Nanowire-Based Programmable Architectures

ANDR´E DEHON

ACM Journal on Emerging Technologies in Computing Systems, Vol. 1, No. 2, July 2005, Pages 109–162

INTRODUCTION

- **Goal :** to develop nanowire-based architectures which can bridge between lithographic and atomic-scale feature sizes and tolerate defective and stochastic assembly of regular arrays
- Using 10nm pitch nanowires, these nanowire-based programmable architectures offer one to two orders of magnitude greater mapped-logic density than defect-free lithographic FPGAs at 22nm

INTRODUCTION

- Do we have an adequate set of capabilities to build logic?
- How do we cope with the regularity demanded by bottom-up assembly?
- How do we accommodate the high defect rates and statistical assembly which accompany bottom-up assembly techniques?
- How do we organize and interconnect these atomic-scale building blocks?

INTRODUCTION

- How do we address nanowires from the lithographic scale for testing, configuration, and IO?
- How do we get logic restoration and inversion?
- What net benefit do these building blocks offer us?

TECHNOLOGY

- Nanowires
  - Atomic-scale nanowires can be engineered to have a variety of conduction properties from insulating to semiconducting to metallic
  - *Growth.* Semiconducting nanowires (NWs) can be grown to controlled dimensions on the nanometer scale using seed catalysts (e.g., gold balls) to define their diameter
  - NWs with diameters down to 3nm have been demonstrated

TECHNOLOGY

- Figure demonstrates growth of Si NWs

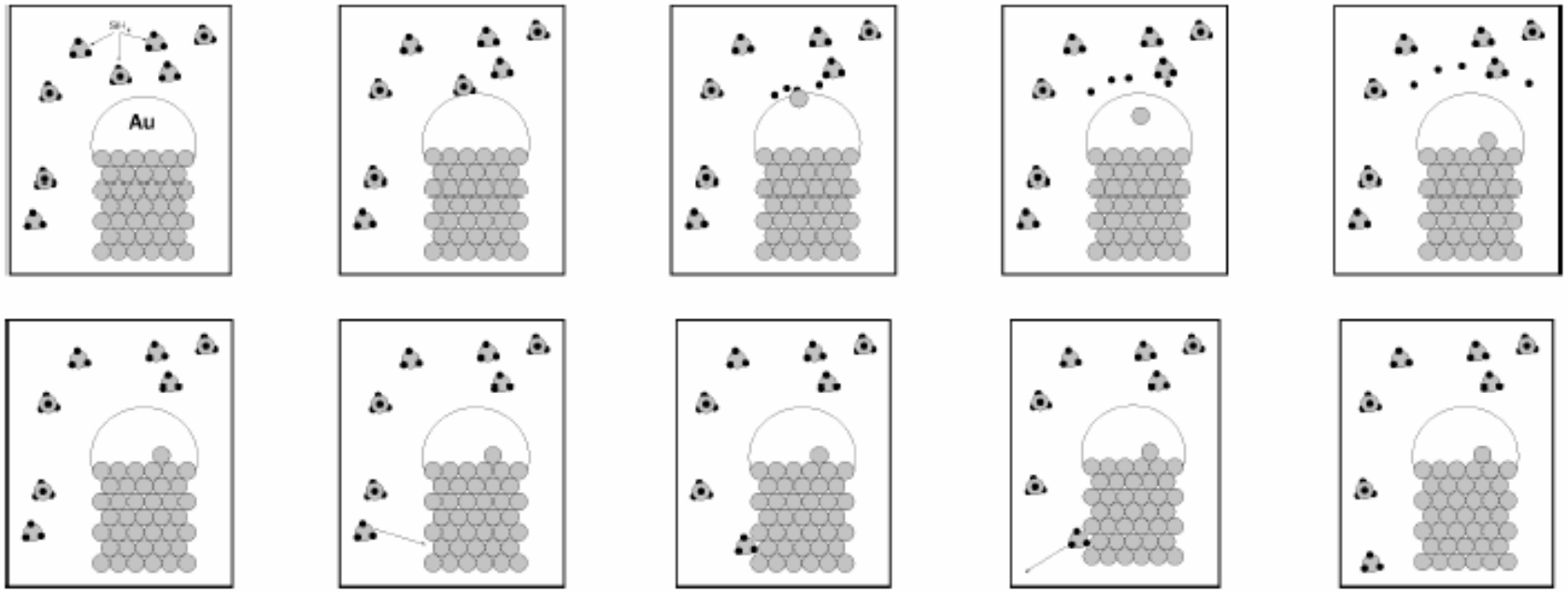


Fig. 1. SiNW growth cartoon. Top row shows how a gold (Au) catalyst allows the SiH_4 molecules to shed their Hydrogen termination so the Silicon can assemble onto the top of the growing NW. Bottom row shows how the Hydrogen termination on the SiH_4 molecule prevents it from assembling onto the edge of the growing NW where the catalyst is not present.

TECHNOLOGY

- *Field-Effect Control.* By controlling the mix of elements in the environment during growth, semiconducting NWs can be doped to control their electrical properties
- Heavily doped NWs are conducting. Conduction through lightly doped NWs can be controlled via an electrical field like Field-Effect Transistors
- Off resistances ($Roff_{fet}$) can be over 10Gs and on resistances (Ron_{fet}) under 0.1M; off/on resistance ratios are at least 10^4

TECHNOLOGY

- *Axial Profile*. The doping profile or material composition along the length of a NW can be controlled, This allows us to construct wires which are gateable in some regions but not gateable in others

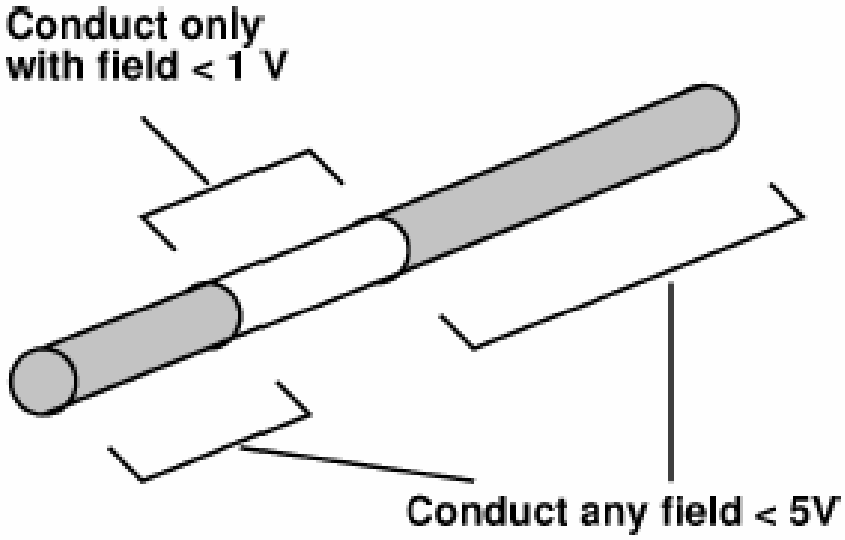


Fig. 2. Axial doping profile places selective gateable regions in a NW.

TECHNOLOGY

- *Radial Profile*. environmental conditions are changed to allow atomic layers to grow over the entire surface of the NW
- This allows us to sheath NWs in insulators (e.g., SiO2) to control spacing between conductors and between gated wires and control wires
- After a NW has been grown, it can be converted into a metal silicide

Fig. 3. Radial doping profile.

TECHNOLOGY

- Assembly
  - Langmuir-Blodgett (LB) flow techniques can be used to align a set of NWs into a single orientation, close pack them, and transfer them onto a surface

Fig. 4. Langmuir Blodgett alignment of NWs.

TECHNOLOGY

- The LB step can be rotated and repeated so that we get multiple layers of NWs such as crossed NWs for building a crossbar array or memory core

➢ Crosspoints

- Many technologies have been demonstrated for nonvolatile, switched crosspoints. Common features include:
  - resistance which changes significantly between on and off states

TECHNOLOGY

- the ability to be made rectifying;
- the ability to turn the device on or off by applying a voltage differential across the junction;
- the ability to be placed within the area of a crossed NW junction

- A typical, CMOS switch might be 2500λ2 [DeHon 1996], compared to a 5λ × 5λ bottom level metal wire crossing, making the crosspoint 100× the area of the wire crossing

CHALLENGES

- As we approach the atomic-scale, Precise location of atoms becomes relevant
- Variations occur due to statistical doping and dopant placement
- Perfect repeatability may be extremely difficult or infeasible for these feature sizes
- These bottom-up approaches, in contrast, promise us finer feature sizes that are controlled by physical phenomena but do not promise perfect, deterministic alignment in three dimensions

CHALLENGES

- This leads us to ask if we can reasonably give up our perfect correlation and complete design freedom in three dimensions in order to exploit smaller feature sizes

➢ Regular Assembly

- The assembly techniques suggest that we can build regular arrays at tight pitch with both NW trace width and trace spacing using controlled NW diameters

CHALLENGES

- we cannot deterministically differentiate features at this scale, that is, we cannot make one particular crosspoint be different in some way from the other crosspoints in the array

➢ Nanowire Lengths

- NWs can be grown to hundreds of microns
- However, at this high length to diameter ratio, they become highly susceptible to bending and ultimately breaking

CHALLENGES

- Consequently, we must limit ourselves to modest NW lengths (10s of microns) in order to yield a large fraction of the NWs in a given array

➢ Defective Wires and Crosspoints

- At this scale, we expect wires and crosspoints to be defective in the 1–10% range
  - NWs may break along their axis during assembly
  - NW to microwire junctions depend on a small number of atomic scale bounds which are statistical in nature and subject to variation in NW properties

CHALLENGES

- Junctions between crossed NWs will be composed of only 10s of atoms or molecules and individual bond formation is statistical in nature
- Statistical doping of NWs may lead to high variation among NWs

- we consider two main defect types:
  - *Wire Defects*
  - *Nonprogrammable Crosspoint Defects*
    - Based on the physical phenomena involved, we consider nonprogrammable junctions to be much more common than shorted junctions

BUILDING BLOCKS

- Crosspoint Arrays
  - assembly processes allow us to create tight-pitch arrays of crossed NWs with switchable diodes at the crosspoints
  - These arrays can serve as:
    - memory cores,
    - programmable, wired-OR planes,
    - programmable crossbar interconnect arrays

