




Machine Learning for Physical Sensing & Perception

CMSC 818V : Fall 2026

Lecture 1: From the Physical World to Sensor Data

Nirupam Roy

COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

Tu-Th 12:30 – 1:45pm

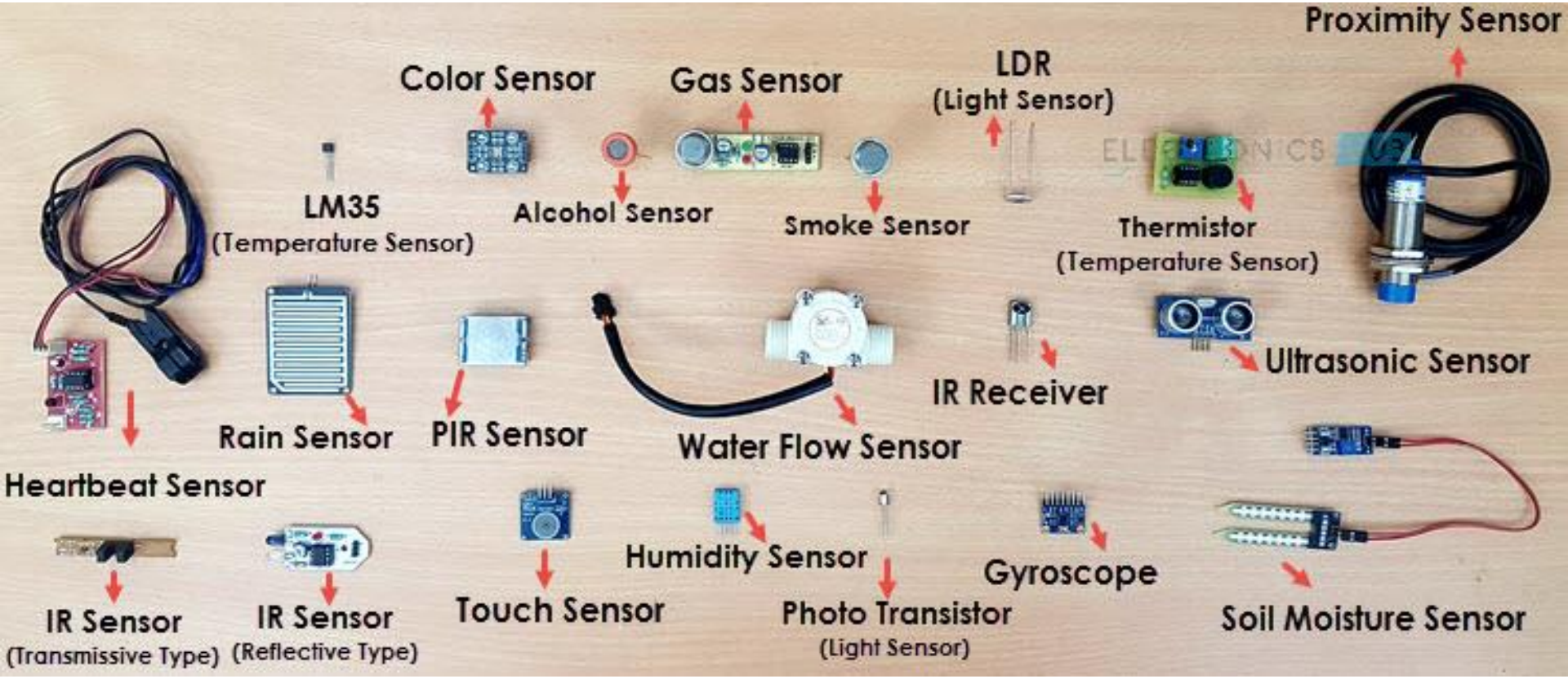
IRB 2107





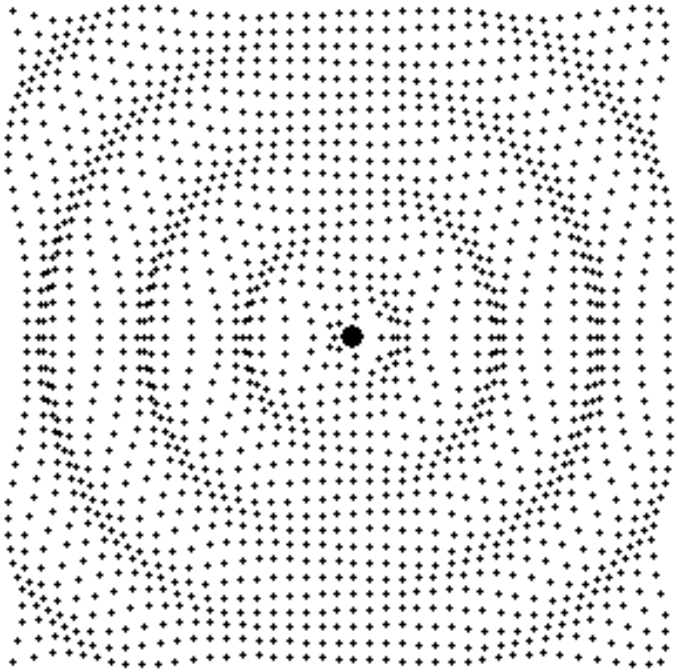


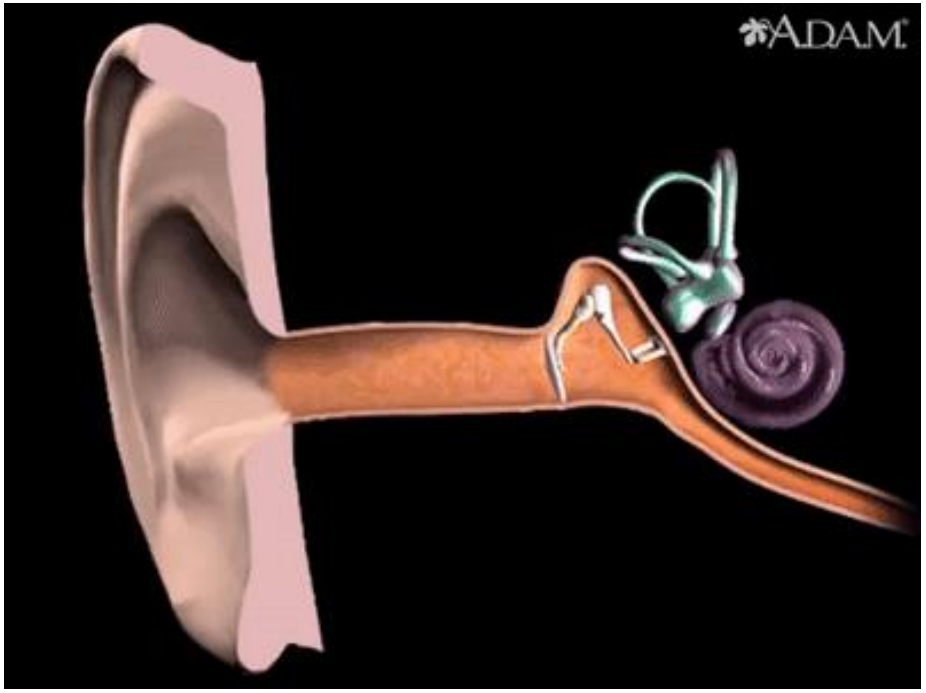
What is a sensor?



Is there a common principle?

How does a sound sensor work?



