

# Optical Encoders

- Any transducer that generates a coded reading of a measurement can be termed an *encoder*.
- *Shaft Encoders* are digital transducers that are used for measuring angular displacements and velocities.
- Relative advantages of digital transducers over their analog counterparts:
  - High resolution (depending on the word size of the encoder output and the number of pulses per revolution of the encoder)
  - High accuracy (particularly due to noise immunity of digital signals and superior construction)

- Relative ease of adaptation in digital control systems (because transducer output is digital) with associated reduction in system cost and improvement of system reliability
- Shaft Encoders can be classified into two categories depending on the nature and method of interpretation of the output:
  - Incremental Encoders
  - Absolute Encoders
- Incremental Encoders
  - Output is a pulse signal that is generated when the transducer disk rotates as a result of the motion that is being measured.

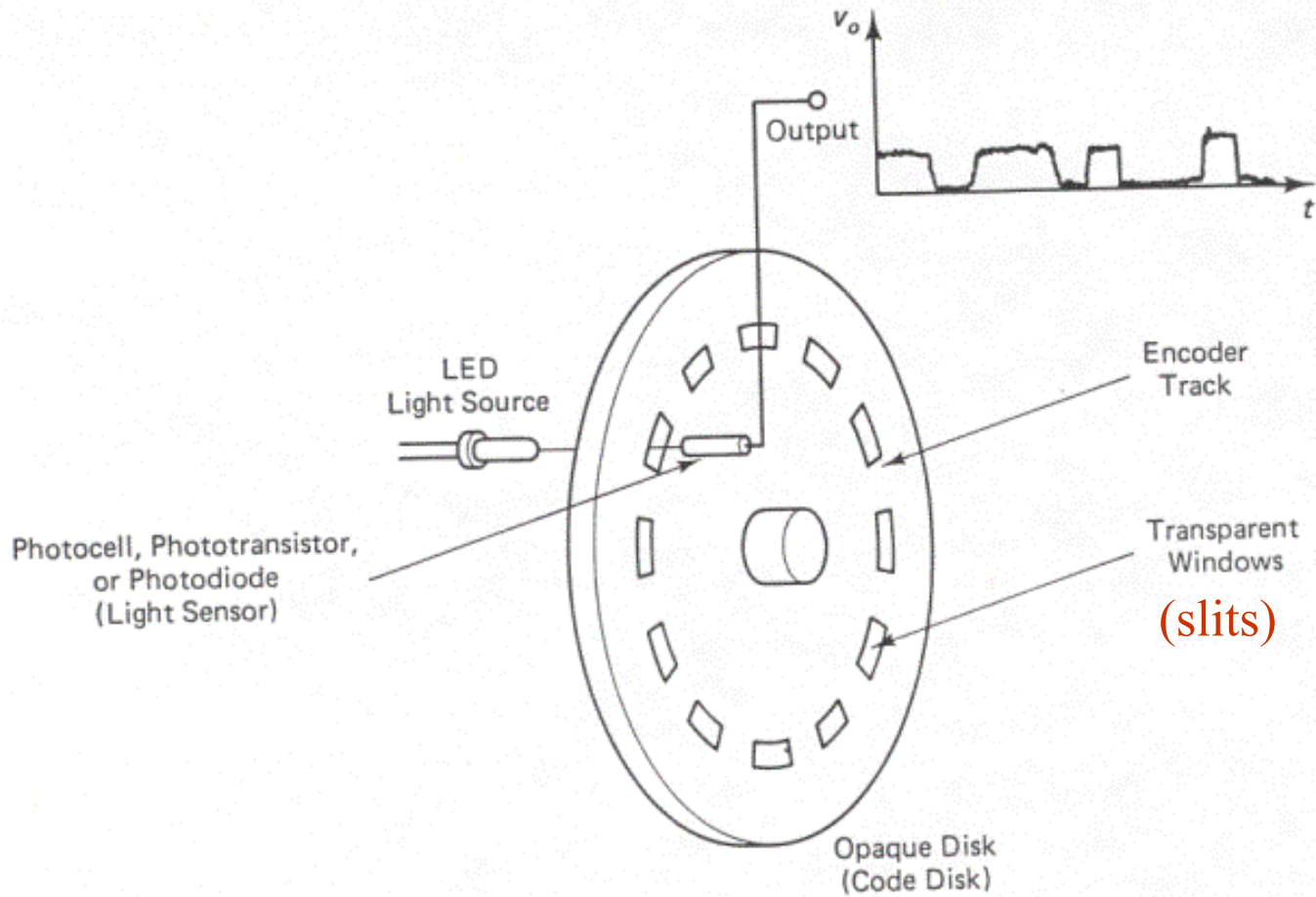
- By counting pulses or by timing the pulse width using a clock signal, both angular displacement and angular velocity can be determined.
- Displacement, however, is obtained with respect to some reference point on the disk, as indicated by a reference pulse (index pulse) generated at that location on the disk. The index pulse count determines the number of full revolutions.

- Absolute Encoders

- An absolute encoder has many pulse tracks on its transducer disk.
- When the disk of an absolute encoder rotates, several pulse trains – equal in number to the tracks on the disk – are generated simultaneously.