BUILDING BLOCKS

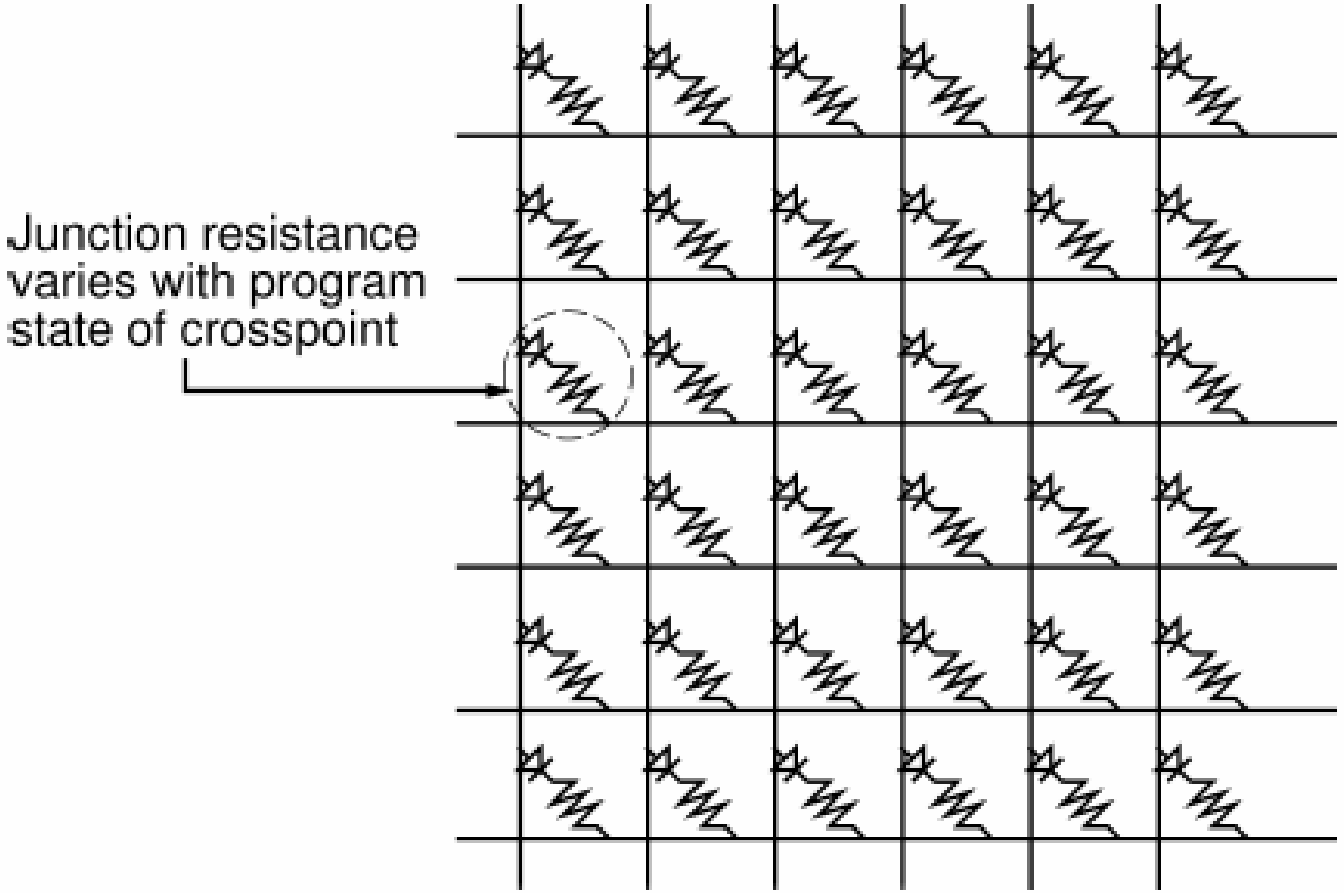


Fig. 5. Logical diode crossbar formed by crossed NWs.

BUILDING BLOCKS

- *Memory Core*
  - by applying a large voltage across a crosspoint junction, the crosspoint can be switched into a high or low resistance state
  - we can operate at a lower voltage without resetting the crosspoint. Consequently, we can read back a crosspoint's state by applying a small, test voltage
- *Programmable, Wired-*OR *Plane*
  - we can program OR logic into a crosspoint array
  - Each row output NW serves as a wired-OR for all of the inputs programmed into the low resistance state
  - outputs will need restoration

BUILDING BLOCKS

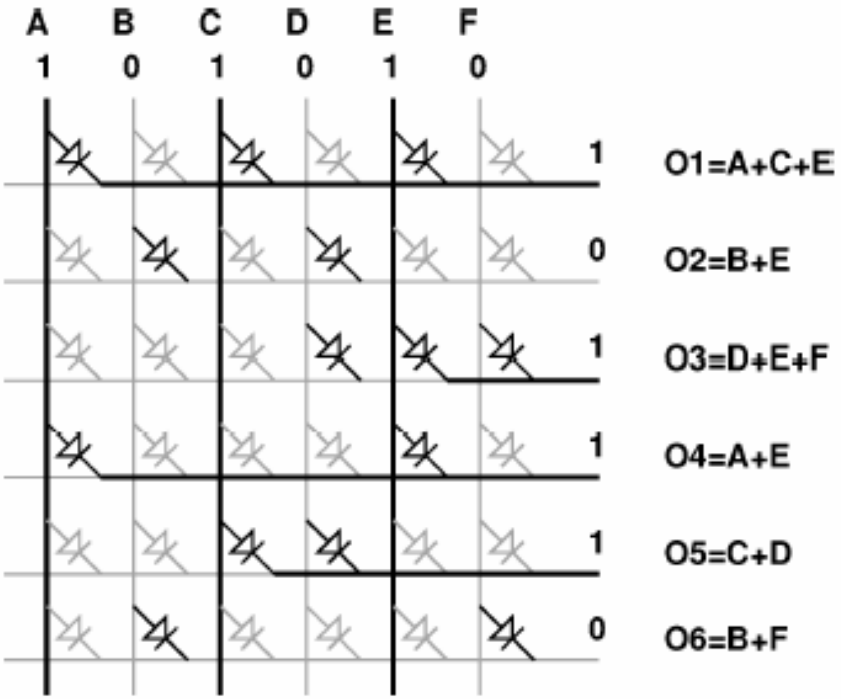


Fig. 6. Wired-OR plane operation. Programmed on crosspoints are shown in black; off crosspoints are shown in grey. Dark lines represent a NW pulled high, while light lines remain low. Output NWs are marked dark, starting at the diode that pulls them high, in order to illustrate current flow; the entire output NW would be pulled high in actual operation.

BUILDING BLOCKS

- *Programmable Crossbar Interconnect Arrays*
  - if we restrict ourselves to connecting a **single** row wire to each column wire, the crosspoint array can serve as a crossbar switch

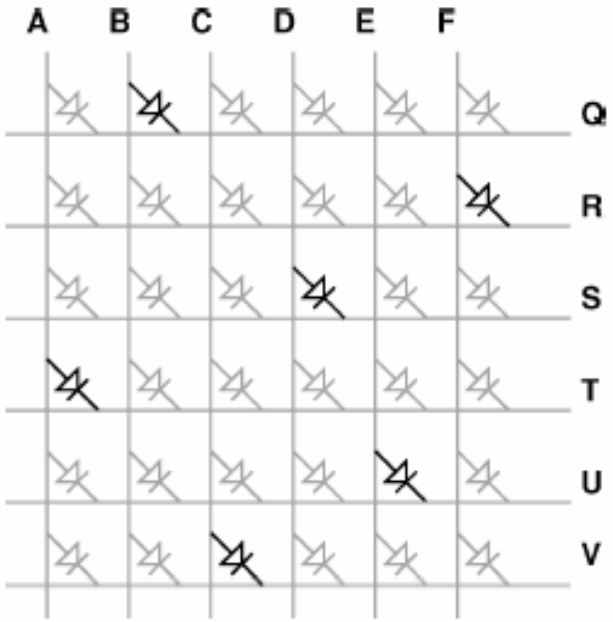


Fig. 7. Example crossbar routing configuration. Programmed on crosspoints are shown in black; off crosspoints are shown in grey. Here we show the crossbar programmed to connect A→T, B→Q, C→V, D→S, E→U, and F→R.

BUILDING BLOCKS

- Decoders
  - A key challenge is bridging the length scale between the lithographic-scale wires and the small diameter NWs
  - By building a decoder between the coarse-pitch lithographic wires and the tight-pitch NWs, we can bridge this length scale and address a single NW
- *NW Coding*
  - One way to build such a decoder is to place an address on each NW using the axial doping or material composition profile

BUILDING BLOCKS

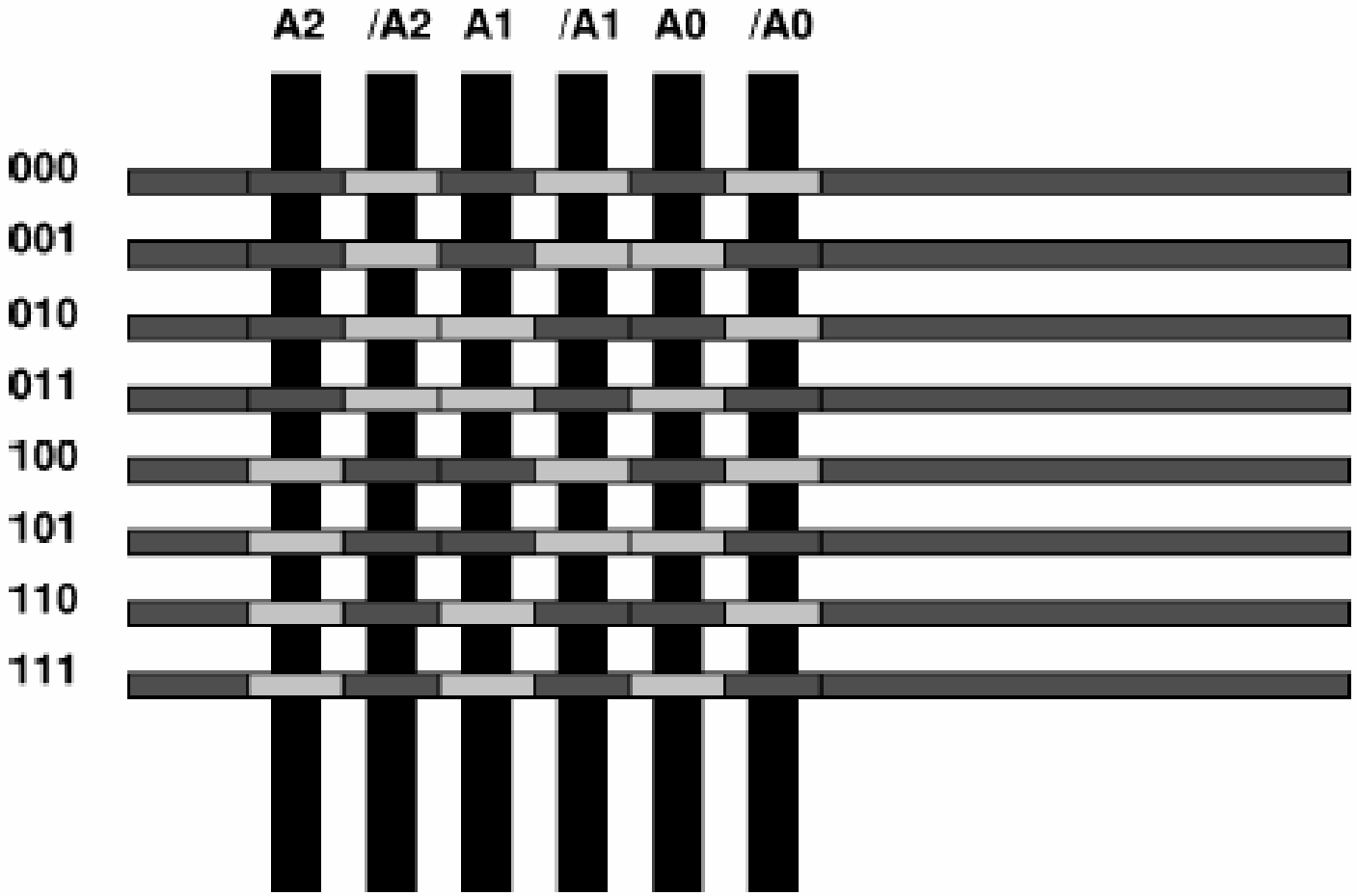
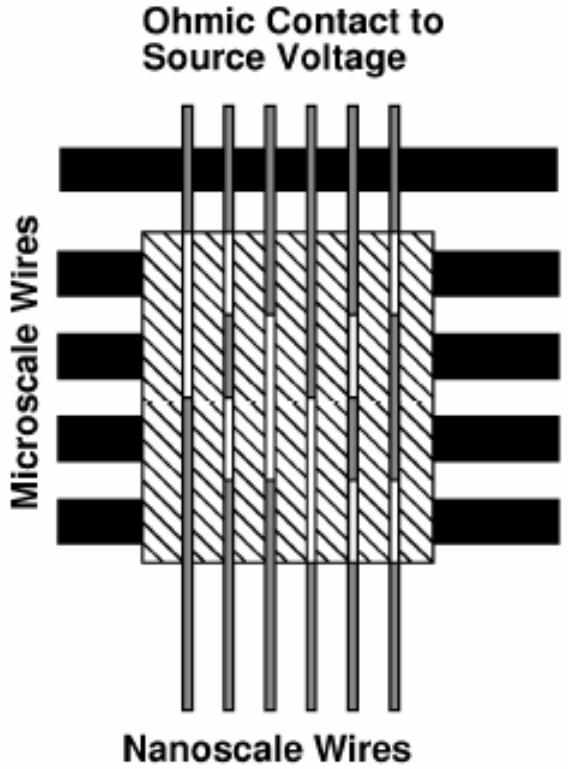


Fig. 8. Dual-rail address coding.