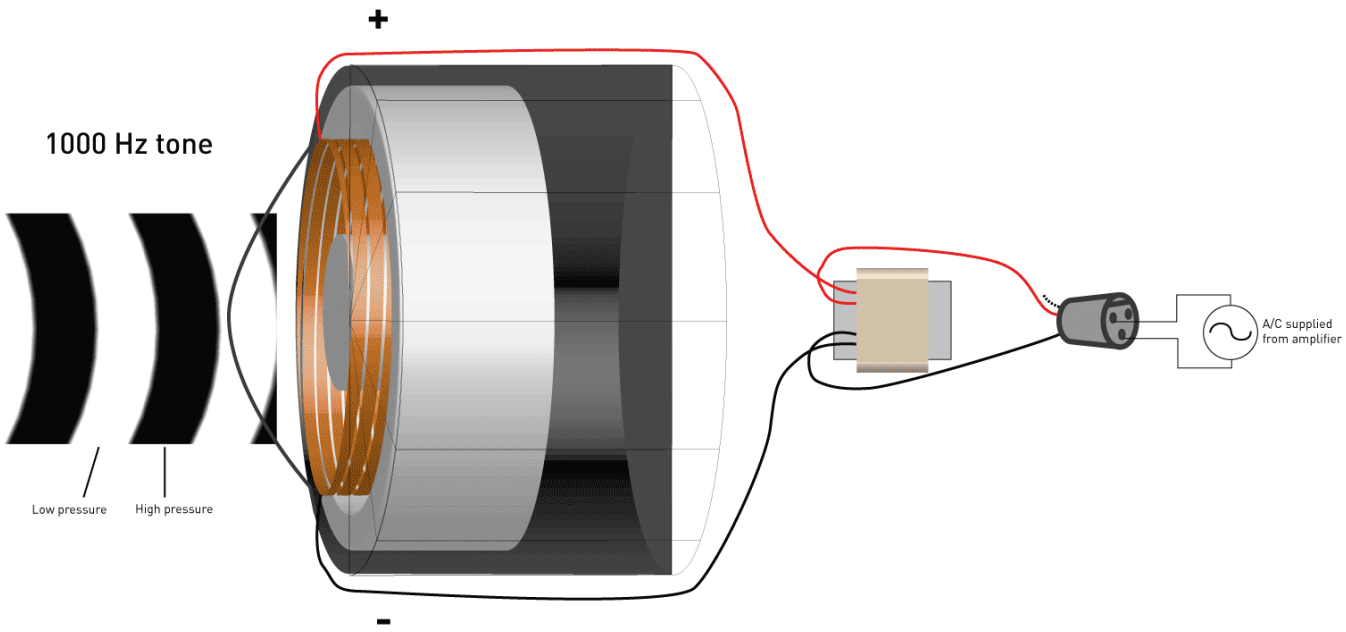


Sound

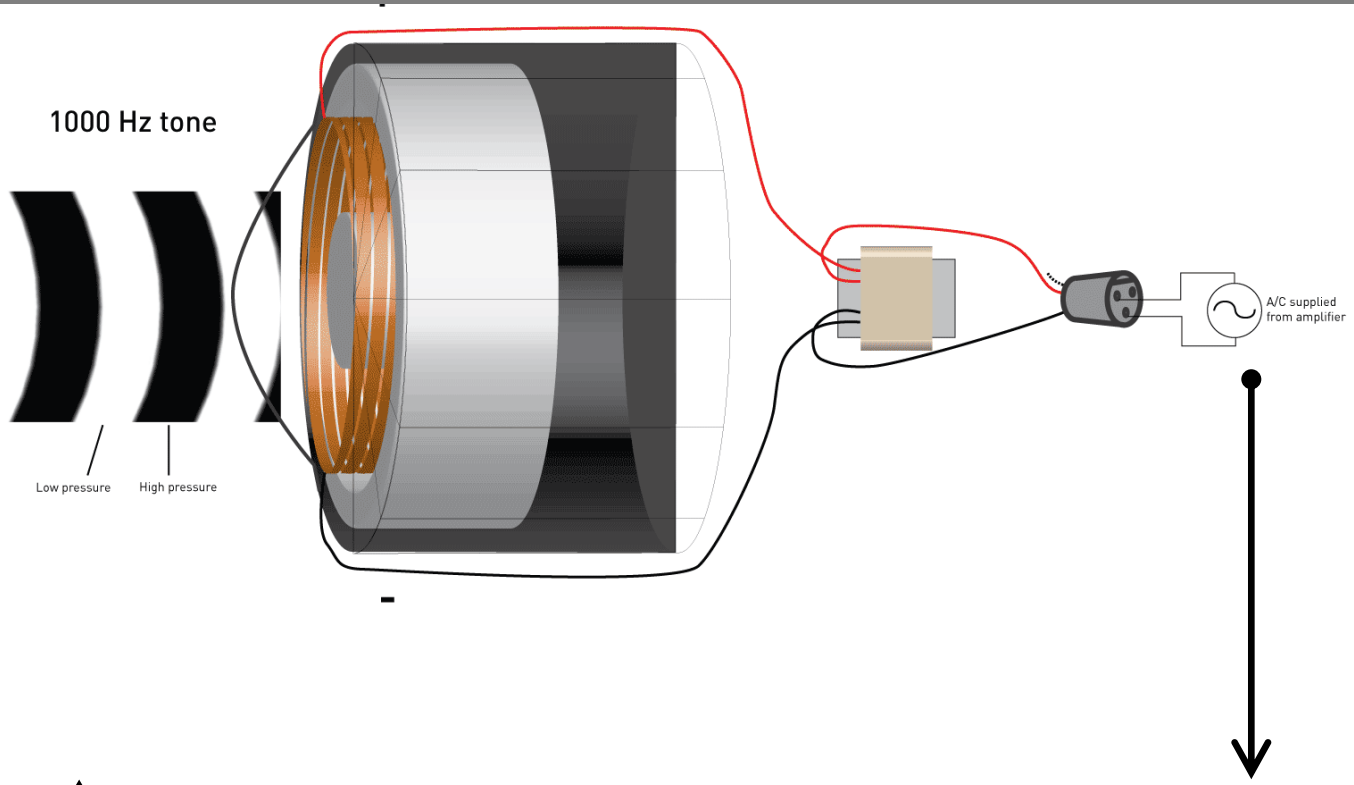
Sensor

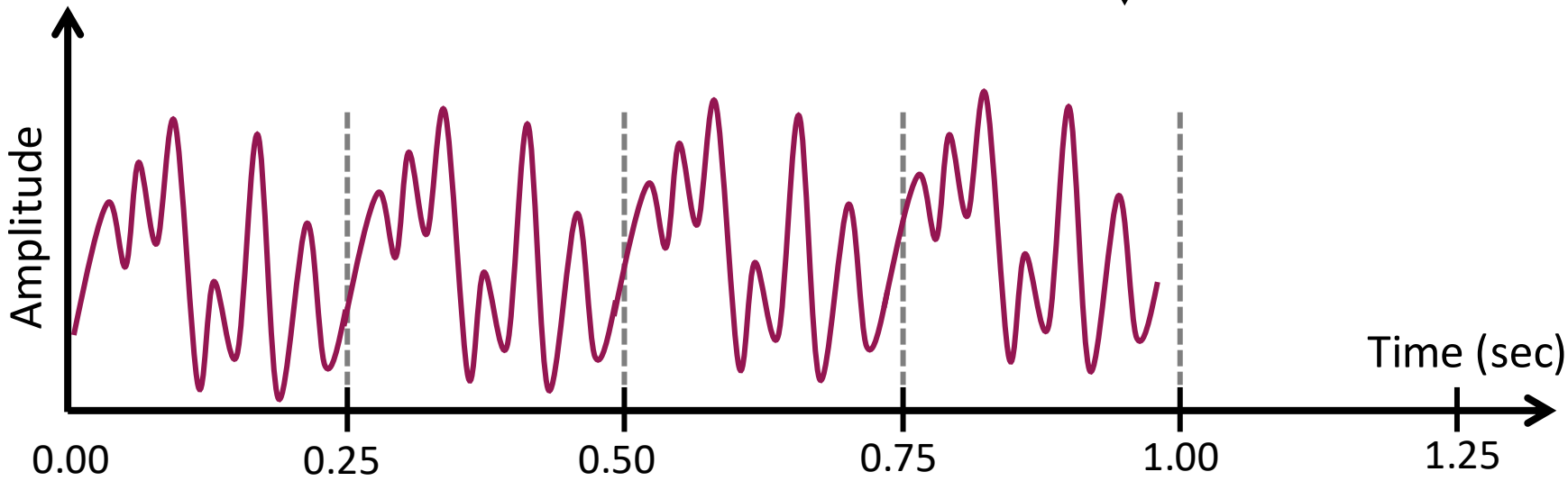