- At a given instant, the magnitude of each pulse signal will have one of two signal levels (i.e., a binary state) as determined by a level detector. This signal level corresponds to a binary digit (0 or 1). Hence, the set of pulse trains gives an encoded binary number at any instant.
- The pulse windows on the tracks can be organized into some pattern (code) so that each of these binary numbers corresponds to the angular position of the encoder disk at the time when the particular binary number is detected.
- Pulse voltage can be made compatible with some form of digital logic (e.g., TTL)
- Direct digital readout of an angular position is possible.

- Absolute encoders are commonly used to measure fractions of a revolution. However, complete revolutions can be measured using an additional track that generates an index pulse, as in the case of an incremental encoder.
- Signal Generation can be accomplished using any one of four techniques:
  - Optical (photosensor) method
  - Sliding contact (electrical conducting) method
  - Magnetic saturation (reluctance) method
  - Proximity sensor method
- Method of signal interpretation and processing is the same for all four types of signal generation.



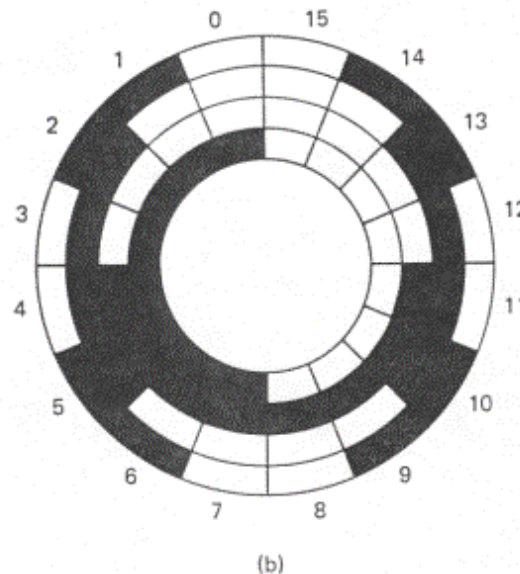
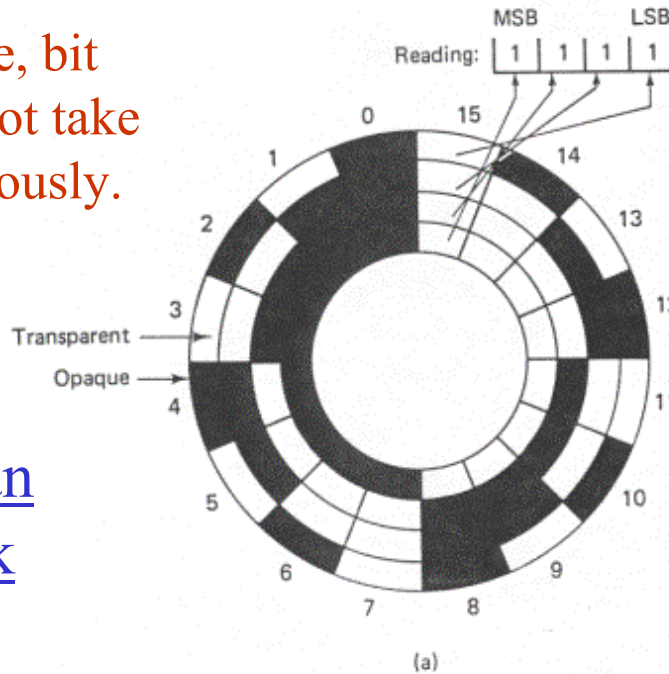
## Schematic Representation of an Optical Encoder One Track and One Pick-Off Sensor Shown

In Binary Code, bit switching may not take place simultaneously.