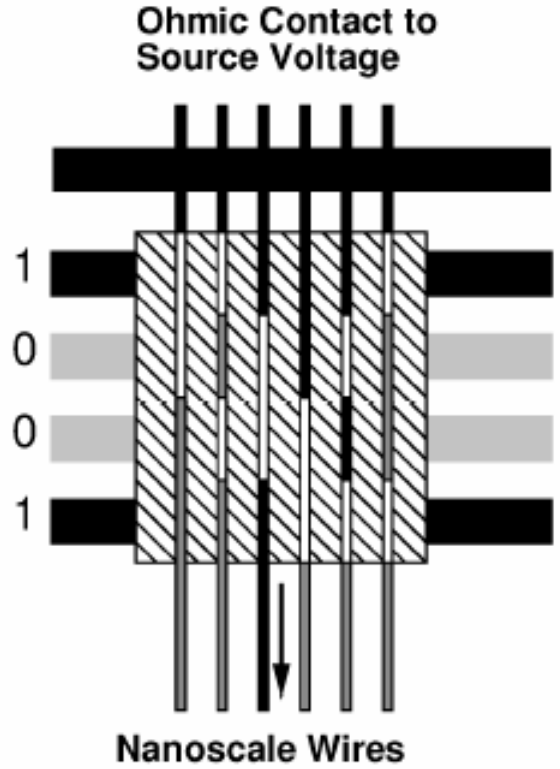
BUILDING BLOCKS

- *Decoder Assembly*
  - We can only arrange to create a tight-pitch parallel ensemble of a collection of NWs
  - We can statistically guarantee with arbitrarily high probability that every NW in an array has a unique address
- *Decoder and Multiplexer Operation*
  - There is no directionality to the decoder. Consequently, this same unit can serve equally well as a multiplexer
  - When we apply an address to the lithographic-scale wires, it allows conduction through the addressing region for only one of the NWs

BUILDING BLOCKS



(a) Decoder Configuration: white NW regions are coded and controllable, while grey regions are not controllable and act as wires.



(b) Dark represents lines driven high; light grey shows lines low or undriven. Only the coding on the third line matches the applied address (1001) and allows conduction. All other cases have a high address voltage crossing a lightly doped region which prevents conduction.

Fig. 9. Coded NW decoder.

BUILDING BLOCKS

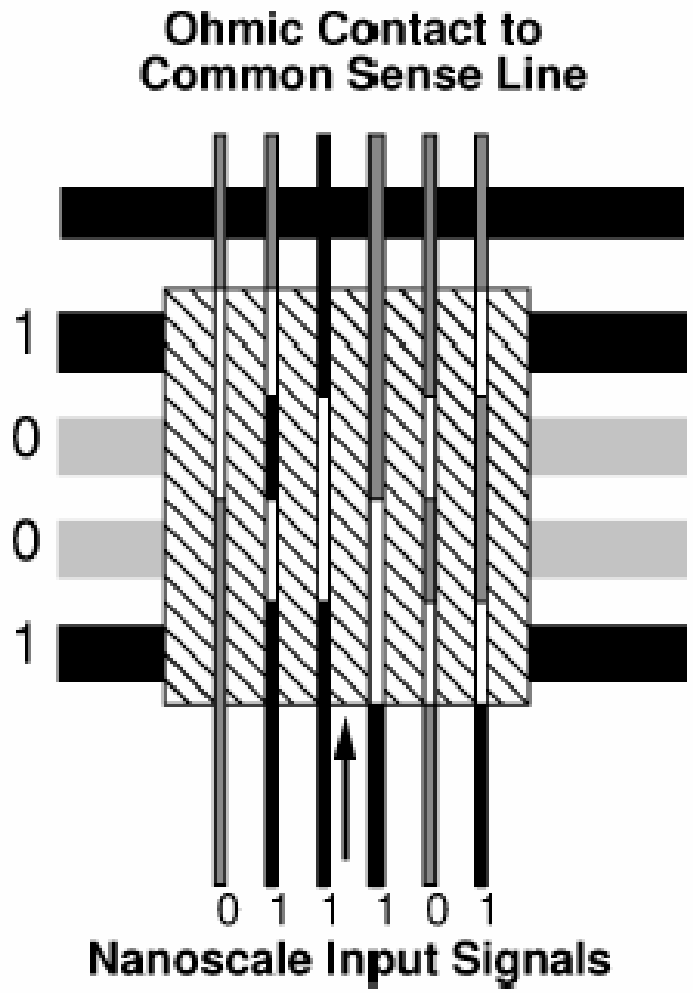


Fig. 10. Coded NW multiplexer operation.

BUILDING BLOCKS

- Restoration and Inversion
  - The programmable, wired-OR logic is passive and nonrestoring, drawing current from the input
  - NWs can be field-effect controlled. This gives us the potential to build FET-like gates for restoration
- NW Inverter and Buffer
  - we can potentially use the field from one NW to control the other NW
  - Figure shows an inverter built using this basic idea

BUILDING BLOCKS

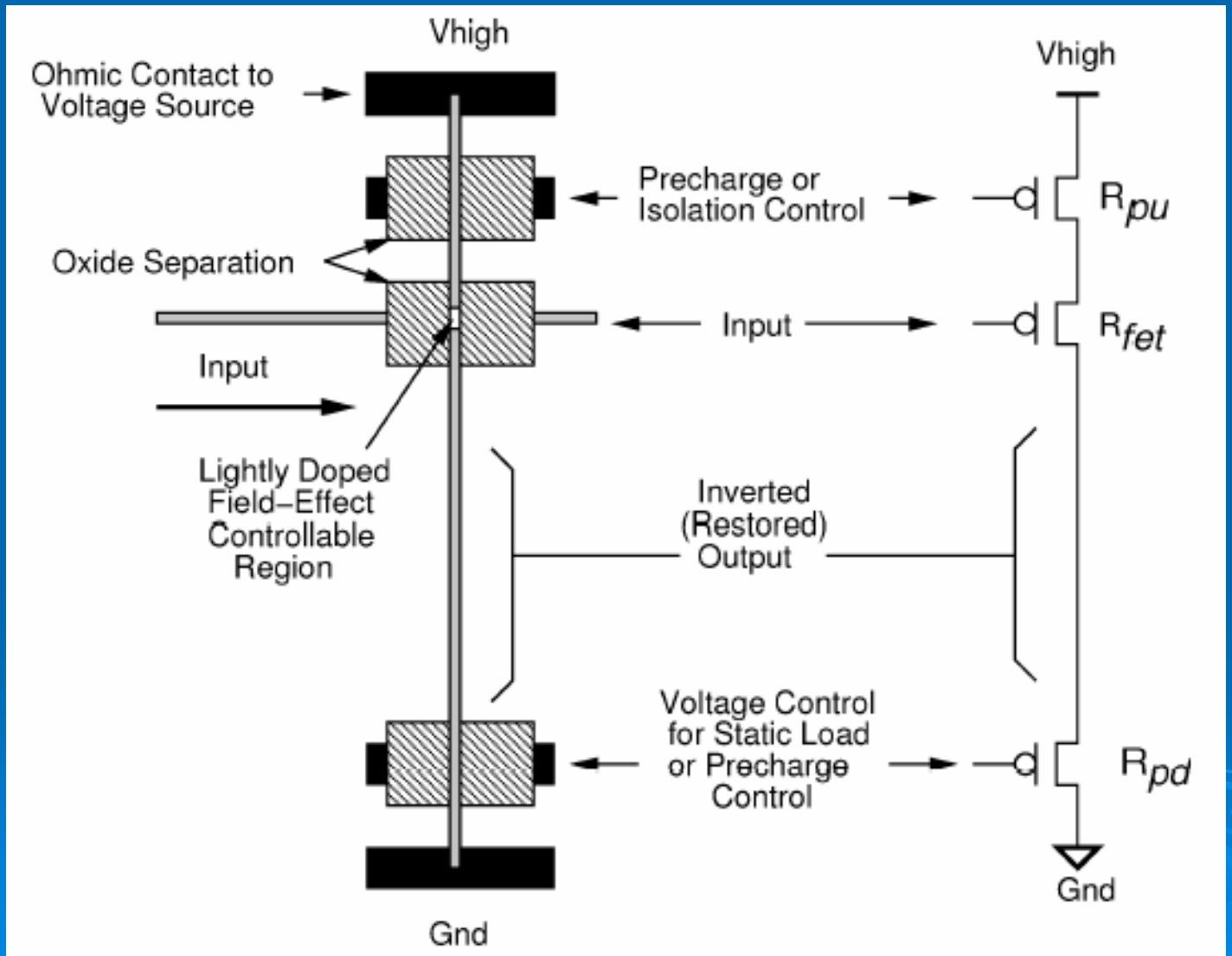


Fig. 11. NW inverter.

BUILDING BLOCKS

- This same arrangement can be used to buffer rather than invert the input
- NW field-effect gating has sufficient nonlinearity so that this gate provides gain to restore logic signal levels

- Ideal Restoration Array
  - we need to restore a set of tight-pitch NWs such as the outputs of a programmable, wired-OR array
  - A restoration array is shown in Figure (a)
  - The only problem here is that we do not have a way to align and place axially doped NWs so that they provide exactly this pattern

BUILDING BLOCKS

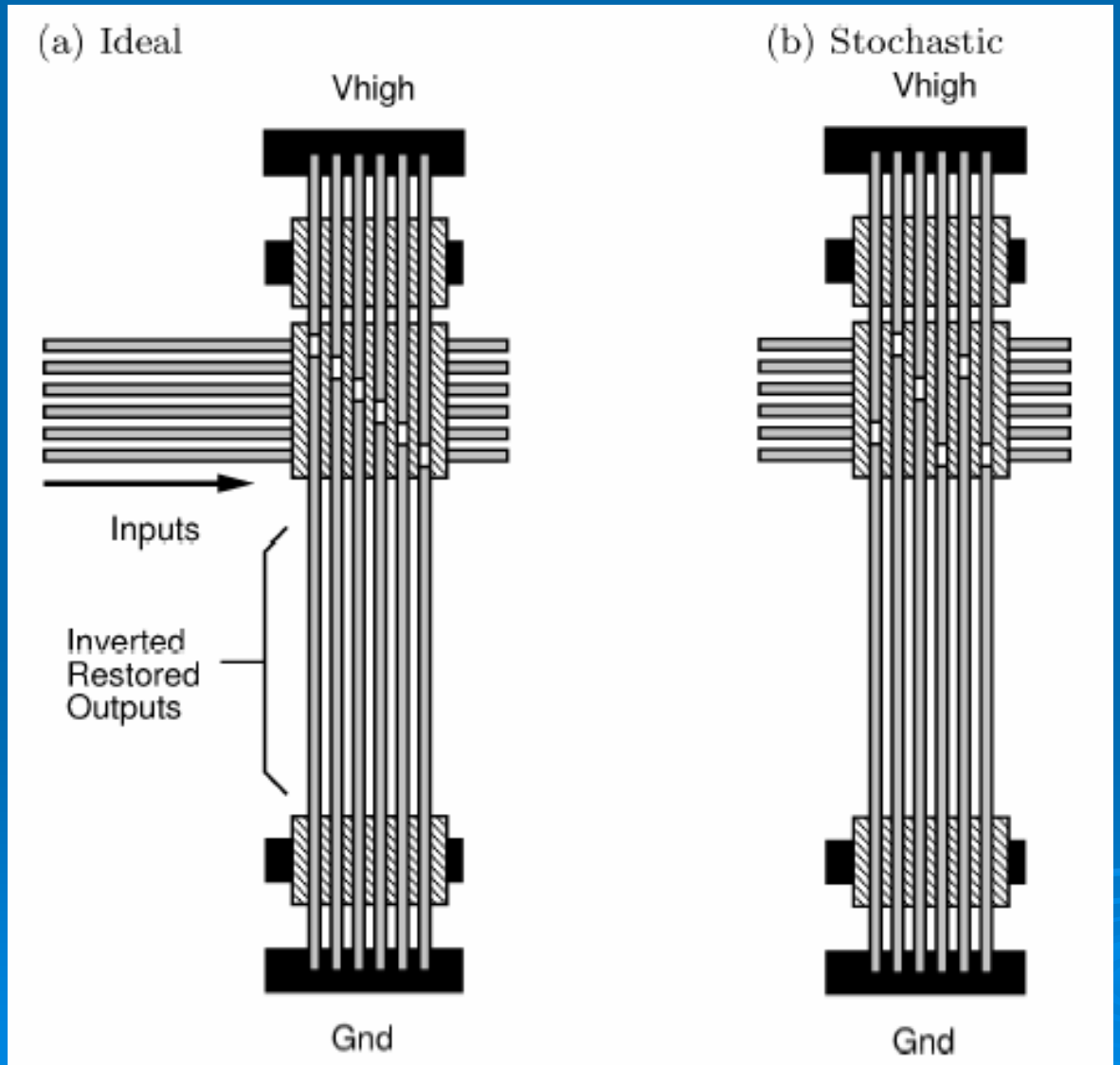


Fig. 12. Restoration array.

