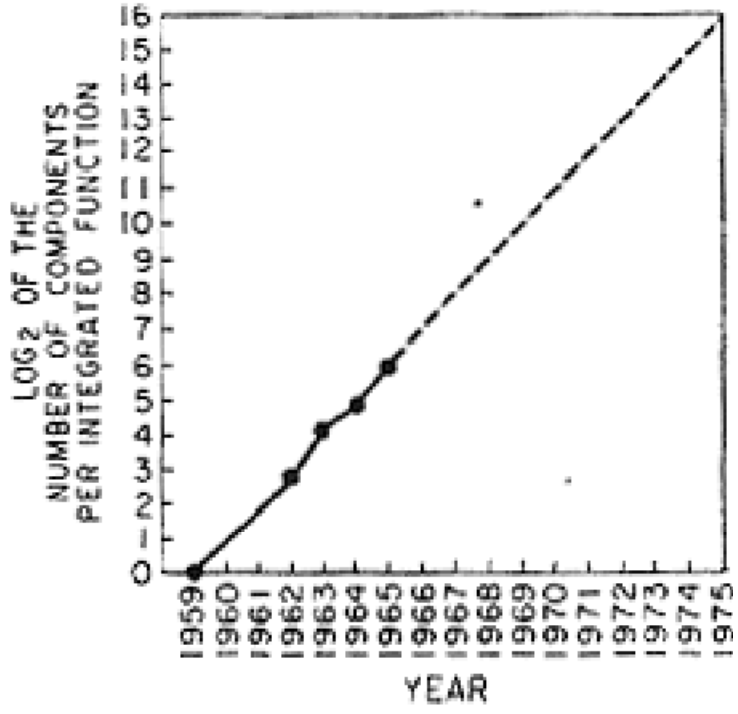
Scaling of MOS circuits

Outline

- Scaling
  - Transistors
  - Interconnect
  - Future Challenges
- VLSI Economics

Moore's Law

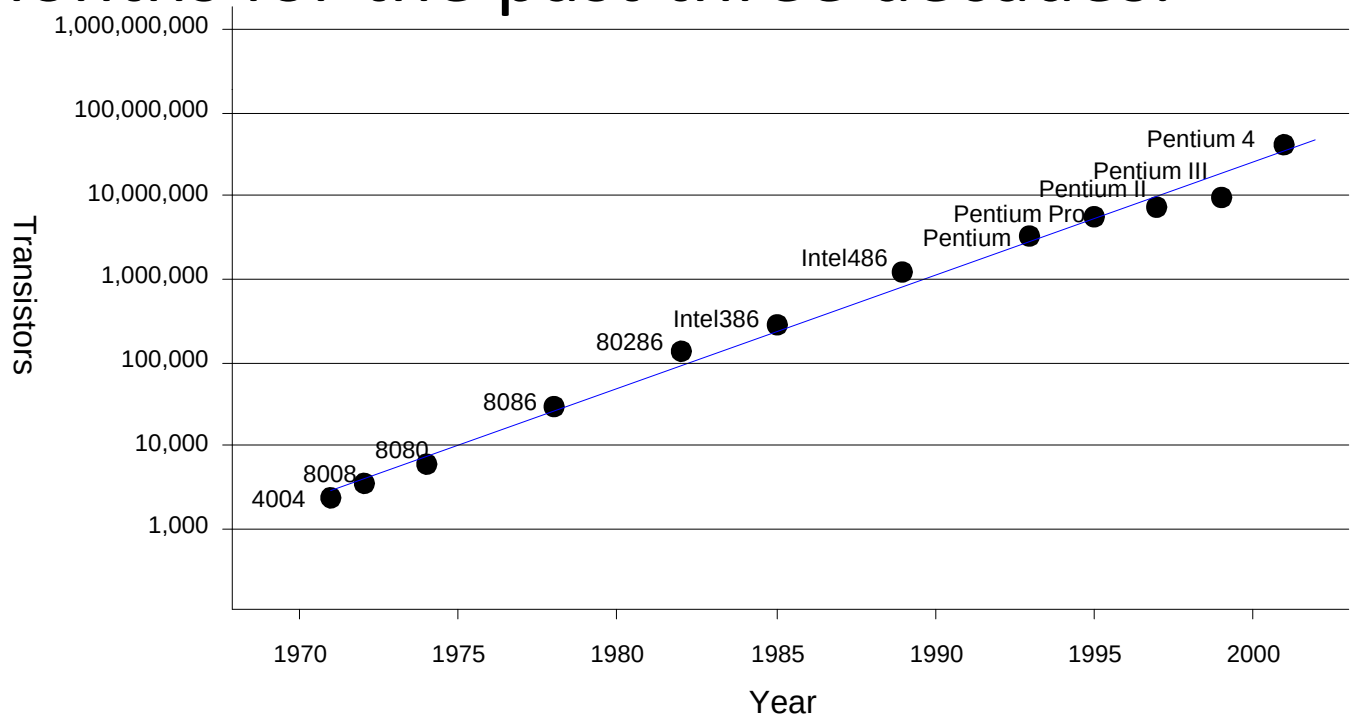
- In 1965, Gordon Moore predicted the exponential growth of the number of transistors on an IC
- Transistor count doubled
  every 18 months
  since invention
- Predicted > 65,000
  transistors by 1975!
- Growth limited by power