How does a sound sensor (microphone) work?

How does a sound sensor (microphone) work?



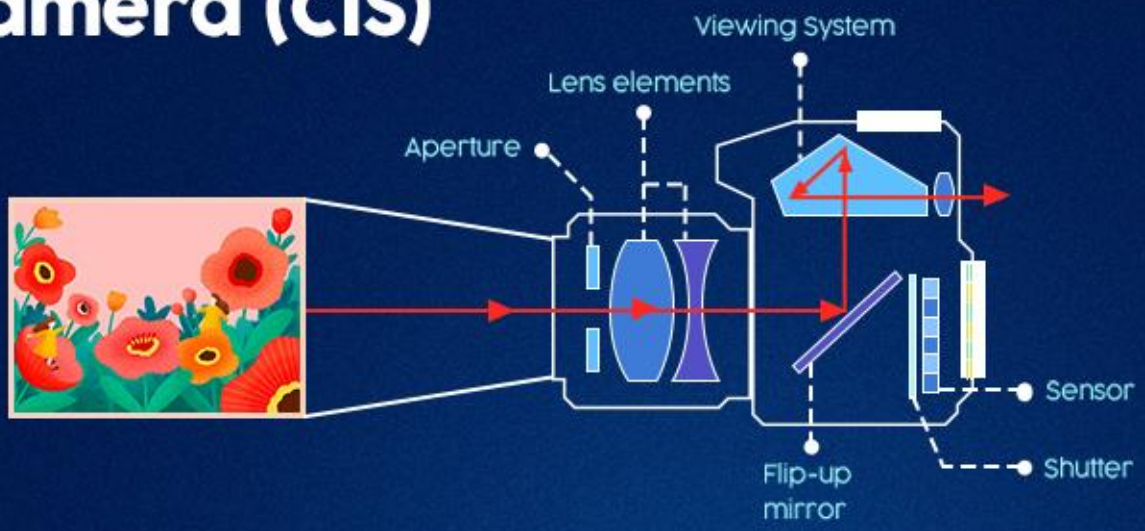
How does a sound sensor (microphone) work?