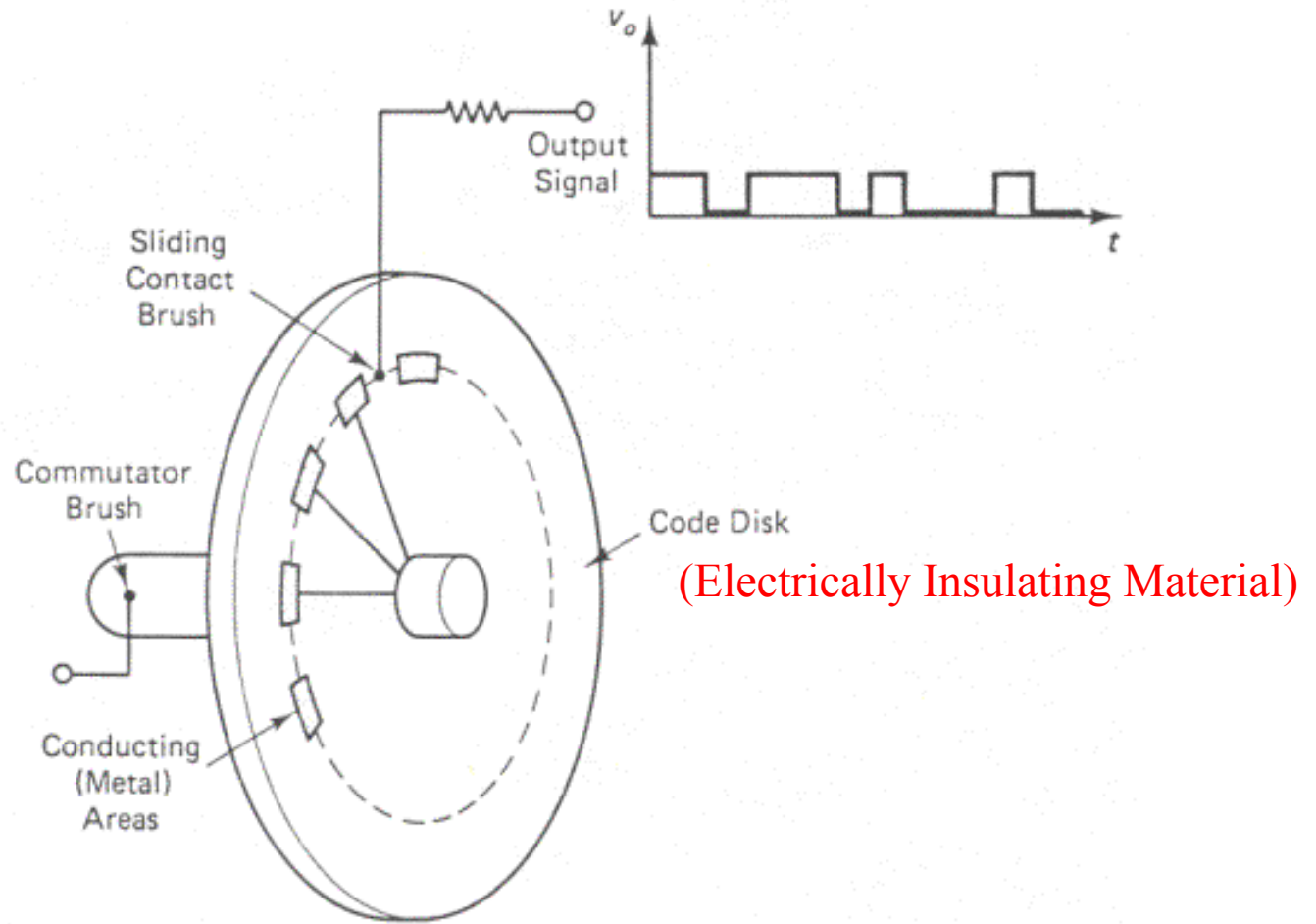
## Schematic Diagram of an Absolute Encoder Disk Pattern

- (a) Binary code
- (b) Gray code

Ambiguities in bit switching can be avoided by using gray code. However, additional logic is needed to convert the gray-coded number to a corresponding binary number.



Absolute Encoders must be powered and monitored only when a reading is taken. Also, if a reading is missed, it will not affect the next reading.