BUILDING BLOCKS

- MEMORY ARRAY
  - Combining the crosspoint memory cores with a pair of decoders, we can build a tight-pitch, NW-based memory array
  - Figure shows how these elements come together in a small memory array
  - Write operations can be performed by driving the appropriate write voltages onto a single row and column line
  - Read operations occur by driving a reference voltage onto the common column line, setting the row and column addresses, and sensing the voltage on the common row read line

MEMORY ARRAY

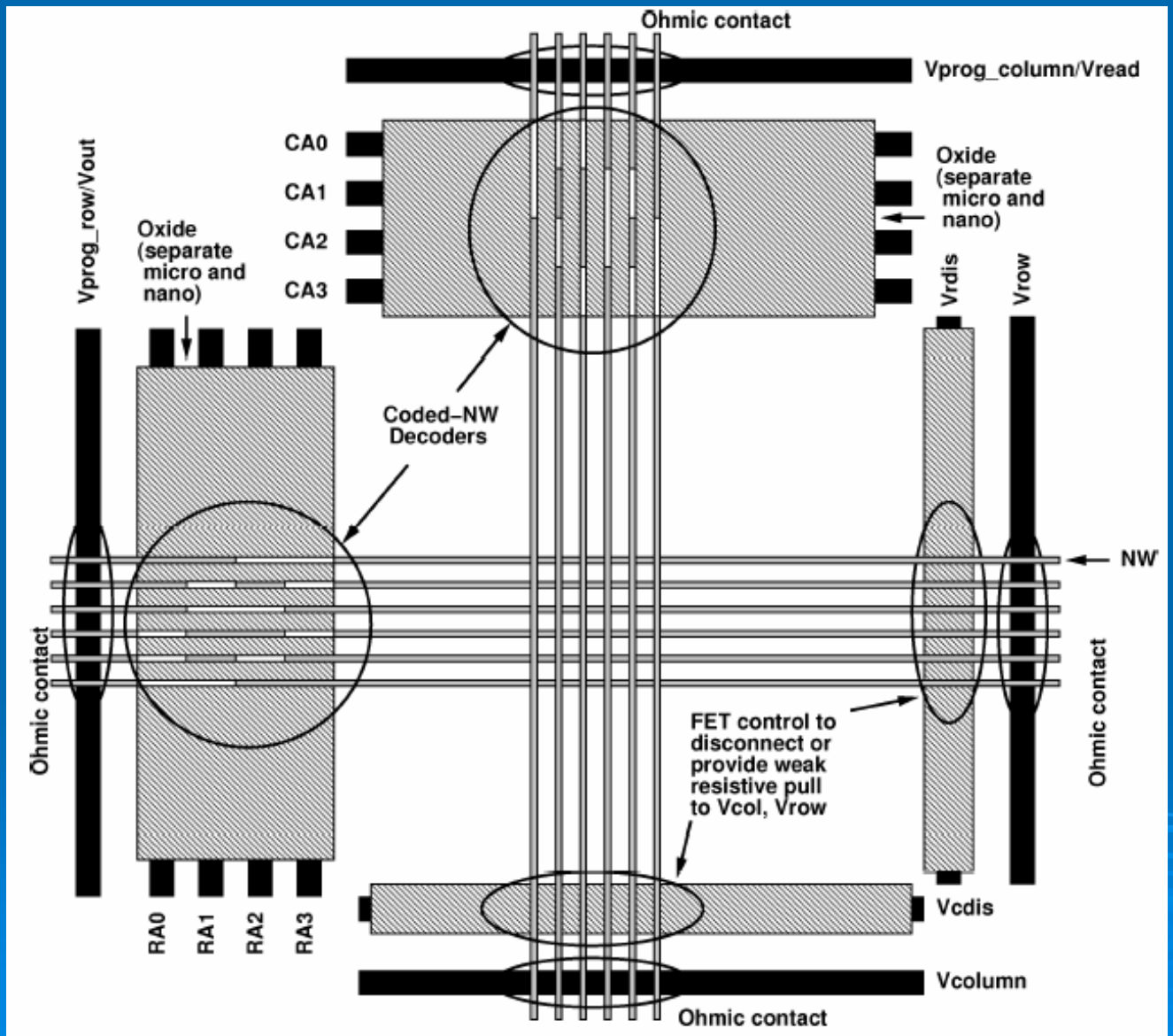


Fig. 14. Memory array built from coded NW decoder and crosspoint memory core.

MEMORY ARRAY

- Limitations on reliable NW length and the capacitance and resistance of long NWs prevent us from building arbitrarily large memory arrays
- we break up large NW memories into banks similar to the banking used in conventional DRAMs
- After accounting for defects, ECC overhead, and lithographic control overhead, net densities on the order of 10^{11} bits/cm^2 seem achievable, using NW pitches around 10nm

MEMORY ARRAY

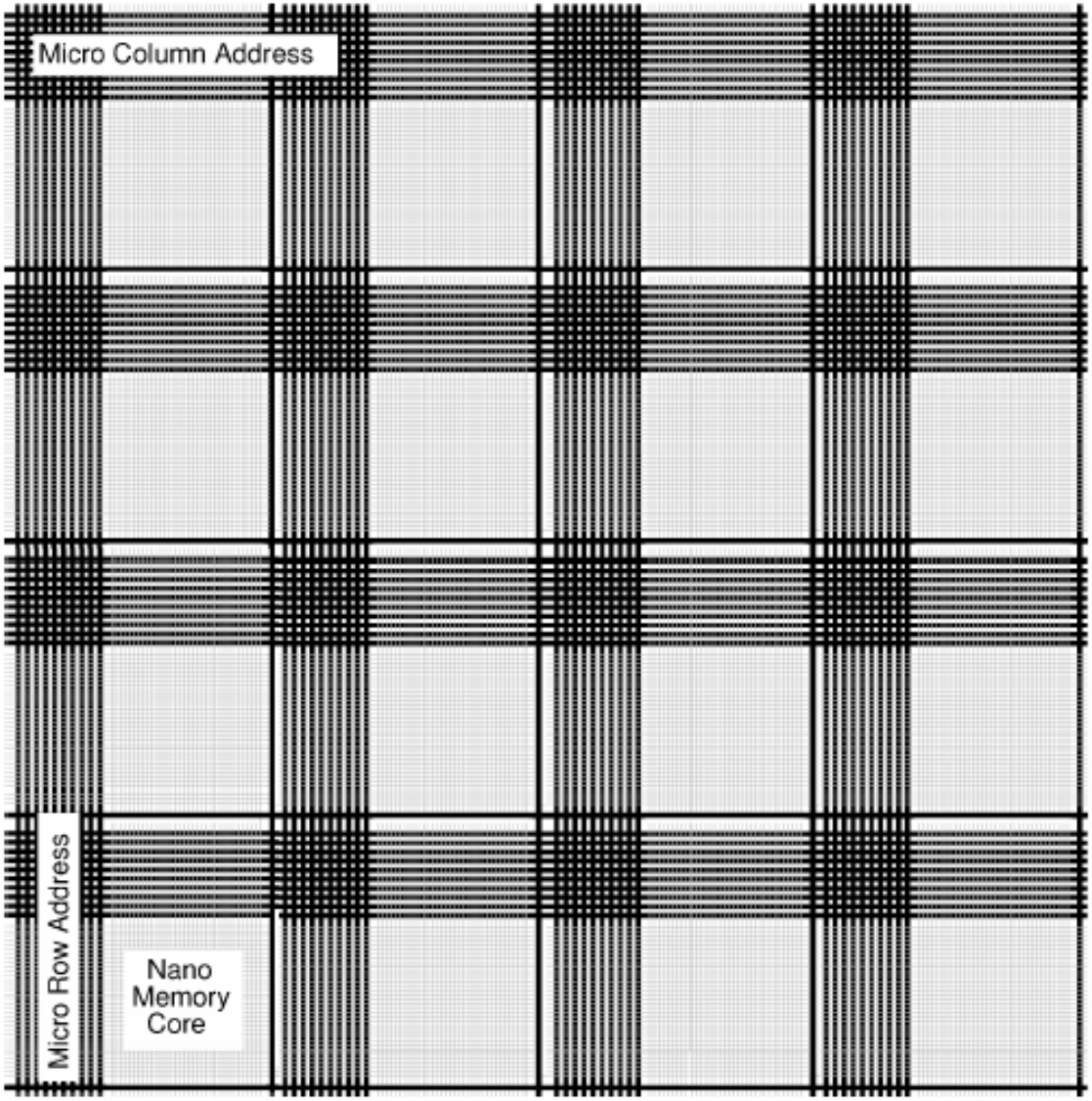


Fig. 15. Tile of NW-based memory banks to construct large-scale memory.

LOGIC ARCHITECTURE

- Figure shows a simple Programmable Logic Array (PLA) built using the building blocks

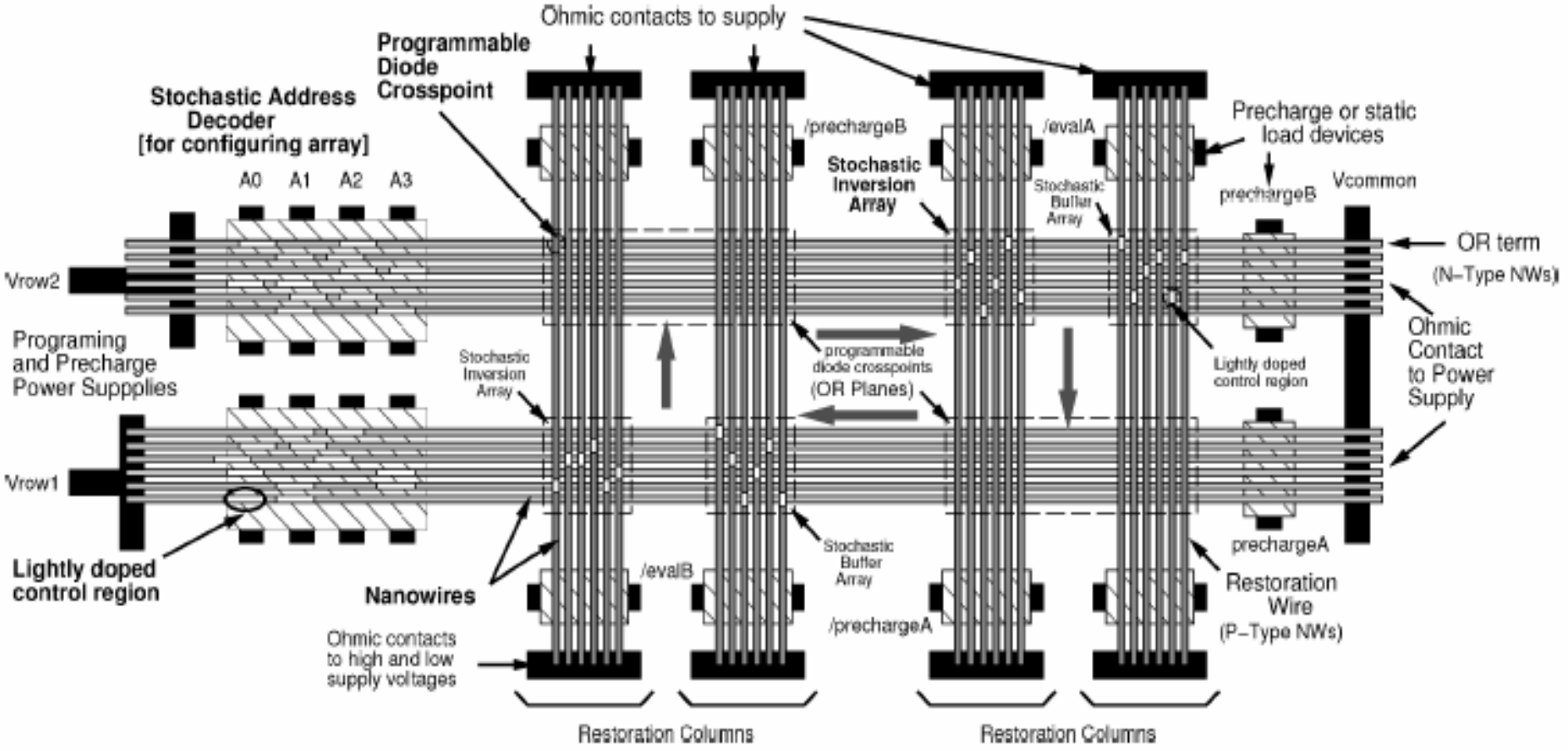


Fig. 16. Simple nanoPLA block.