[Moore65]

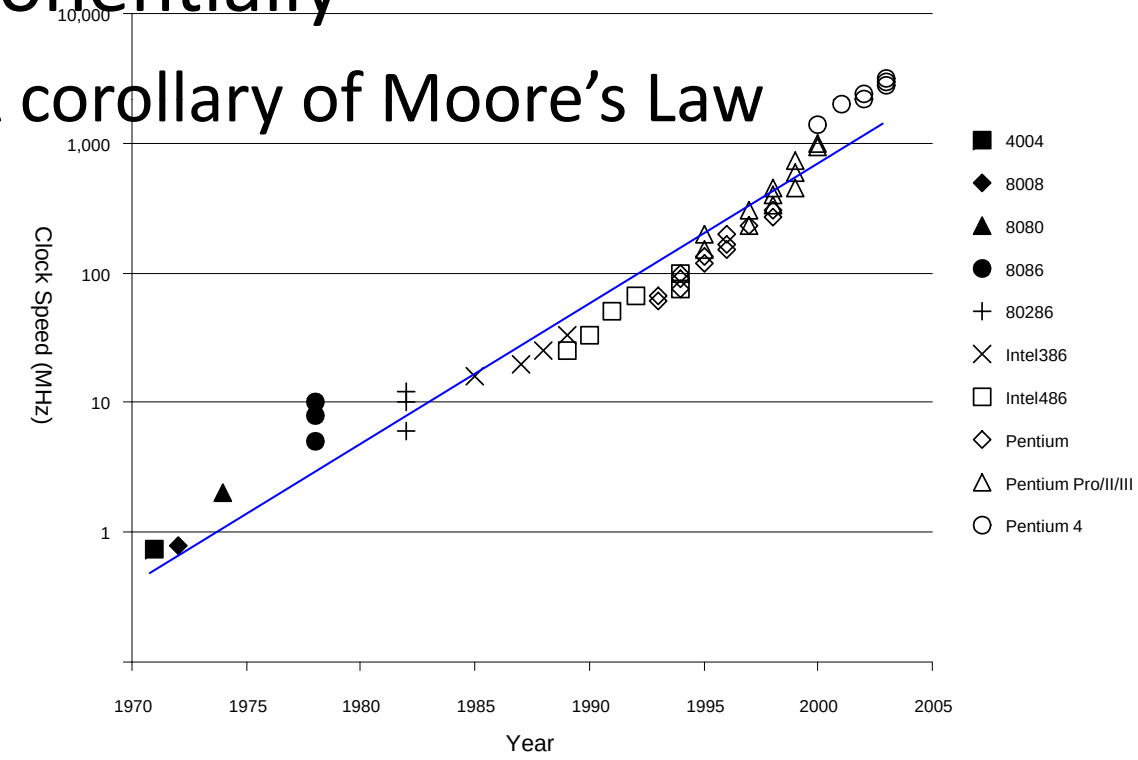
More Moore

- Transistor counts have doubled every 26 months for the past three decades.



Speed Improvement

- Clock frequencies have also increased exponentially
  - A corollary of Moore's Law



Why?

- Why more transistors per IC?

- Why faster computers?

Why?