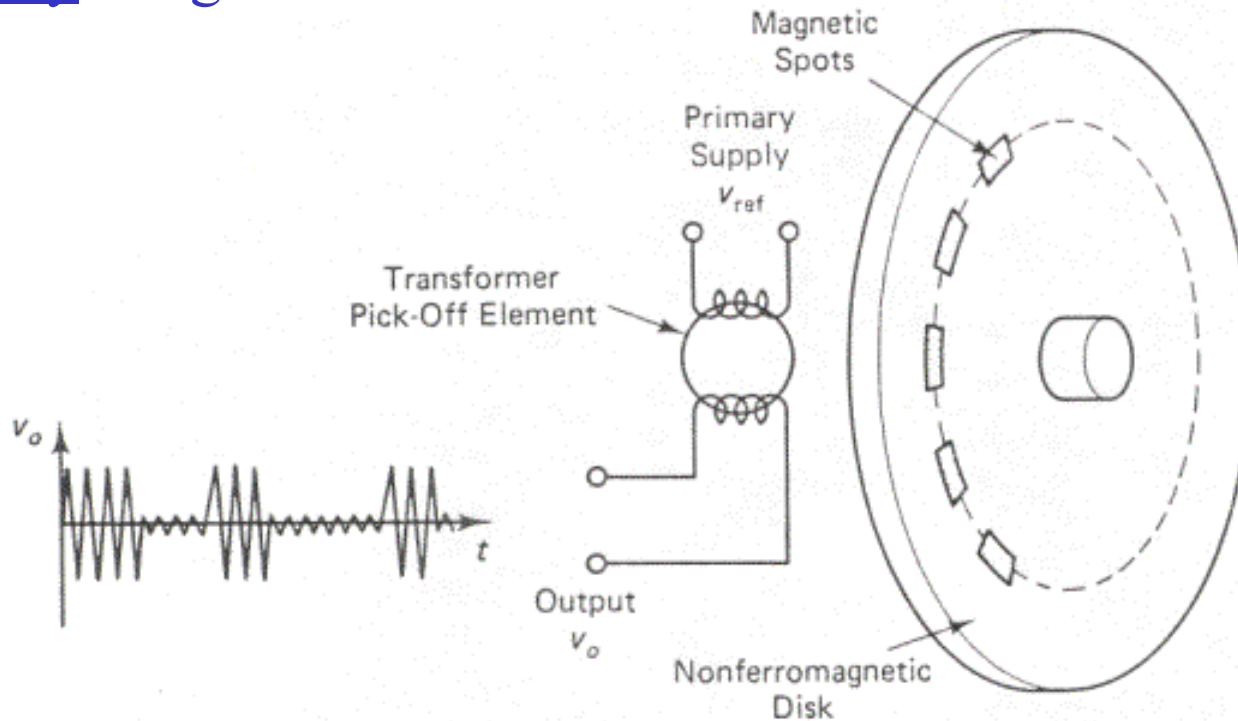


## Schematic Representation of a Sliding Contact Encoder



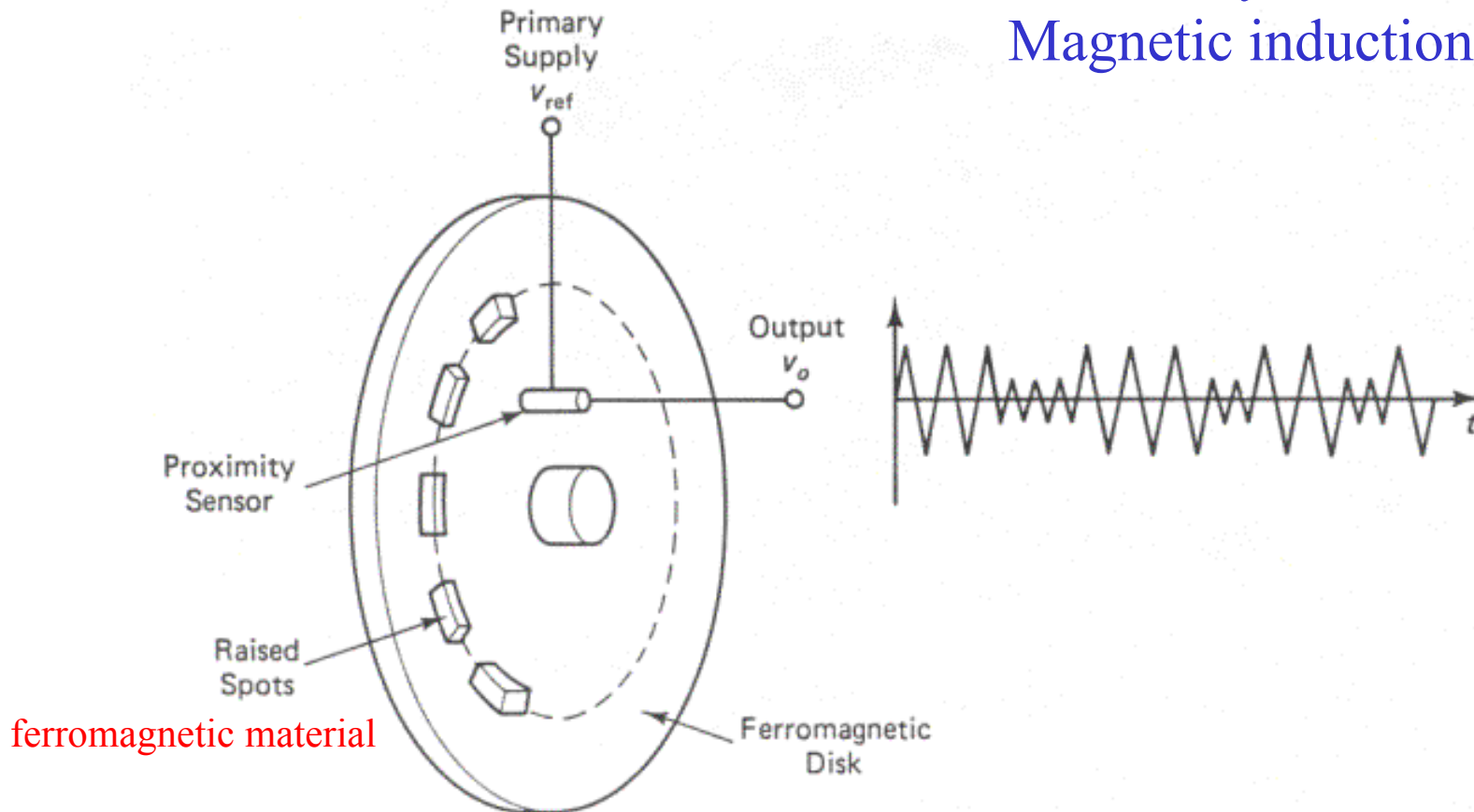
Pulse peak: nonmagnetic area

Pulse valley: magnetic area



## Schematic Representation of a Magnetic Encoder

## Proximity sensor: Magnetic induction



## Schematic Representation of a Proximity Probe Encoder

- Elements of the Optical Encoder

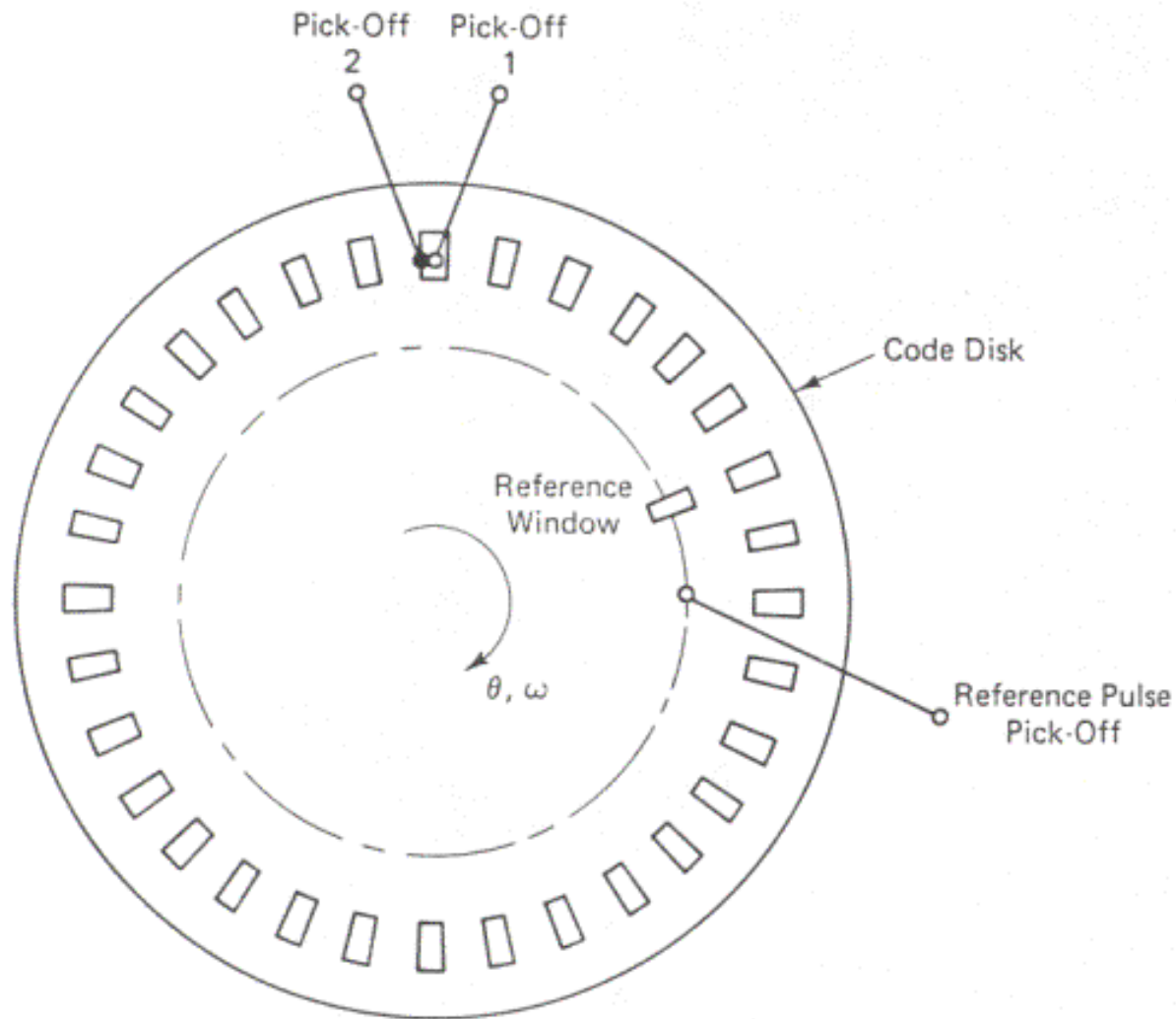
- The optical encoder uses an opaque disk (code disk) that has one or more circular tracks, with some arrangement of identical transparent windows (slits) in each track.
- A parallel beam of light (e.g., from a set of light-emitting diodes) is projected to all tracks from one side of the disk.
- The transmitted light is picked off using a bank of photosensors on the other side of the disk that typically has one sensor for each track.
- The light sensor could be a silicon photodiode, a phototransistor, or a photovoltaic cell.

- Since the light from the source is interrupted by the opaque areas of the track, the output signal from the probe is a series of voltage pulses. This signal can be interpreted to obtain the angular position and angular velocity of the disk.
- Note that an incremental encoder disk requires only one primary track that has equally spaced and identical window (pick-off) areas. The window area is equal to the area of the inter-window gap. Usually, a reference track that has just one window is also present in order to generate a pulse (known as the index pulse) to initiate pulse counting for angular position measurement and to detect complete revolutions.

- In contrast, absolute encoder disks have several rows of tracks, equal in number to the bit size of the output data word. Furthermore, the track windows are not equally spaced but are arranged in a specific pattern on each track so as to obtain a binary code (or gray code) for the output data from the transducer.
- It follows that absolute encoders need as least as many signal pick-off sensors as there are tracks, whereas incremental encoders need one pick-off sensor to detect the magnitude of rotation and an additional sensor at a quarter-pitch separation (pitch = center-to-center distance between adjacent windows) to identify the direction of rotation, i.e., the *offset sensor configuration*.

- Some designs of incremental encoders have two identical tracks, one a quarter-pitch offset from the other, and the two pick-off sensors are placed radially without any circumferential offset, i.e., the *offset track configuration*.
- A pick-off sensor for a reference pulse is also used.
- *Signal interpretation* depends on whether the particular optical encoder is an incremental device or an absolute device.
  - We will focus on the incremental optical encoder.
  - The output signals from either the offset sensor configuration or the offset track configuration are the same.