LOGIC ARCHITECTURE

- Two interconnected logic planes, Each plane is composed of a programmable wired-OR array, followed by a restoration array
- Two restoration arrays: one providing the inverted sense of the OR-term logic and one providing the non-inverted buffered sense
- The entire construction is a set of crossed NWs as allowed by assembly constraints
- The logic gates in each plane are composed of a diode-programmable wired-OR NW, followed by a field-effect buffer or inverter NW

LOGIC ARCHITECTURE

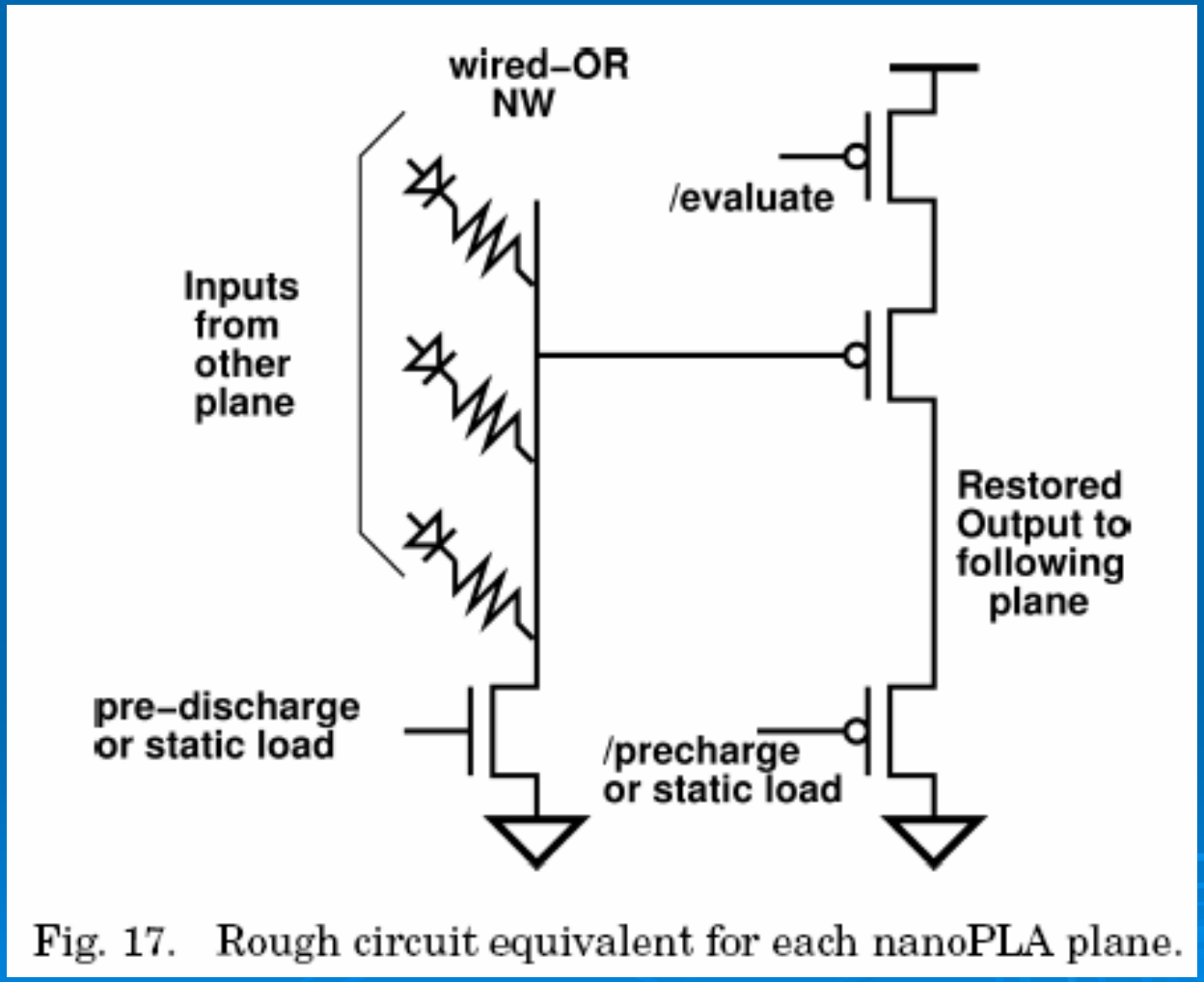


Fig. 17. Rough circuit equivalent for each nanoPLA plane.

LOGIC ARCHITECTURE

- *Basic Clocking*
  - The basic nanoPLA is simply two restoring logic stages back-to-back
  - If we turn off all three of the control transistors in restoring stages, there is no current path from the input to the diode output stage
  - The output stage is capacitively loaded, so it will hold its value
  - This is the same strategy as two-phase clocking in conventional VLSI

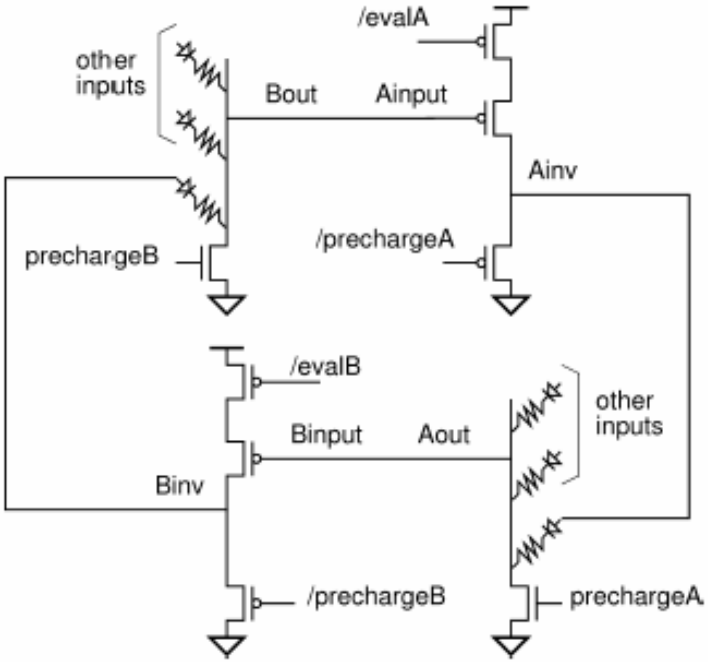


Fig. 18. Precharge clocked INV-OR-INV-OR (NOR-NOR, AND-OR) cycle.

LOGIC ARCHITECTURE

- Interconnect
  - We know from VLSI that large PLAs do not always allow us to exploit the structure which exists in logic
  - limitation on NW length bounds the size of the PLAs we can reasonably build
  - to scale up to large capacity logic devices, we must interconnect modest size nanoPLA blocks
  - The key idea for interconnecting nanoPLA blocks is to overlap the restored output NWs from each such block with the wired-OR input region of adjacent nanoPLA blocks

LOGIC ARCHITECTURE

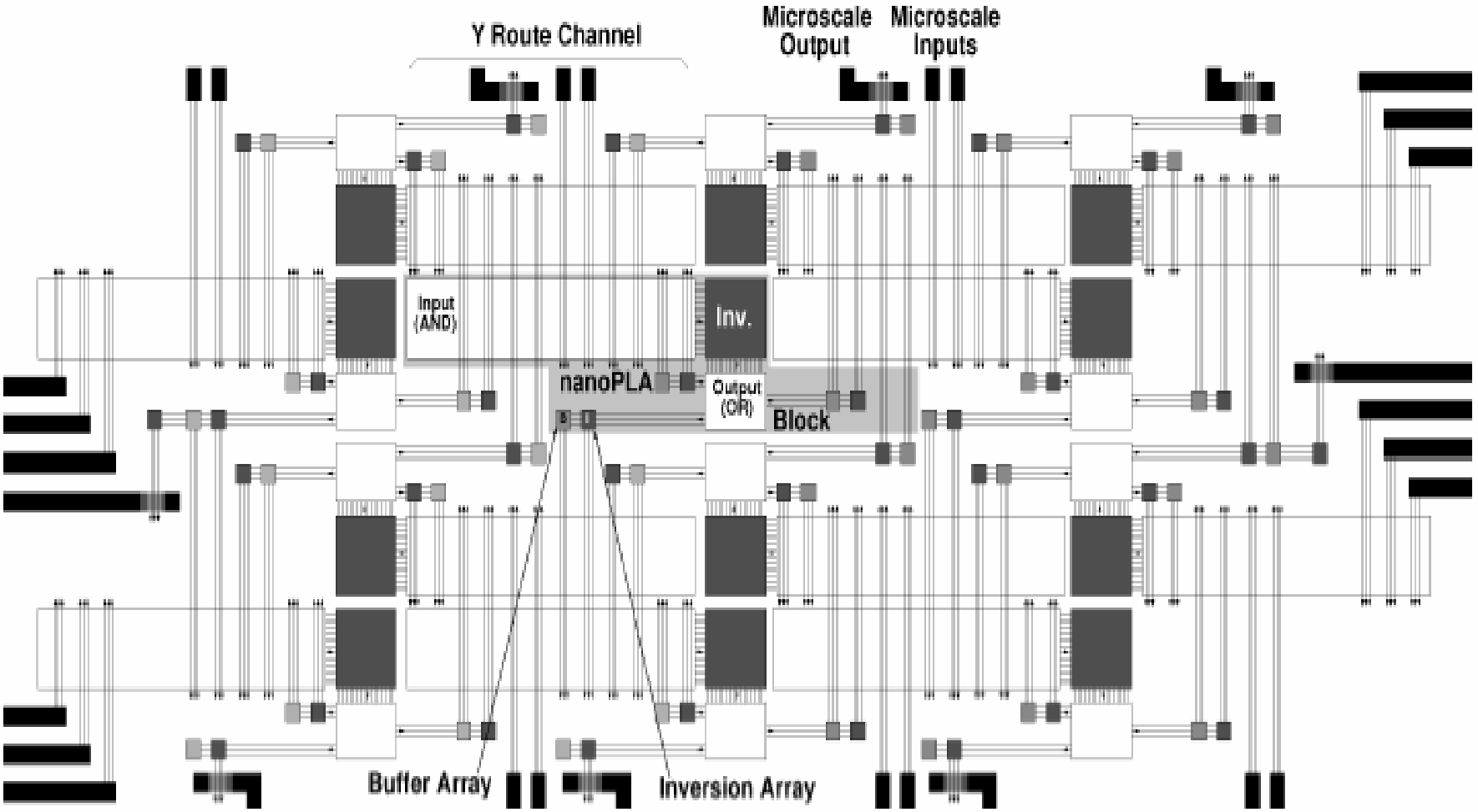


Fig. 20. nanoPLA block tiling with edge IO to lithographic scale.