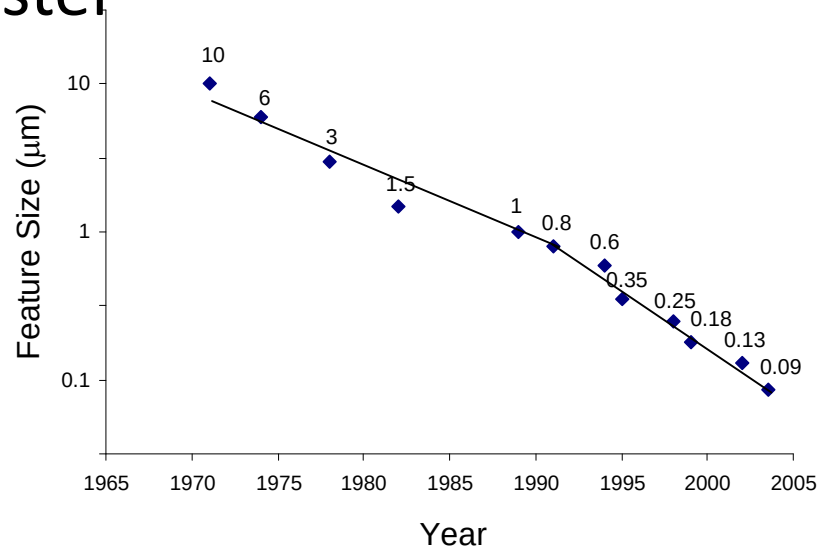
- Why more transistors per IC?
  - Smaller transistors
  - Larger dice
- Why faster computers?

Why?

- Why more transistors per IC?
  - Smaller transistors
  - Larger dice
- Why faster computers?
  - Smaller, faster transistors
  - Better microarchitecture (more IPC)
  - Fewer gate delays per cycle

Scaling

- The only constant in VLSI is constant change
- Feature size shrinks by 30% every 2-3 years
  - Transistors become cheaper
  - Transistors become faster
  - Wires do not improve (and may get worse)
- Scale factor $S = \sqrt{2}$
  - Typically
  - Technology nodes



Scaling Assumptions

- What changes between technology nodes?
- Constant Field Scaling
  - All dimensions (x, y, z => W, L, t_{ox})
  - Voltage (V_{DD})
  - Doping levels
- Lateral Scaling
  - Only gate length L
  - Often done as a quick gate shrink (S = 1.05)

SCALING FACTORS FOR DEVICE PARAMETERS

- It is important that you understand how the following parameters are effected by scaling
- Gate Area
- Gate Capacitance per unit area
- Gate Capacitance
- Charge in Channel
- Channel Resistance
- Transistor Delay
- Maximum Operating Frequency
- Transistor Current
- Switching Energy
- Power Dissipation Per Gate (Static and Dynamic)
- Power Dissipation Per Unit Area
- Power - Speed Product

Parameters		*Combined V and D*	*Constant E*	*Constant V*
V_{DD}	Supply voltage	$1/\beta$	$1/\alpha$	1
L	Channel length	$1/\alpha$	$1/\alpha$	$1/\alpha$
W	Channel width	$1/\alpha$	$1/\alpha$	$1/\alpha$
D	Gate oxide thickness	$1/\beta$	$1/\alpha$	1
A_g	Gate area	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha^2$
C_0(or C_{ox})	Gate C per unit area	β	α	1
C_g	Gate capacitance	β/α^2	$1/\alpha$	$1/\alpha^2$
C_x	Parasitic capacitance	$1/\alpha$	$1/\alpha$	$1/\alpha$
Q_{on}	Carrier density	1	1	1
R_{on}	Channel resistance	1	1	1
I_{dss}	Saturation current	$1/\beta$	$1/\alpha$	1
A_c	Conductor X-section area	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha^2$
I	Current density	α^2/β	α	α^2
V_g	Logic 1 level	$1/\beta$	$1/\alpha$	1
E_g	Switching energy	$1/\alpha^2.\beta$	$1/\alpha^3$	$1/\alpha^2$
P_g	Power dispn per gate	$1/\beta^2$	$1/\alpha^2$	1
N	Gates per unit area	α^2	α^2	α^2
P_a	Power dispn per unit area	α^2/β^2	1	α^2
T_d	Gate delay	β/α^2	$1/\alpha$	$1/\alpha^2$
f_0	Max. operating frequency	α^2/β	α	α^2
P_T	Power-speed product	$1/\alpha^2.\beta$	$1/\alpha^3$	$1/\alpha^2$