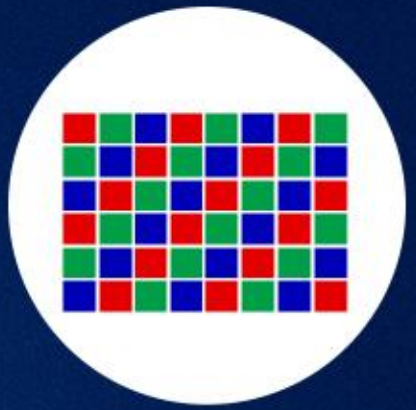




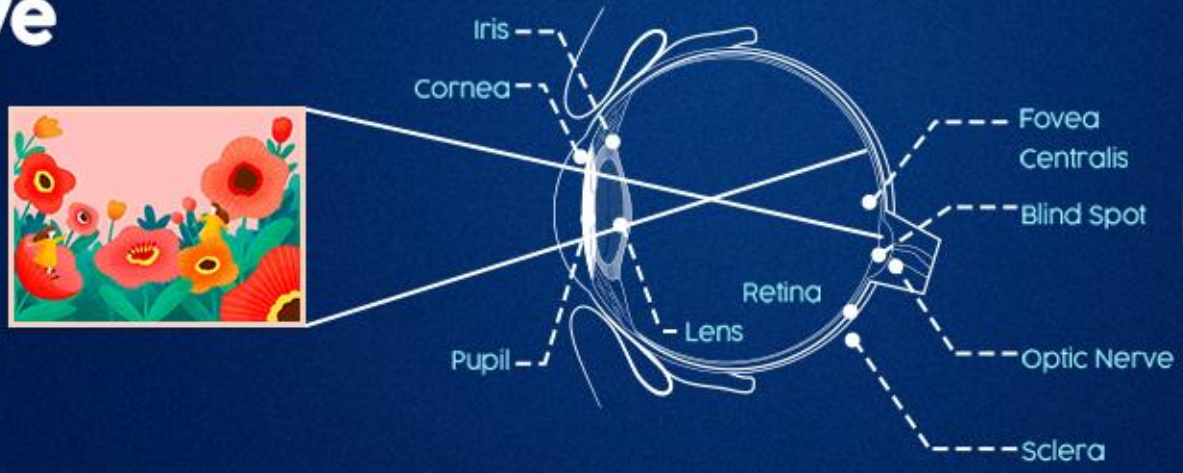
How does a camera sensor work?

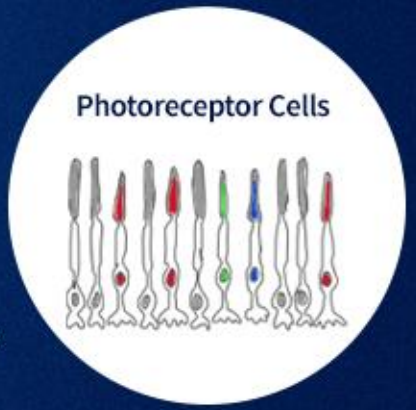
Camera (CIS)