LOGIC ARCHITECTURE

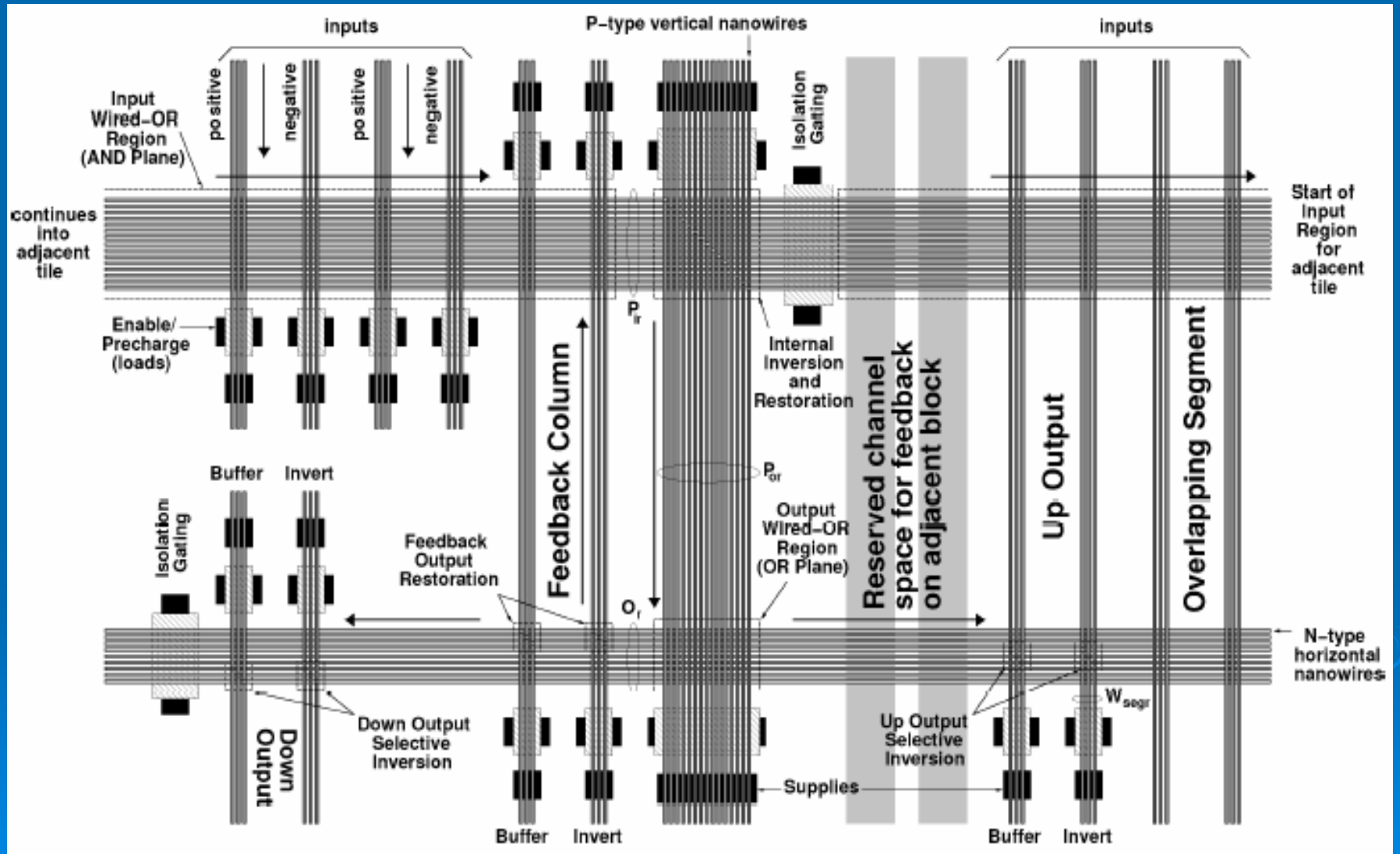


Fig. 21. nanoPLA block tile.

LOGIC ARCHITECTURE

- CMOS IO
  - NanoPLAs will be built on top of a lithographic substrate
  - Lithographic circuitry and wiring provides a way to probe the NWs, to map their defects and to configure the logic
  - we can provide IO blocks to connect the nanoscale logic to lithographic-scale

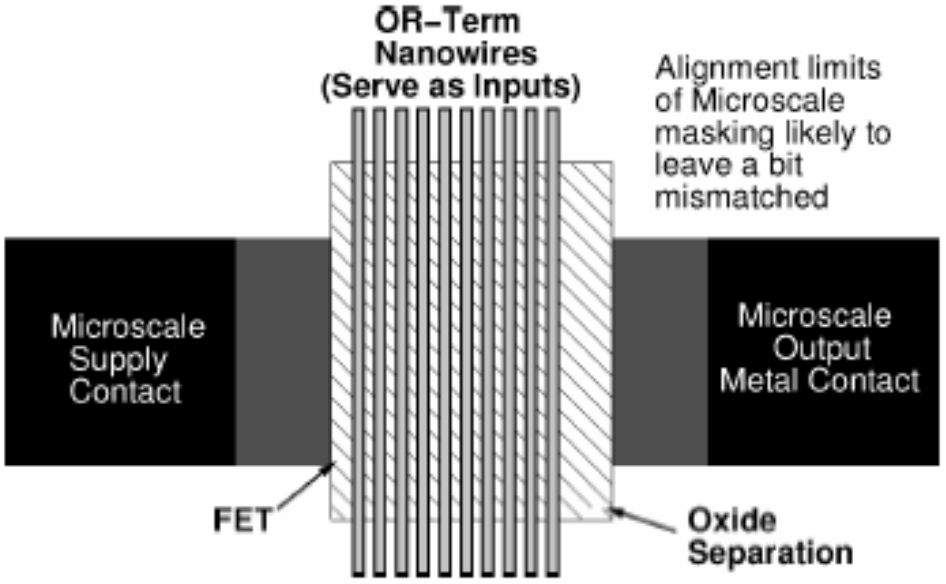


Fig. 23. Nanoscale to lithographic-scale FET output structure.

DEFECT TOLERANCE

- Small percentage of wires are defective and crosspoints are nonprogrammable
- Due to stochastic assembly and misalignment a percentage of NWs are unusable
- We can provide spare NWs in an array, test NWs for usability, and configure the array using only the non-defective NWs

➢ NW Sparing

- The probability of having exactly *i* restored OR-terms is:

$$P_{yield}(N,i) = \left(\binom{N}{i} (P_{or})^{i} (1 - P_{or})^{N-i} \right)$$

DEFECT TOLERANCE

- Probability of having at least M non-defective wires out of N:

$$P_{M\,of\,N} = \sum_{M \le i \le N} \left(\binom{N}{i} (P_{or})^i \left(1 - P_{or}\right)^{N-i} \right)$$

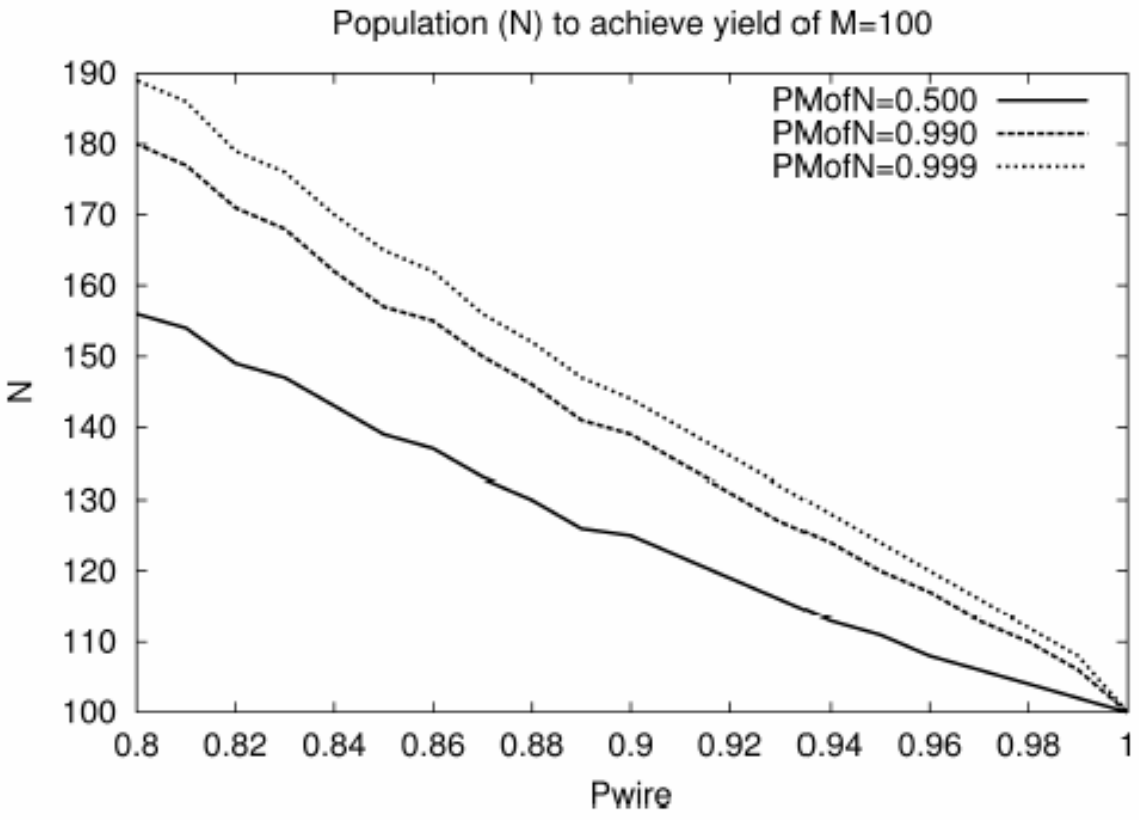


Fig. 26. Physical population (N) of wires to achieve 100 restored OR-terms (M).

BOOTSTRAP TESTING

- Discovery
  - we will need to discover the live addresses and their restoration polarity
  - We must identify which NWs are usable and which are not
- Programming
  - To program any diode crosspoint, we drive one address into the top address decoder and the second into the bottom
  - We effectively place the desired programming voltage differential across a single crosspoint

BOOTSTRAP TESTING

- *Programming Diode Crosspoints*
  - Knowing which polarities are available from each of the present addresses, we can program up the intended function

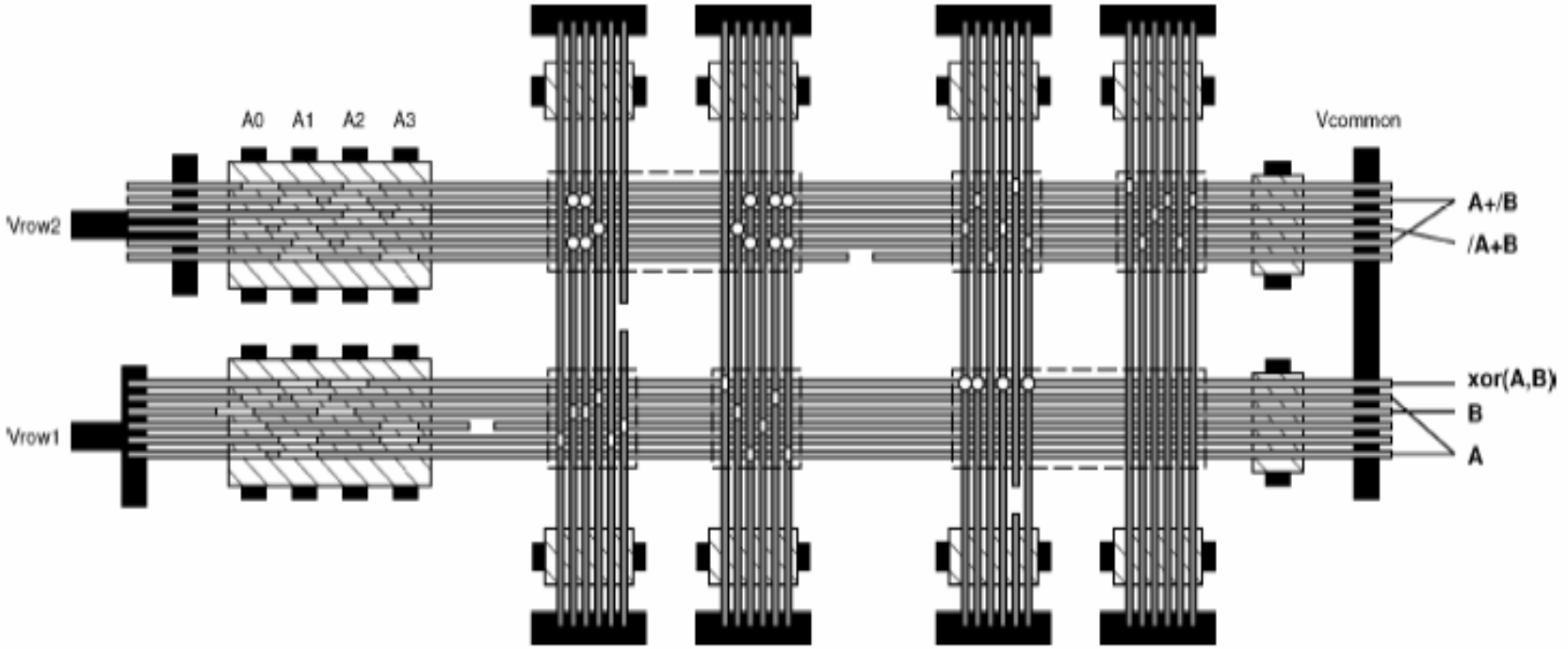
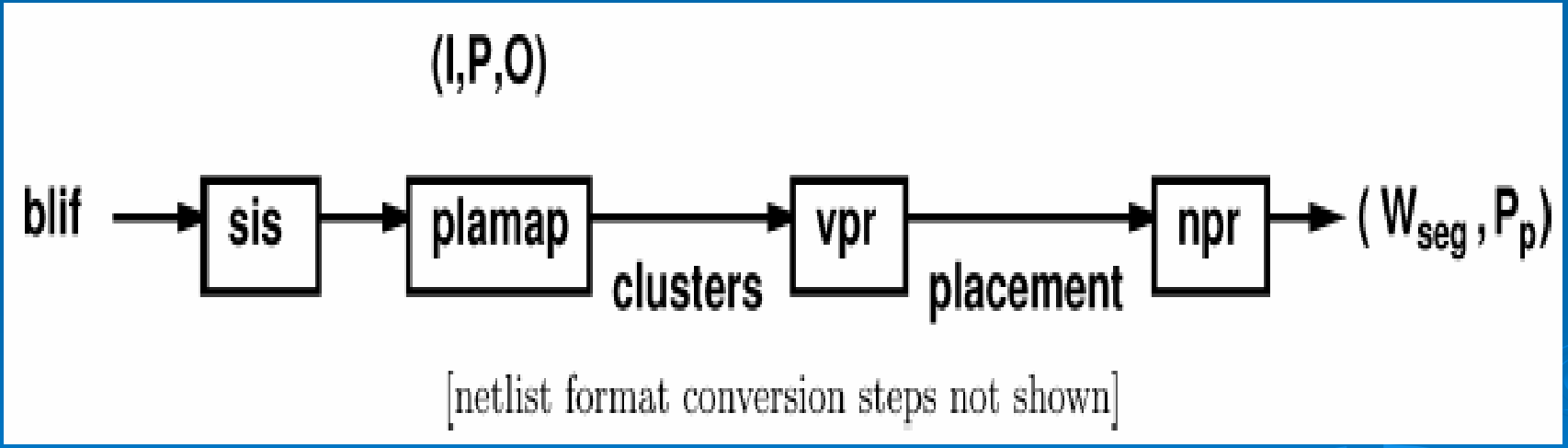