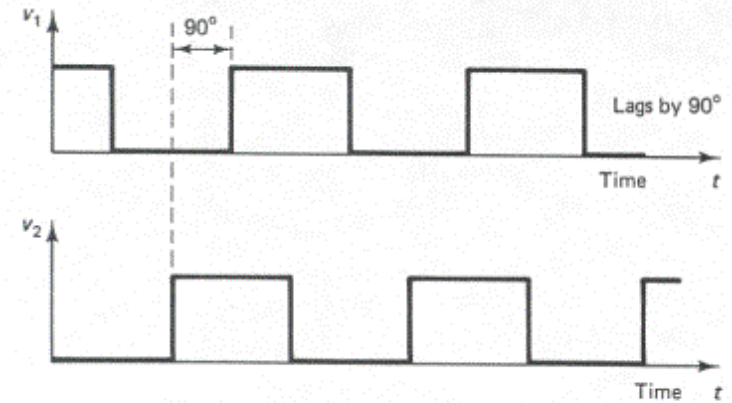
- Note that the pulse width and pulse-to-pulse period (encoder cycle) are constant in each sensor output when the disk rotates at constant angular velocity. When the disk accelerates, the pulse width decreases continuously; when the disk decelerates, the pulse width increases continuously.
- The quarter-pitch offset in sensor location or track position is used to determine the direction of rotation of the disk. It is obtained by determining the phase difference of the two output signals, using phase-detection circuitry. One method for determining the phase difference is to time the pulses using a high-frequency clock signal.



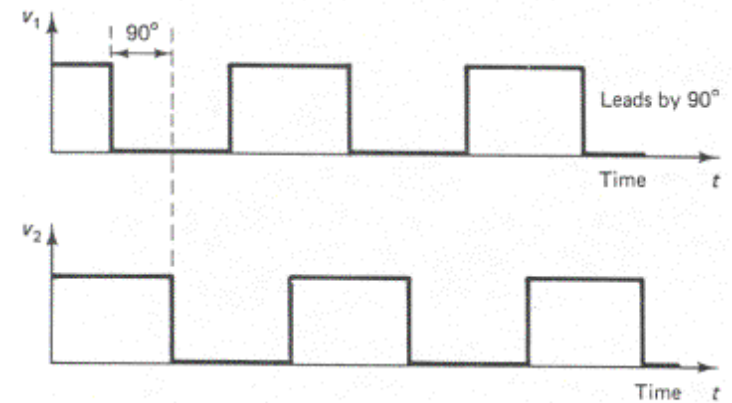
## Incremental Optical Encoder Disk Offset-Sensor Configuration



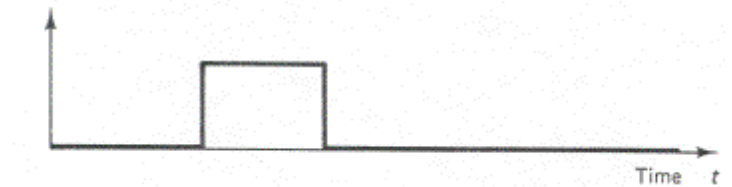
Clockwise (CW) rotation:  
 $V_1$  lags  $V_2$  by a quarter of a cycle  
 (i.e., a phase lag of  $90^\circ$ )  
 Counterclockwise (CCW) rotation:  
 $V_1$  leads  $V_2$  by a quarter of a cycle



(a)



(b)



(c)

## Incremental Encoder Pulse Signals

(a) CW rotation    (b) CCW rotation    (c) reference

- Displacement Computation

- Maximum count possible:  $M$  pulses

$$\theta = \frac{n \text{ pulses}}{M} \theta_{\max}$$

- Range of the encoder:  $\pm \theta_{\max}$

- If the data size is  $r$  bits, allowing for a sign bit,  $M = 2^{r-1}$ , where zero count is also included.

- If zero count is not included,  $M = 2^{r-1} - 1$

- If  $\theta_{\max}$  is  $2\pi$  and  $\theta_{\min}$  is zero, then  $\theta_{\max}$  and  $\theta_{\min}$  will correspond to the same position of the code disk. To avoid this ambiguity, we use

$$\theta_{\min} = \frac{\theta_{\max}}{2^{r-1}}$$

- The conventional definition for digital resolution is:

$$\frac{(\theta_{\max} - \theta_{\min})}{(2^{r-1} - 1)}$$

- Two methods are available for determining velocities using an incremental encoder:
  - *pulse-counting method*
  - *pulse-timing method*
- Pulse-Counting Method
  - The pulse count over the sampling period of the digital processor is measured and is used to calculate the angular velocity. For a given sampling period, there is a lower speed limit below which this method is not very accurate.

- To compute the angular velocity  $\omega$ , suppose that the count during a sample period  $T$  is  $n$  pulses. Hence, the average time for one pulse is  $T/n$ . If there are  $N$  windows on the disk, the average time for one revolution is  $NT/n$ . Hence  $\omega$  (*rad/s*) =  $2\pi n/NT$ .

- **Pulse-Timing Method**

- The time for one encoder cycle is measured using a high-frequency clock signal. This method is particularly suitable for measuring low speeds accurately.
- Suppose that the clock frequency is  $f$  Hz. If  $m$  cycles of the clock signal are counted during an encoder period (interval between two adjacent windows), the time for that encoder cycle (i.e., the time to rotate through one encoder pitch) is given by  $m/f$ .

