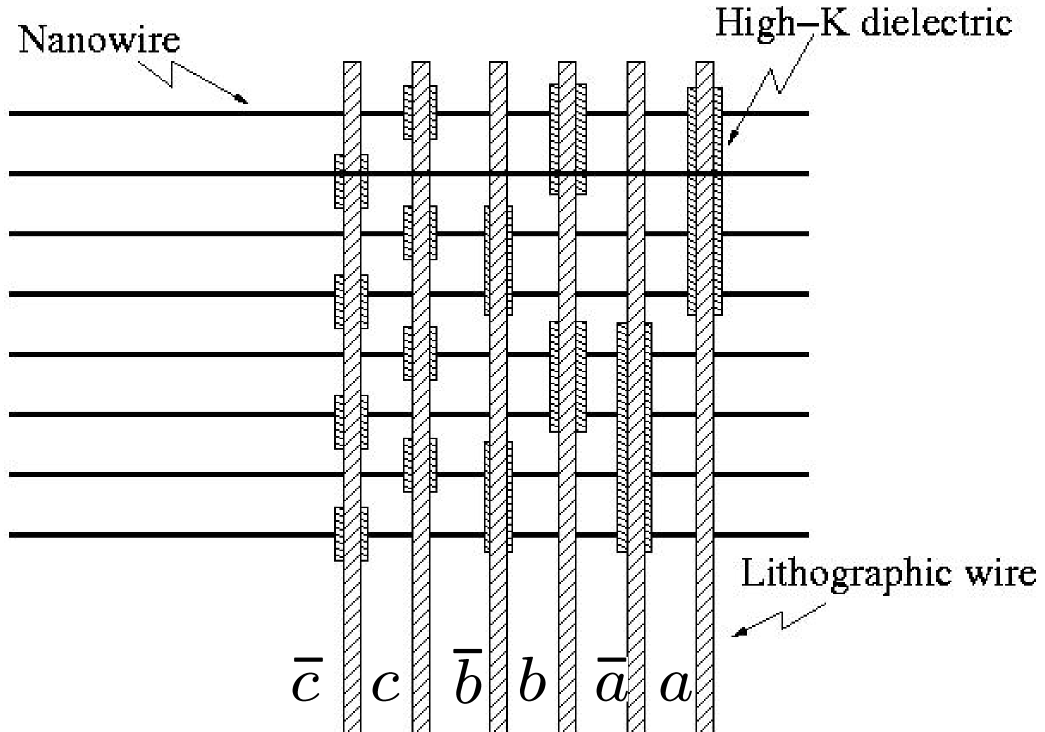
Conclusions About RCDs

- An area efficient NW RCD addressing strategy identified.
- Analysis shows the impact of faults.
- RCD shown to be a competitive decoder.
  - May be easier to implement than other methods.
- Because **Take What You Get** needs no more than M = 13, simulation is possible.
  - Simulation shows that M ~= 10 suffices!
- The importance of analysis firmly established.

Randomized Mask-Based Decoder

Logarithmic Mask-Based Decoder

- High-K dielectric regions couple NWs & MWs
- Deposit high-K dielectric regions under MWs