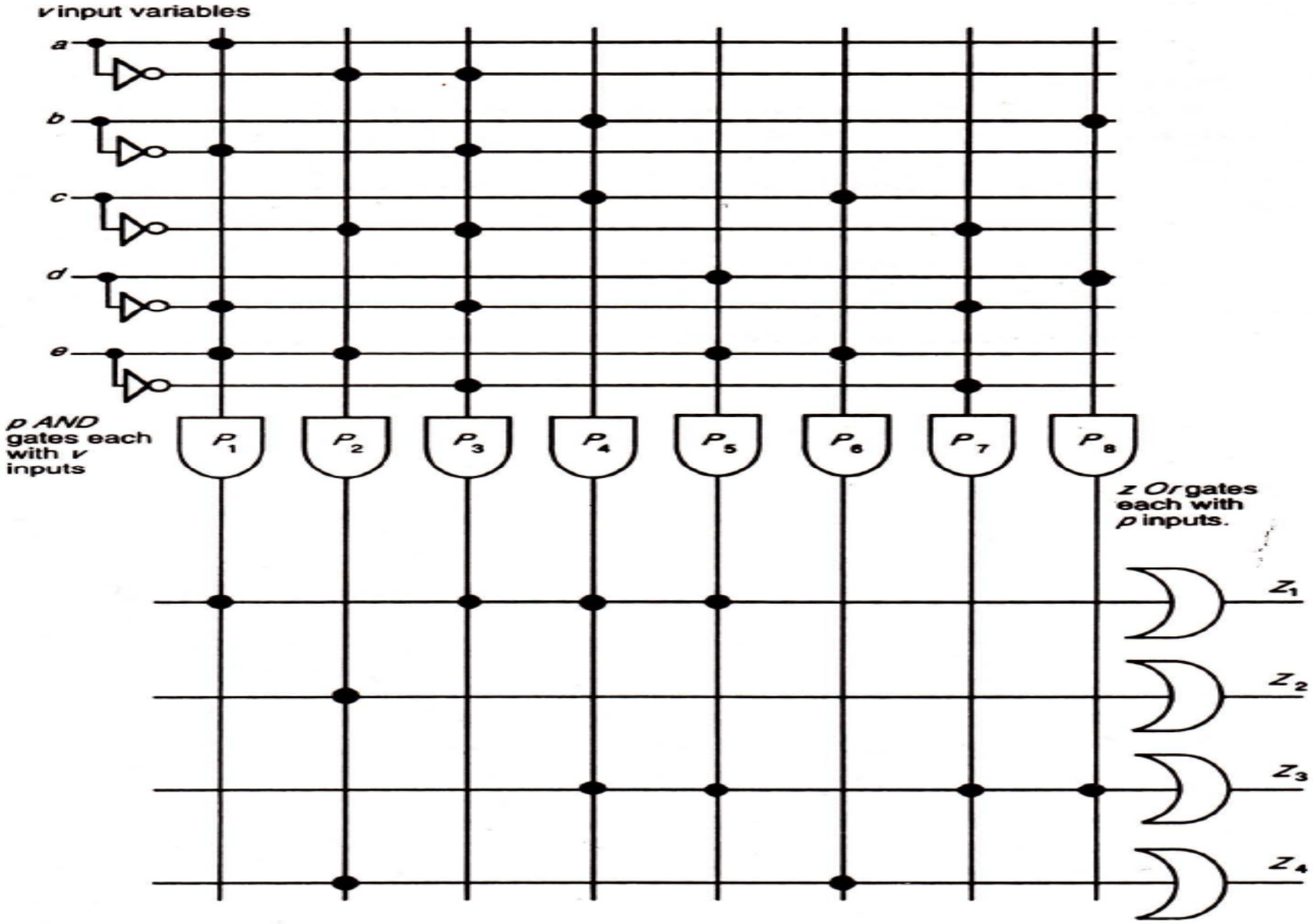
Constant E:$\beta = \alpha$; Constant V:$\beta = 1$

LIMITATIONS OF SCALING

- **Substrate Doping**
- **Limits of Miniaturization**
- Limits of Interconnect and Contact Resistance
- **LIMITS DUE TO SUBTHRESHOLD CURRENTS**
- **LIMITS ON LOGIC LEVELS AND SUPPLY VOLTAGE DUE TO NOISE**
- **LIMITS DUE TO CURRENT DENSITY**

PLA

- In VLSI design our objective is to map circuits onto silicon to meet particular specifications.
- The way in which a PLA maps onto the chip may be indicated by a 'floor plan' which gives
- the notional areas and relative disposition of the particular circuits and subsystems.
- the *Nor gate is an And gate to inverted input levels.*
- Obviously, the output *Or functions of the PLA can be realized with Nor gates each* followed by an inverter. Thus, the requirements and floor plan of the PLA may be adapted to *Nor gate form*



Note: 5 × 8 × 4 PLA shown symbolically and programmed for:

$Z_1 = p_1 + p_3 + p_4 + p_5 \therefore Z_1 = a\bar{b}d\bar{e} + \bar{a}\bar{b}c\bar{d}\bar{e} + bc + de$