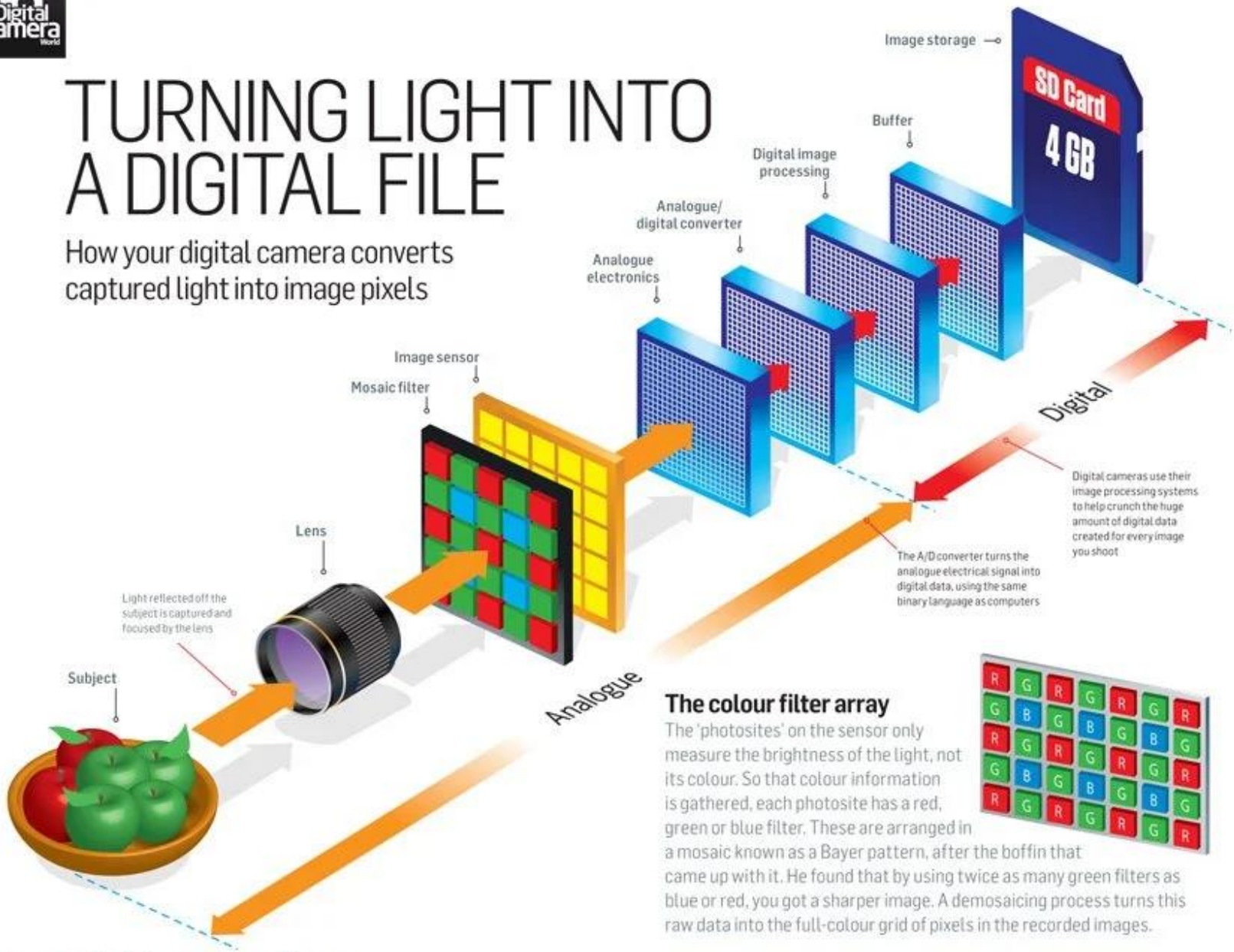
Eye





TURNING LIGHT INTO A DIGITAL FILE

How your digital camera converts captured light into image pixels



The colour filter array

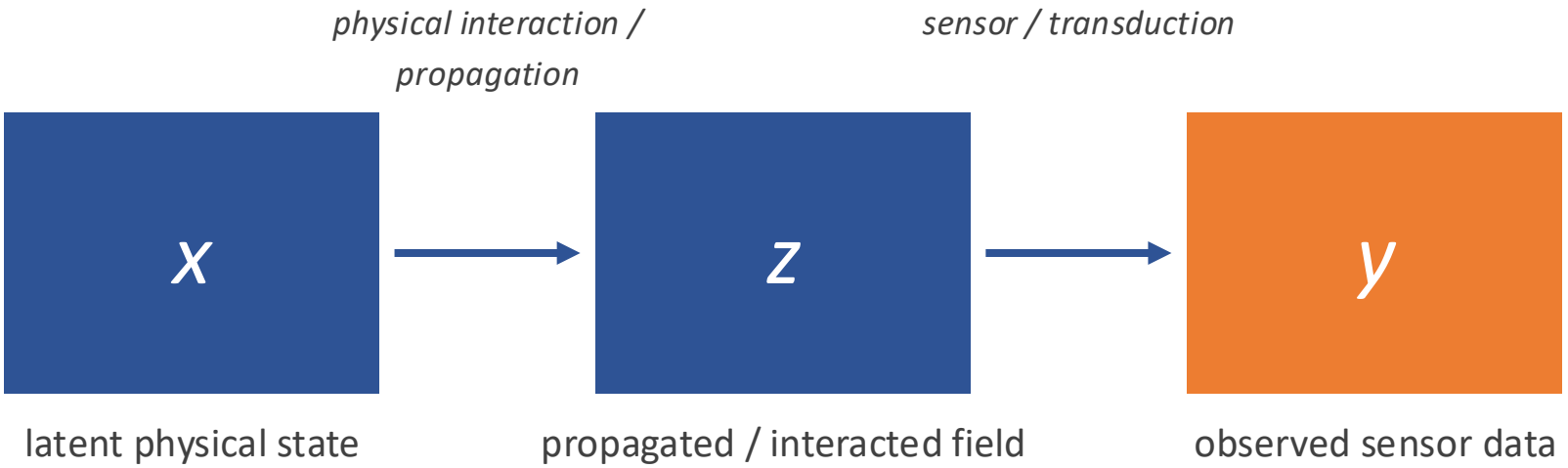
The 'photosites' on the sensor only measure the brightness of the light, not its colour. So that colour information is gathered, each photosite has a red, green or blue filter. These are arranged in a mosaic known as a Bayer pattern, after the boffin that came up with it. He found that by using twice as many green filters as blue or red, you got a sharper image. A demosaicing process turns this raw data into the full-colour grid of pixels in the recorded images.

Sensor data is *not* the physical world.

It is the result of a physical process acting on the world.

- A photograph is photons at pixels — *not "flower."*
- A recording is air pressure at a diaphragm — *not "voice."*
- An IQ sample is a reflected echo — *not "car."*

From physical state to sensor data



Every sensing system, however different its physics, follows this same three-stage abstraction. We will map camera, microphone, and radar onto it repeatedly today.

The forward measurement model

$$y = F(x; \theta) + n$$

x	F	θ	y	n
latent physical state	forward / measurement process	environmental or sensor parameters	observed sensor data	measurement noise

This is the equation the whole semester keeps coming back to — for every sensing system we study, we will identify each of these five terms.

Physical State vs. Sensor Observation

What does a sensor actually measure?

A photograph of a flower

What did the camera actually measure?