Problems with Logarithmic Mask-Based Decoder

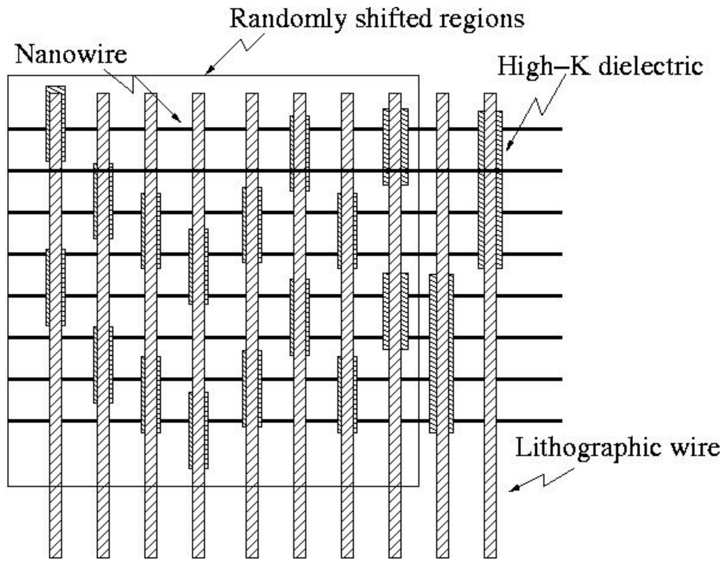
- Can't make regions as small as NW pitch
  - Lithography can't reach nm dimensions
- Can't position regions deterministically
  - At nanometer scales, positional inaccuracy is large
  - Inaccuracy is fractions to multiples of a NW pitch
- Approach: exploit natural randomness

Role of Logarithmic Decoder

- Use standard decoder to resolve uncertainty from N NWs to sets of *w* NWs.
  - #MWs = $2 \log_2 (N/w)$ for mask-based decoder
- Use randomized *linear* decoder (coming) to resolve decoding down to one NW.
  - Method can guarantee success to within some predetermined probability

Randomized Linear Decoder

- Randomly shift smallest litho regions (LRs).
  - Placement of LRs via masks is random
- *w* is width and separation of LRs.
  - *w* is fixed!



Model of Linear Decoder

- Deterministic logarithmic decoder resolves set of conducting NWs down to *2w* NWs
  - It deterministical leaves *2w* NWs conducting
  - The remaining NWs are non-conducting

- Random linear decoder resolves uncertainty down to one NW with high probability
  - It uses multiple randomly displaced LRs

Controllability of NWs by MWs Due to Placement of LRs

- A NW region is **controllable** by a MW if an LR under it covers the NW enough so a MW field can turn it off
- A NW region is **noncontrollable** by a MW if an LR under the MW doesn't cover the NW enough that a MW field can turn it off
  - This condition can be avoided by making LRs long enough
- A NW region is **ambiguous** w.r.t. a MW if it is neither controllable or noncontrollable.

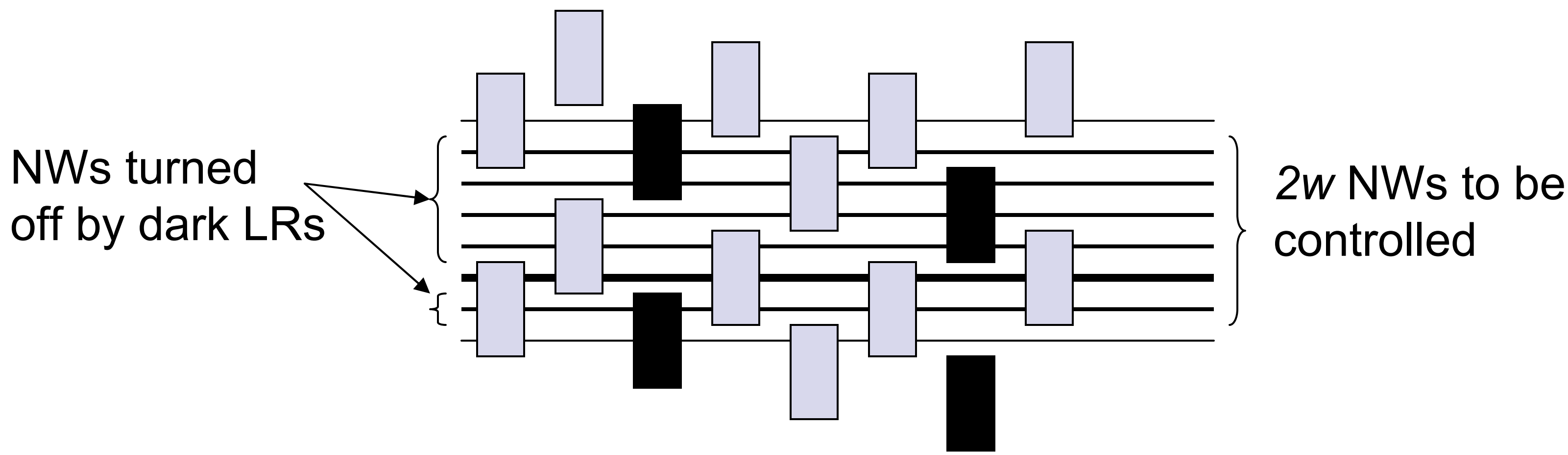
Conditions for Individually Controllable Linear Decoder NWs