Fig. 31. Assign OR-terms for Computing XOR2.

CAD MAPPING

- To map from standard logic netlists (e.g., BLIF) to the nanoPLA arrays, a combination of conventional and custom tools as shown in the figure are used



CAD MAPPING

- SIS performs standard technology independent optimizations and decomposes the logic into small fanin nodes for covering.
- PLAMAP covers the logic into (I,P,O) PLA clusters
- These clusters can then be placed with VPR
- We developed our own nanoPLA router (npr) for routing

COST MODELS

- Text presents estimations and calculations for area and delay of the proposed architecture
- It also gives formulas for NW capacitance and resistance
- Power and energy relations are also given

Key NanoPLA parameters

- *Wseg* is the number of NWs in each output group
- *Lseg* is the number of nanoPLA block heights up or down which each output
- Crosses
- *F* is the number of NWs in the feedback group
- *P* is the number of logical PTERMS in the input (AND) plane of the nanoPLA logic block
- *Op* is the number of total outputs in the OR plane
- *Pp* is the number of total PTERMS in the input (AND) plane
- Since these are also used for route-through connections, this is larger than the number of logical PTERMS in each logic block:

$$Pp \leq P + 2 \times Wseg + F$$

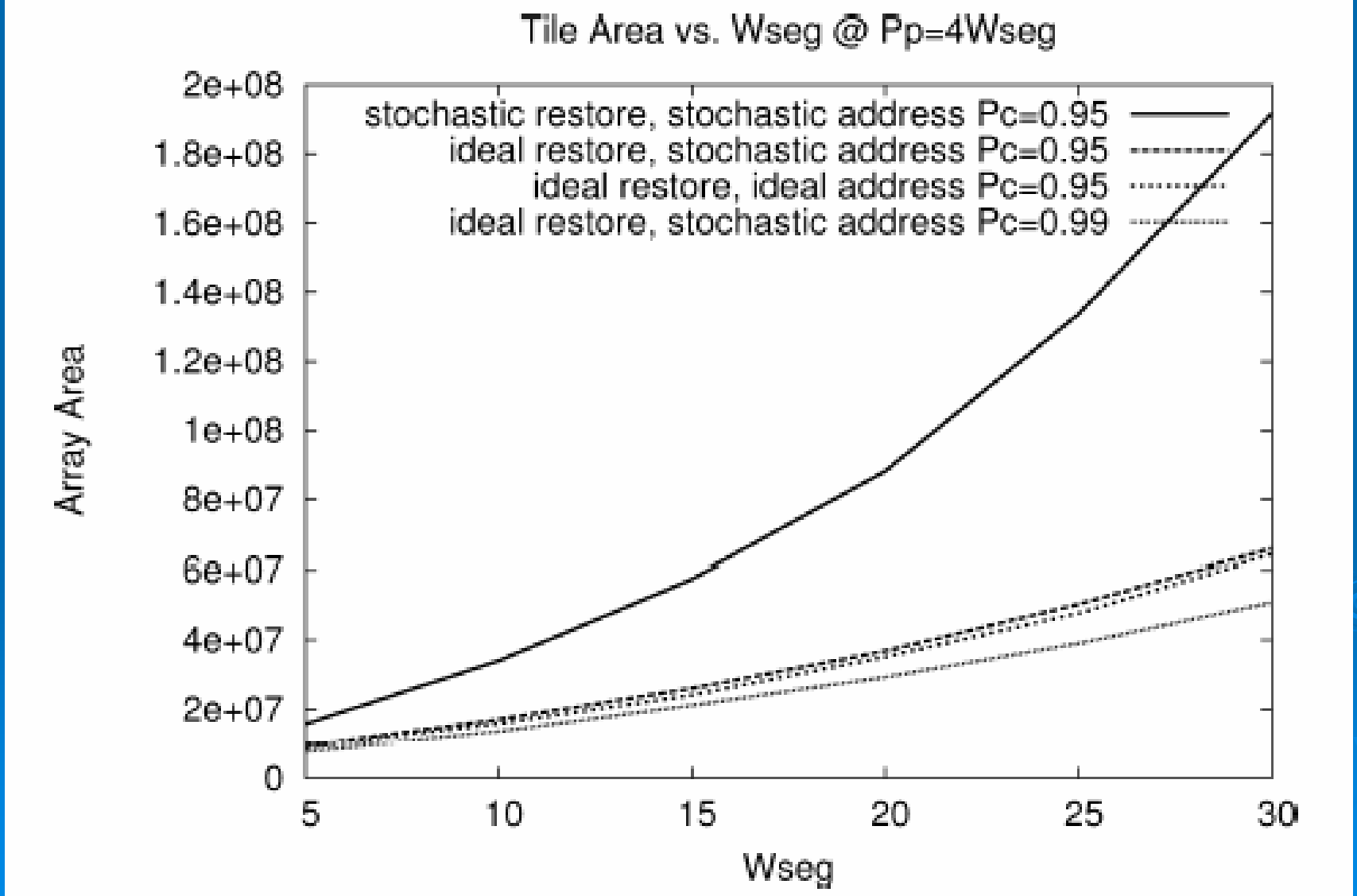
DESIGN SPACE EXPLORATION

Table III. Area Minimizing Design Points (Ideal Restoration, $W_{litho} = 105$nm, $W_{fnano} = 10$nm, $W_{dnano} = 10$nm)

Design	Parameters					nanoPLA			Electrical Length (μm)
	Map			Physical		array org.	Area $\times 10^8$nm^2	Area Ratio	
	I	P	O	P_p	W_{seg}				
alu4	18	44	2	60	8	5 × 5	4.5	339	11.2
apex2	20	24	8	54	15	14 × 14	47.2	39	12.5
apex4	12	48	2	62	7	6 × 6	6.0	208	10.8
bigkey	16	24	8	44	13	11 × 11	24.7	69	11.3
clma	20	48	8	104	28	23 × 23	278.6	30	19.3
des	18	28	8	78	25	12 × 12	59.6	26	16.8
diffeq	16	44	8	86	21	11 × 11	46.5	32	16.0
dsip	20	24	6	58	18	9 × 9	23.0	59	13.7
elliptic	18	24	8	78	27	17 × 17	130.1	27	17.7
ex1010	20	48	4	66	9	9 × 9	16.0	287	11.9
ex5p	12	32	8	67	18	3 × 3	2.7	389	14.2
frisc	18	24	8	92	34	18 × 18	198.5	17	21.2
misex3	18	48	4	64	8	7 × 7	9.1	153	11.5
pdc	16	48	8	74	13	7 × 7	12.6	363	12.8
s298	18	48	8	79	15	8 × 8	18.3	105	13.8
s38417	14	32	8	76	22	23 × 23	199.3	32	15.9
seq	20	36	8	72	18	9 × 9	25.2	69	14.4
spla	20	44	8	68	12	5 × 5	5.8	632	12.1
tseng	16	28	8	78	25	11 × 11	50.1	20	16.8