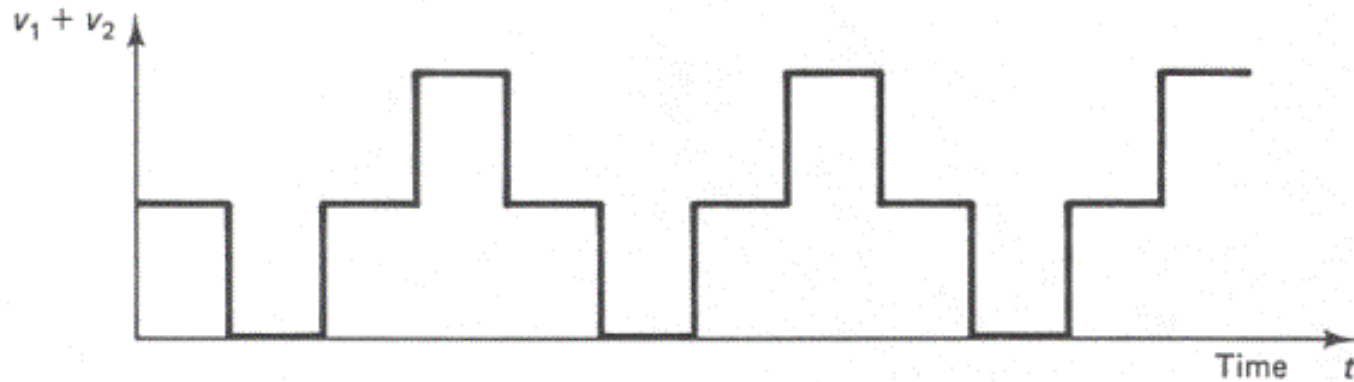
- With a total of  $N$  windows on the track, the average time for one revolution of the disk is  $Nm/f$ . Hence  $\omega = 2\pi f/Nm$ .
- Resolution of an Encoder
  - The resolution of an encoder represents the smallest change in measurement that can be measured realistically. Since an encoder can be used to measure both displacement and velocity, a resolution can be identified for each case.
  - Displacement Resolution
    - *Displacement resolution* is governed by the number of windows  $N$  in the code disk and the digital size (number of bits)  $r$  of the buffer (counter output). The *physical resolution* is determined by  $N$ . If only one pulse signal is used (i.e., no direction sensing), and if the rising edges of the pulses are detected (i.e., full cycles of the encoder are counted), the physical resolution is given by  $(360/N)^\circ$ .

- But if both pulse signals (quadrature signals) are available and the capability to detect rising and falling edges of a pulse is also present, four counts can be made per encoder cycle, thereby improving the resolution by a factor of four.
- Hence, the physical resolution is given by  $(360/4N)^\circ$ . When the two signals  $V_1$  and  $V_2$  are added, the resulting signal has a transition at every quarter of the encoder cycle. By detecting each transition (through edge detection or level detection), four pulses can be counted within every main cycle.
- Assuming that the maximum angle measured is  $360^\circ$  (or  $\pm 180^\circ$ ), the *digital resolution* is given by: 
$$\Delta\theta_d = \frac{180^\circ}{2^{r-1}} = \frac{360^\circ}{2^r}$$
- A digital word containing  $r$  bits can represent  $2^r$  different values (unsigned). Note that  $0^\circ$  and  $360^\circ$  represent the same position of the code disk. An ambiguity does not arise if we take the minimum value of  $\theta$  to be  $360^\circ/2^r$ , not zero.

- *The larger of the two resolutions governs the displacement resolution of the encoder:*

$$\Delta\theta_p = \frac{360^\circ}{4N} \quad \Delta\theta_d = \frac{360^\circ}{2^r}$$

- The physical resolution of an encoder can be improved by using step-up gearing so that one rotation of the moving object that is being monitored corresponds to several rotations of the code disk of the encoder. This improvement is directly proportional to the gear ratio.
- In summary, the *displacement resolution* of the incremental encoder depends on the following factors:
- Number of windows on the code track
  - Gear ratio
  - Word size of the measurement buffer



## Quadrature Signal Addition to Improve Physical Resolution



## – *Velocity Resolution*

- An incremental encoder is also a velocity-measuring device. The *velocity resolution* of an incremental encoder depends on the method that is employed to determine velocity. Since the pulse-counting method and the pulse-timing method are both based on counting, the resolution corresponds to the change in angular velocity that results from changing (incrementing or decrementing) the count by one. If the pulse-counting method is employed, a unity change in the count  $n$  corresponds to a speed change of

$$\Delta\omega_c = \frac{2\pi}{NT}$$

- $N$  is the number of windows in the code track and  $T$  is the sampling period.
- This velocity resolution is independent of the angular velocity itself. The resolution improves, however, with the number of windows and the sampling period.

- But under transient conditions, the accuracy of the velocity reading decreases with increasing  $T$  (the sampling frequency has to be at least double the highest frequency of interest in the velocity signal). Hence, the sampling period should not be increased indiscriminately.
- If the pulse-timing method is employed, the velocity resolution is given by:
 
$$\Delta\omega_t = \frac{2\pi f}{Nm} - \frac{2\pi f}{N(m+1)} = \frac{2\pi f}{Nm(m+1)}$$
- $f$  is the clock frequency. For large  $m$ ,  $(m+1)$  can be approximated by  $m$ . Then
 
$$\Delta\omega_t = \frac{N\omega^2}{2\pi f}$$
- Note that in this case, the resolution degrades quadratically with speed. For a given speed, the resolution degrades with increasing  $N$ . The resolution can be improved, however, by increasing the clock frequency. Gearing up has a detrimental effect on the speed resolution in the pulse-timing method, but it has a favorable effect in the pulse-counting method.