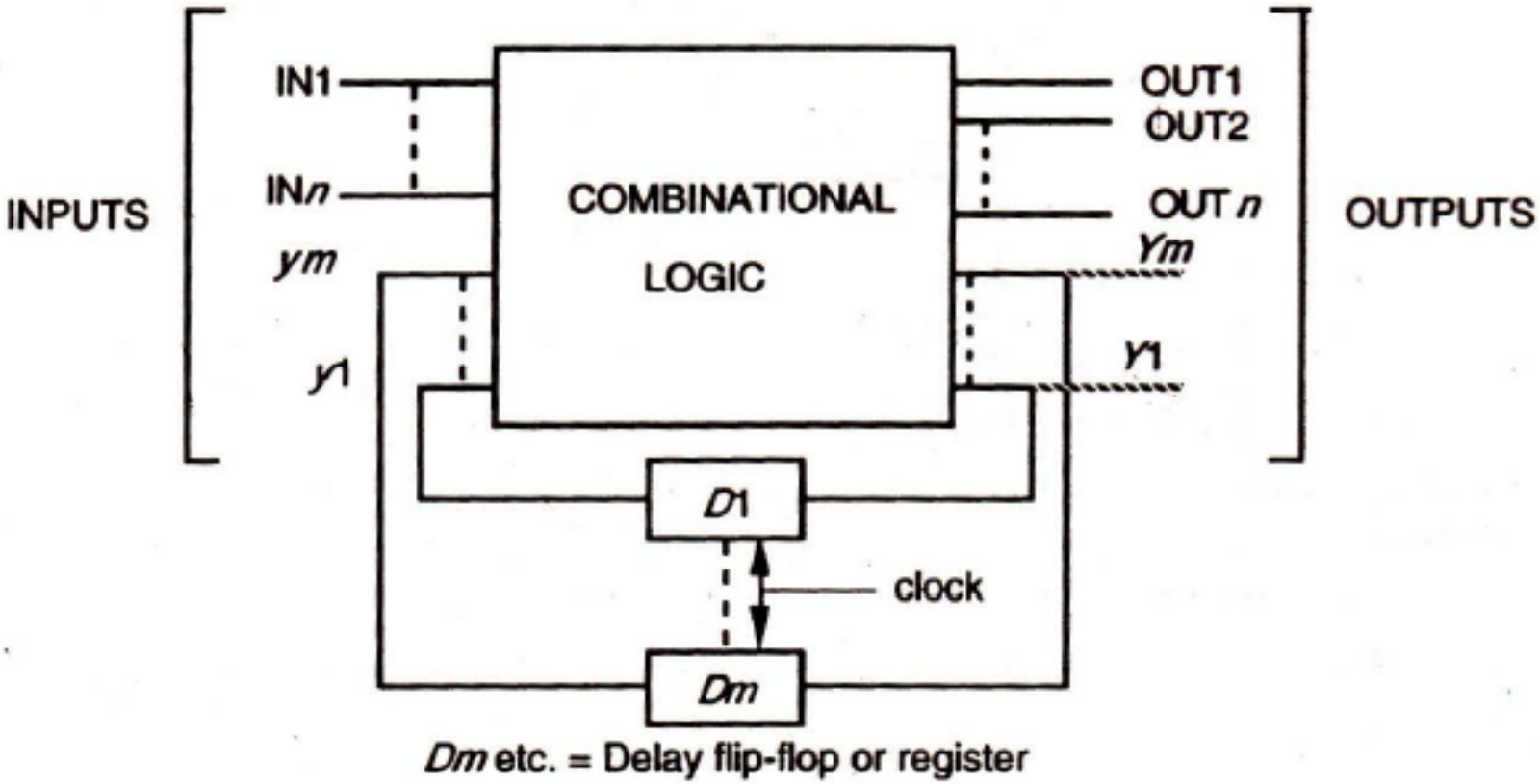
$Z_2 = p_2 \qquad \therefore Z_2 = \bar{a}\bar{c}e$

$Z_3 = p_4 + p_5 + p_7 + p_8 \therefore Z_3 = bc + de + \bar{c}\bar{d}\bar{e} + bd$

$Z_4 = p_2 + p_6 \qquad \therefore Z_4 = \bar{a}\bar{c}e + ce$

FIGURE C.1 $v \times p \times z$ PLA.

Finite state Machines.



- The 'm' feedback variables constitute the state vector and determine the maximum number of finite states which may be assumed by the circuit.
- In the most general case, the next state and the output are both functions of the present state and the independent inputs.
- The delay elements are generally assumed to be associated with the feedback path and, for clocked systems,
- the basic delay elements are flip-flops, although, in asynchronous circuits in particular, the delays may be contributed by circuit propagation delays.
- The test generation for a sequential circuit is a very complicated task since the test signals must not only be logically correct but must also occur at the correct time relative to other signals.