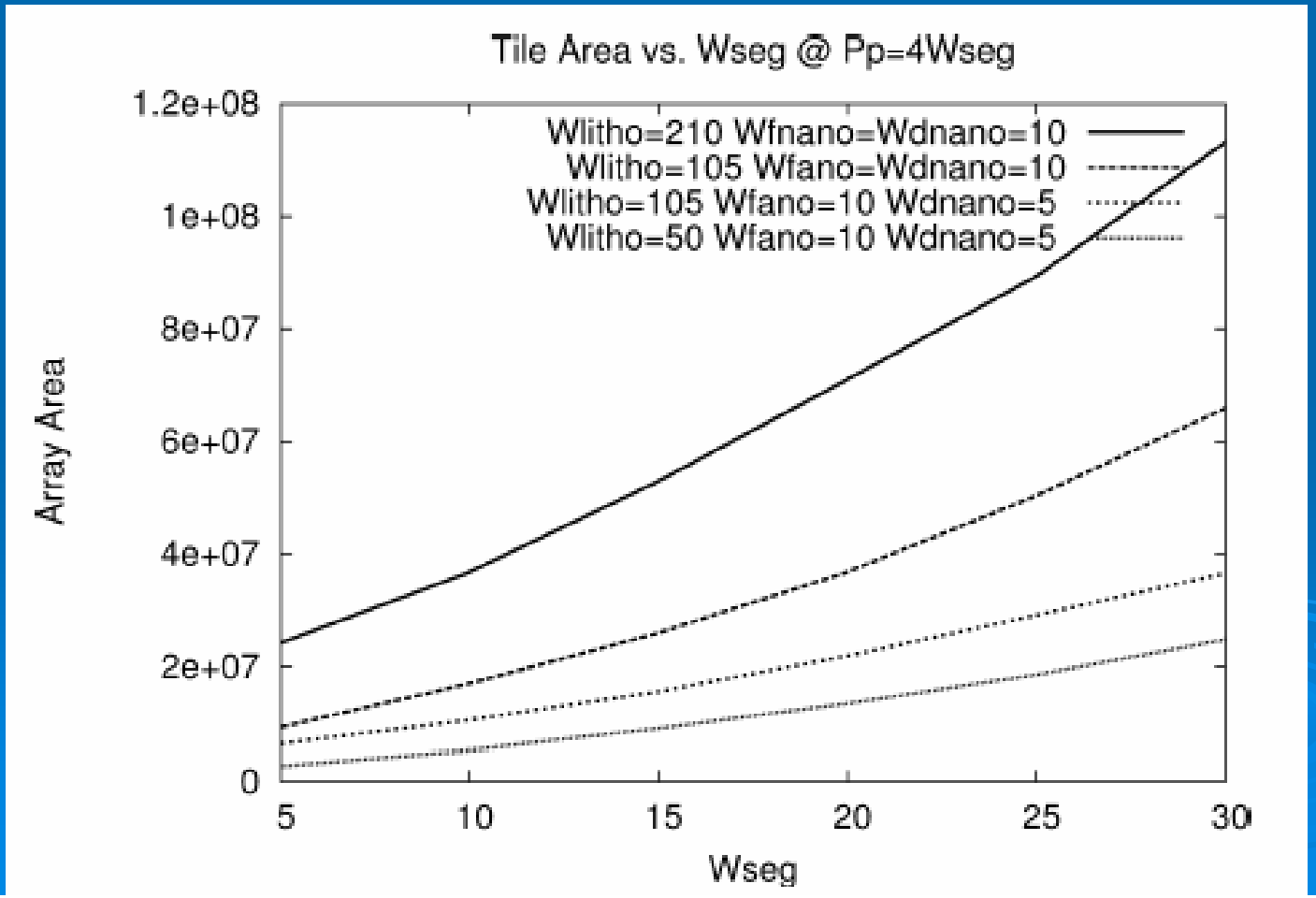
DESIGN SPACE EXPLORATION

- *Stochastic vs. Deterministic Construction.*



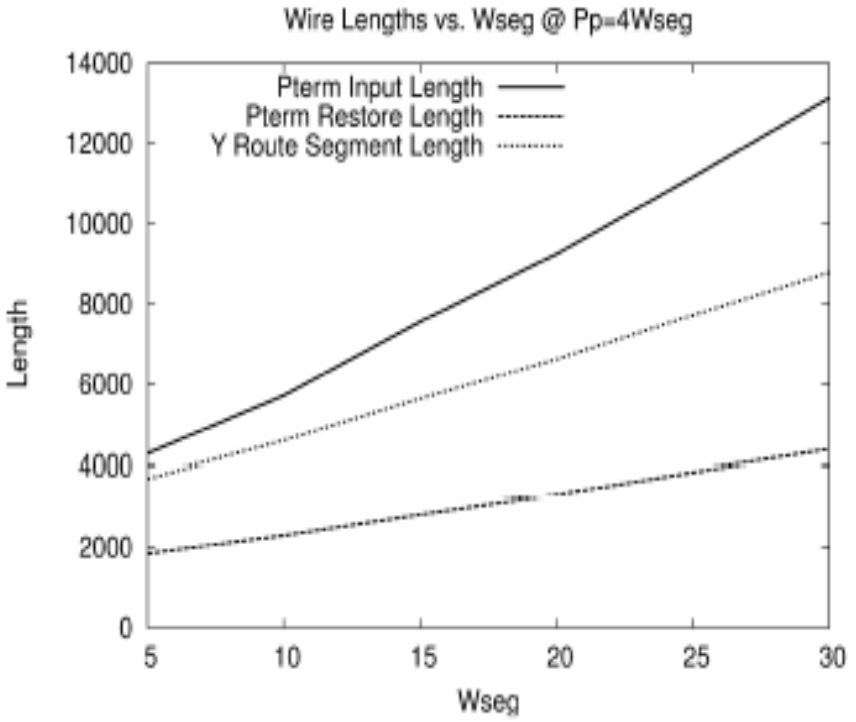
DESIGN SPACE EXPLORATION

- *Feature Sizes*. Figure shows the impact of lithographic support technology and reduced diode pitch.

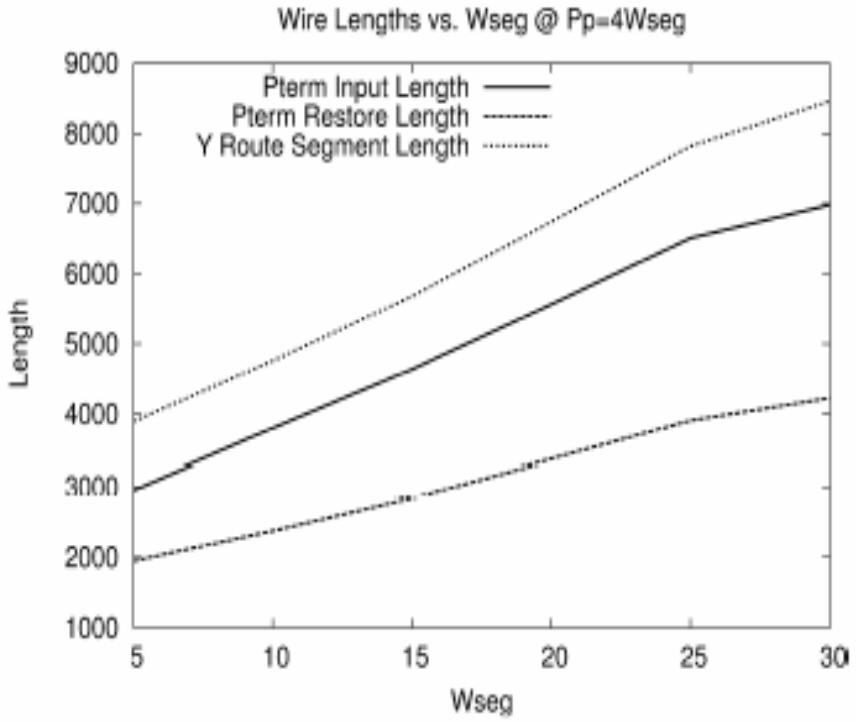


DESIGN SPACE EXPLORATION

- *NW Lengths:* Figure shows the lengths of the key NW features



$W_{dnano} = 10\text{nm}$



$W_{dnano} = 5\text{nm}$

Fig. 36. NW lengths as a function of W_{seg} for ideal restore, stochastic address case with $W_{litho} = 105\text{nm}$, $W_{fnano} = 10\text{nm}$.

DESIGN SPACE EXPLORATION

- *Delay*

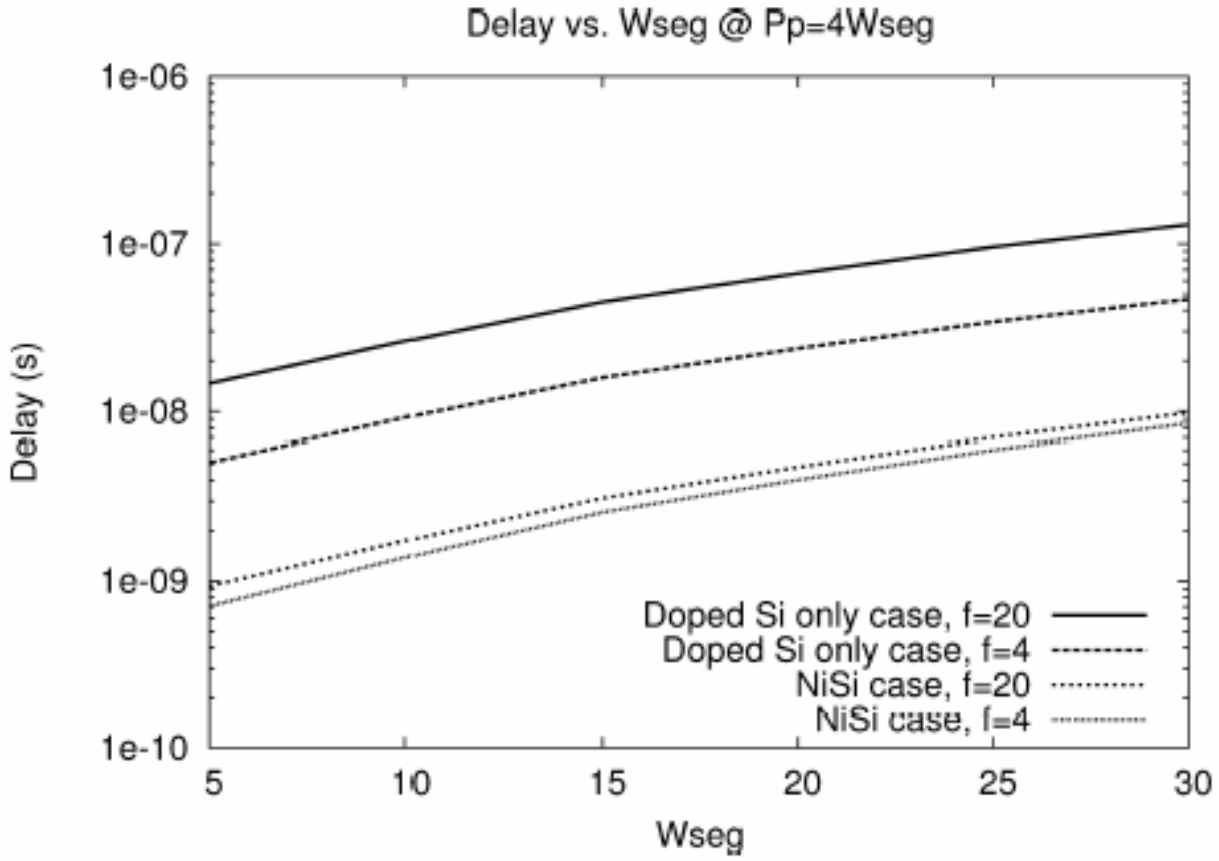


Fig. 37. Delay as a function of W_{seg} for ideal restore, stochastic address case with $W_{litho} = 105\text{nm}$, $W_{fnano} = 10\text{nm}$, $W_{dnano} = 10\text{nm}$.

DESIGN SPACE EXPLORATION

- Power Density

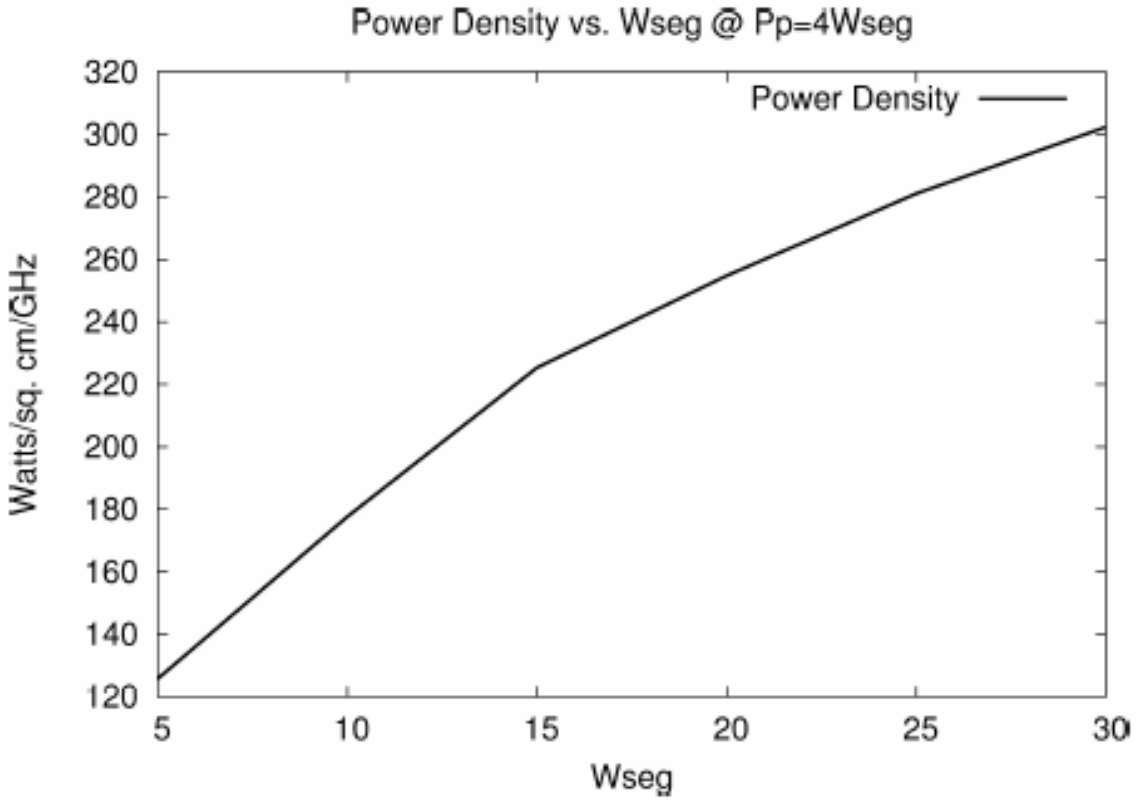


Fig. 38. Power density as a function of W_{seg} for ideal restore, stochastic address case with $W_{litho} = 105\text{nm}$, $W_{fnano} = 10\text{nm}$, $W_{dnano} = 10\text{nm}$.