- Under what conditions can a NW be turned on without turning on other NWs?
  - Let $I_{a,j}$ be intersection between NW n_a and MW j.
  - Let $C(I_{a,j}) = 0\ (1)$ if MW j can (cannot) control n_a.
- Let $J_a = \{j \mid C(I_{a,j}) = 0\}$
- If $J_a \subseteq J_b$ and n_a is on, then n_b must also be on

- Thus, all NWs can be individually addressed if for no two NWs n_a and n_b is $J_a \subseteq J_b$

Restating Controllability Conditions

- When are all *2w* NWs controllable?
  - $J_a \subseteq J_b$ cannot hold if there are top and bottom ends of LRs between every pair of NWs.
  - For any NW, all NWs above it can be turned off (see dark LRs). Same for NWs below given NW.



Coupon Collector Problem

- C coupon types
- Each box equally likely to contain any type of coupon
- How many boxes should be purchased to collect all C coupons with probability at least $1-\varepsilon$?

Equivalence to Coupon Collector Problem

- LR top (bottom) endpoint coupons:
  - A coupon corresponds to the space between a pair of consecutive NW.
  - There are 2w-1 top (bottom) endpoint coupons
- Failure coupon:
  - Failure corresponds to an ambiguous NW, which occurs when an LR endpoint lands on a NW
  - p_f is probability of failure

LR Displacement Model

- Simple model for the randomized mask-based decoder:
  - The LRs are equally likely to fall anywhere.
  - $p_f \approx 0.5$ and $p_s = 1 - p_f \approx 0.5$
  - Probability that *i*th coupon collected $p_i = (1-p_f)/C$
  - $C = 2w-1$, the number of consecutive NW pairs

Coupon Collector Problem with Failures

Theorem Let T = number trials to ensure all C coupons collected with probability = 1-ε when trial fails with prob $1-p_s$ and *i*th coupon collected with prob $p_i = p_s/C$. T satisfies

$$\frac{C}{p_s(1+p_s/C)} \ln\left(\frac{C}{\epsilon(1+\epsilon)}\right) \leq T \leq \frac{C}{p_s} \ln\left(\frac{C}{\epsilon}\right)$$

- This result bounds # MWs in linear decoder.

Performance of Mask-Based Decoder

- $2 \log_2 (N/(2w))$ MWs in logarithmic decoder
- Approx $\frac{C}{p_s} \ln \left(\frac{C}{\epsilon}\right)$ MWs in linear decoder where $C = 2w-1$.
- Mask-based decoder uses $M = 2 \log_2 (N/(2w)) + ((2w-1)/p_s) \log_2 (2w-1)/\varepsilon$ MWs.
- When $p_s = .5$, $w = 10$, $\varepsilon = .01$, $\mathbf{M = 2\log_2 N + 320}$ *MWs* needed to control *N* NW!
  - This improves to $M = 156$ if we assume that standard decoder used for contact groups and we tighten bounds. (See **Analysis of Mask-Based Nanowire Decoders** 1, Eric Rachlin, John E. Savage, to appear IEEE Transactions on Computers, 2007.)