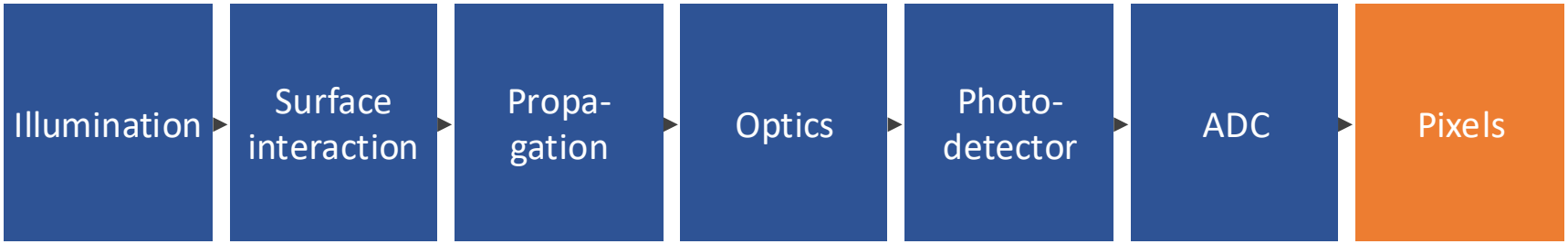
It did NOT measure…

- "Flower" (a semantic category)
- Shape
- Material
- Color as an abstract property
- Distance

It DID measure…

Photons converted into electrical charge at pixels.

The imaging pipeline, physically



Latent scene: $x = \{ \text{geometry, material, illumination, pose} \}$

Observation: $y = \text{pixel intensities}$

The photograph is a highly indirect trace of x — filtered through illumination, surfaces, propagation, and the sensor itself.

cf. Berkeley CS 180 — pinhole camera → pixels → physics of light → image formation → computation.

Three concepts to fix today

1

Latent variable / physical state

A quantity that exists physically but is not directly observed.

2

Observable

What the sensor actually measures.

3

Measurement operator

$y = F(x)$ — the mapping from latent state to measurement.

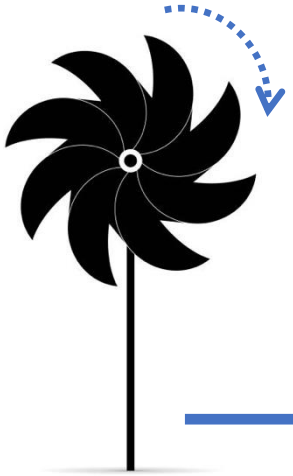
This distinction — latent state vs. observable, linked by an operator — is foundational for inverse problems covered later lectures.

The Universal Sensing Pipeline

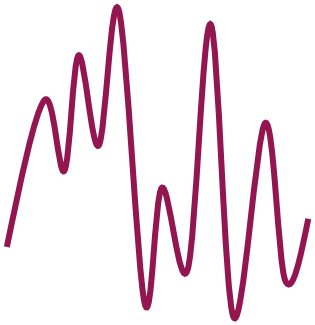
Fundamental principle of sensors

Air flow





Electrical fluctuations proportional to the speed/volume of air-flow



Mapping

10 Miles per hour



A physical phenomenon
We want to measure

SENSOR

A structure or material that reacts to the physical phenomenon

A circuit that converts the reaction to measurable electrical signal (voltage, current, or energy)

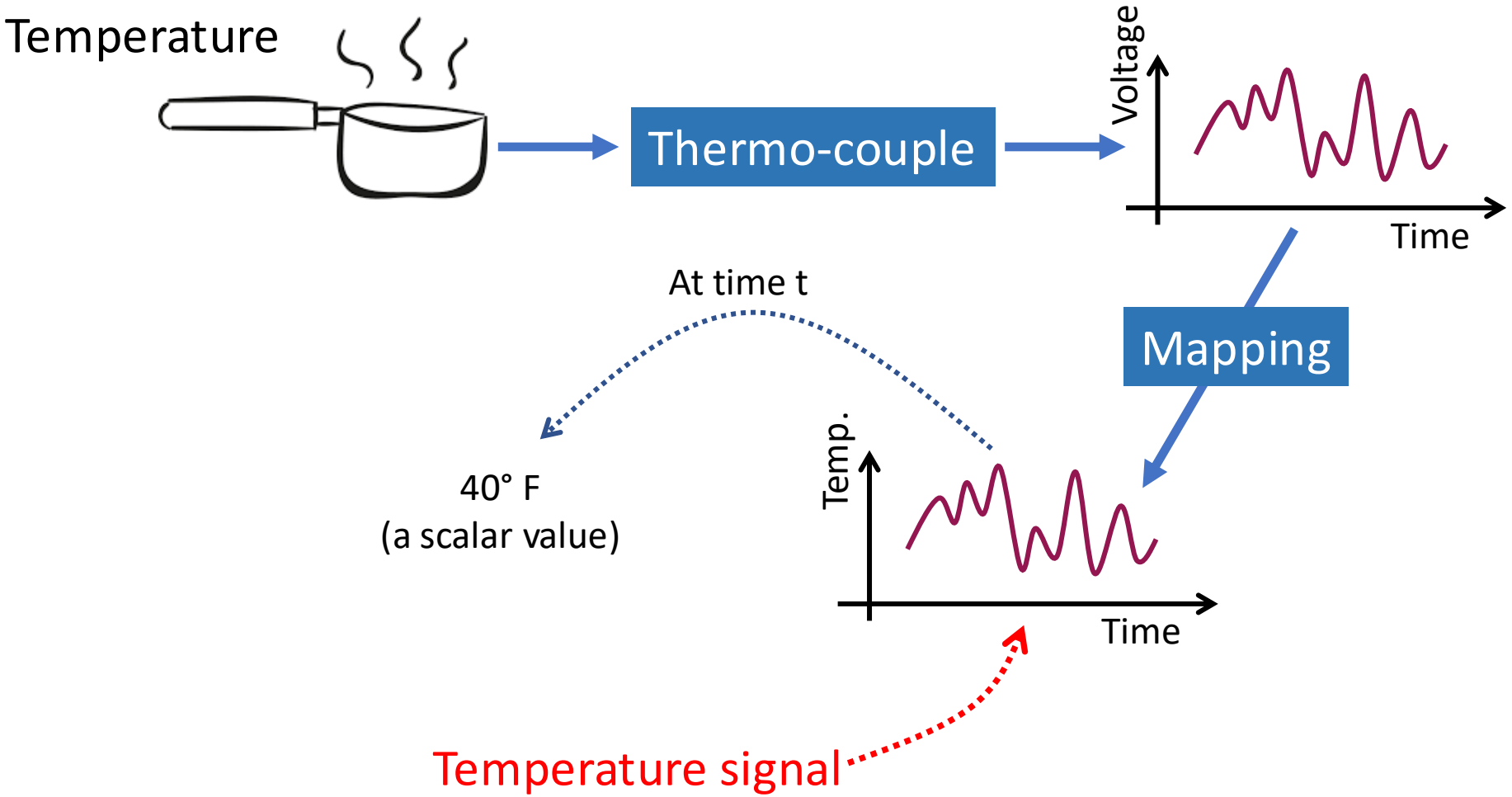
Interpretation of the signal, storing the data