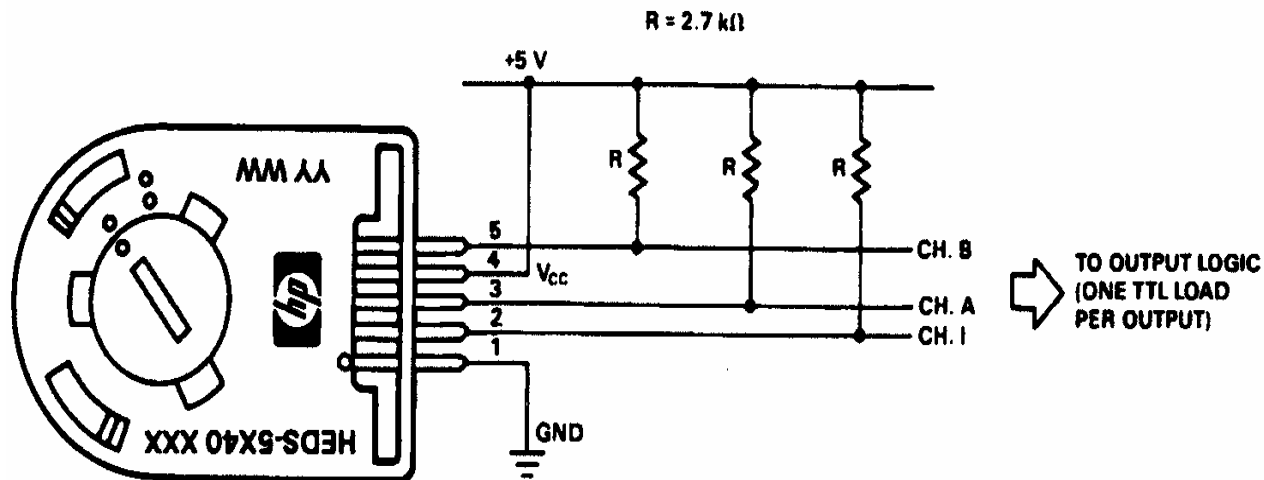
- In summary, the *speed resolution* of an incremental encoder depends on the following factors:
  - number of windows  $N$
  - sampling period  $T$
  - clock frequency  $f$
  - speed  $\omega$
  - gear ratio
- Errors in shaft encoder readings can come from several factors:
  - Quantization error (due to digital word size limitations)
  - Assembly error (eccentricity, etc.)
  - Coupling error (gear backlash, belt slippage, loose fit, etc.)

- Structural limitations (disk deformation and shaft deformation due to loading)
- Manufacturing tolerances (errors from inaccurately imprinted code patterns, inexact positioning of the pick-off sensors, limitations and irregularities in signal generation and sensing components, etc.)
- Ambient effects (vibration, temperature, light noise, humidity, dirt, smoke, etc.)
- These factors can result in erroneous displacement and velocity readings and inexact direction detection.
- One form of error in an encoder reading is the hysteresis.

- For a given position of the moving object, if the encoder reading depends on the direction of motion, the measurement has a hysteresis error.
- In that case, if the object rotates from position A to position B and back to position A, for example, the initial and final readings of the encoder will not match.
- The causes of hysteresis include backlash in gear couplings, loose fits, mechanical deformation in the code disk and shaft, delays in electronic circuitry (electrical time constants), and noisy pulse signals that make the detection of pulses (by level detection or edge detection) less accurate.

## Properties of the HP HEDS-5505-A14 Incremental Optical Encoder

Specification	Values
Supply Voltage	5 Volts
Supply Current	40 mA
Resolution	500 pulses per revolution
Resolution with quadrature	2000 pulses per revolution
Number of Channels	2
Index	none
Maximum Velocity	30000 RPM
Maximum Acceleration	250000 rad/sec <sup>2</sup>



- Problem # 1

- Explain how resolution of a shaft encoder could be improved by pulse interpolation.
- Suppose that a pulse generated from an incremental encoder can be approximated by:

$$v = v_o \left( 1 + \sin \frac{2\pi\theta}{\Delta\theta} \right)$$

- Here  $\theta$  denotes the angular position of the encoder window with respect to the photosensor position.

- Let us consider rotations of a half pitch or smaller, i.e.,  $0 \leq \theta \leq \Delta\theta/2$ , where  $\Delta\theta$  is the window pitch angle.
- By using this sinusoidal approximation for a pulse, show that we can improve the resolution of an encoder indefinitely simply by measuring the shape of each pulse at clock cycle intervals using a high-frequency clock signal.



- Problem # 2

- Consider the two quadrature pulse signals (say, A and B) from an incremental encoder. Using sketches of these signals, show that in one direction of rotation, signal B is at a high level during the up-transition of signal A, and in the opposite direction of rotation, signal B is at a low level during the up-transition of signal A. Note that the direction of motion can be determined in this manner by using level detection of one signal during the up-transition of the other signal.

- Problem # 3

- A shaft encoder that has  $N$  windows per track is connected to a shaft through a gear system with gear ratio  $p$ . Derive formulas for calculating angular velocity of the shaft by (a) the pulse-counting method, and (b) the pulse-timing method. What is the speed resolution in each case? What effect does the step-up gearing have on the speed resolution?