Fundamental principle of sensors

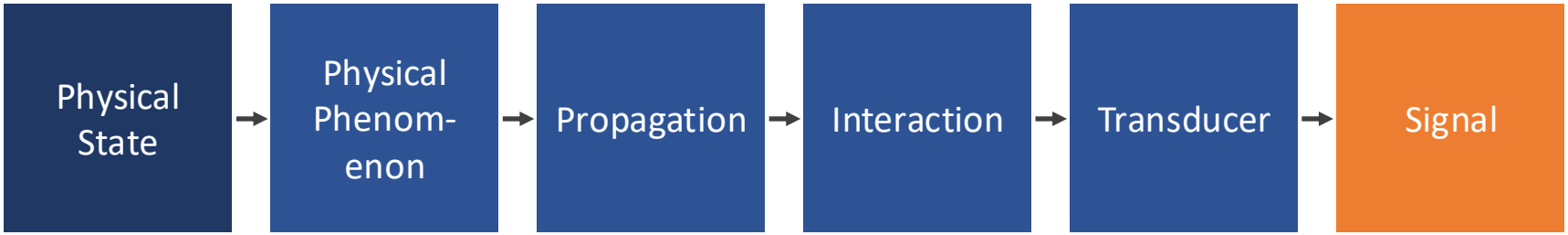
"Sensors are transducers"

A transducer is a device that converts energy from one physical form into another

Fundamental principle of sensors



Generalizing beyond the camera



- Every sensing system — camera, microphone, radar, or otherwise — instantiates this same six-stage chain.
- What differs across modalities is the physics of propagation and interaction, and what counts as the transducer.
- This is the general form of x → z → y from before, expanded one level.

Sensor to ”sense” distance





Passive vs. active sensing

Passive sensing

The sensor observes energy it did not emit.

Camera: illumination → scene → camera

Microphone: source → acoustic field → mic

The sensor has no control over the illumination or the source — it only listens.

Active sensing

The sensor emits energy and measures what returns.

Radar: transmitter → EM propagation → target → EM propagation → receiver

Also: sonar, ultrasound, LiDAR, structured light.

Because we control the transmitted signal, we control part of F itself.

Active sensing → sensing system design

$$y = F(x;\, u)$$

u = a design variable we control: transmitted waveform, illumination pattern, frequency, sensor position, …

An idea worth planting early:

Machine learning may eventually optimize u — the sensing process itself — not merely infer x from fixed measurements.

Three Physical Phenomena, Three Sensing Systems

Light → camera · Sound → microphone · RF → radar

A. Light → camera: from radiance to pixels



Reflection, at the level we need today:

$$I_{measured} \propto (I_{illumination} \times \rho)$$

- Diffuse reflection
- Specular reflection
- Wavelength-dependent reflectance (ρ)

Pixel intensity depends jointly on object, illumination, geometry, and sensor — so two images of the same object need not have the same y.

B. Sound → microphone: from pressure to a waveform

Free field: a speaker creates a pressure variation

$$p(r) \propto 1 / r$$

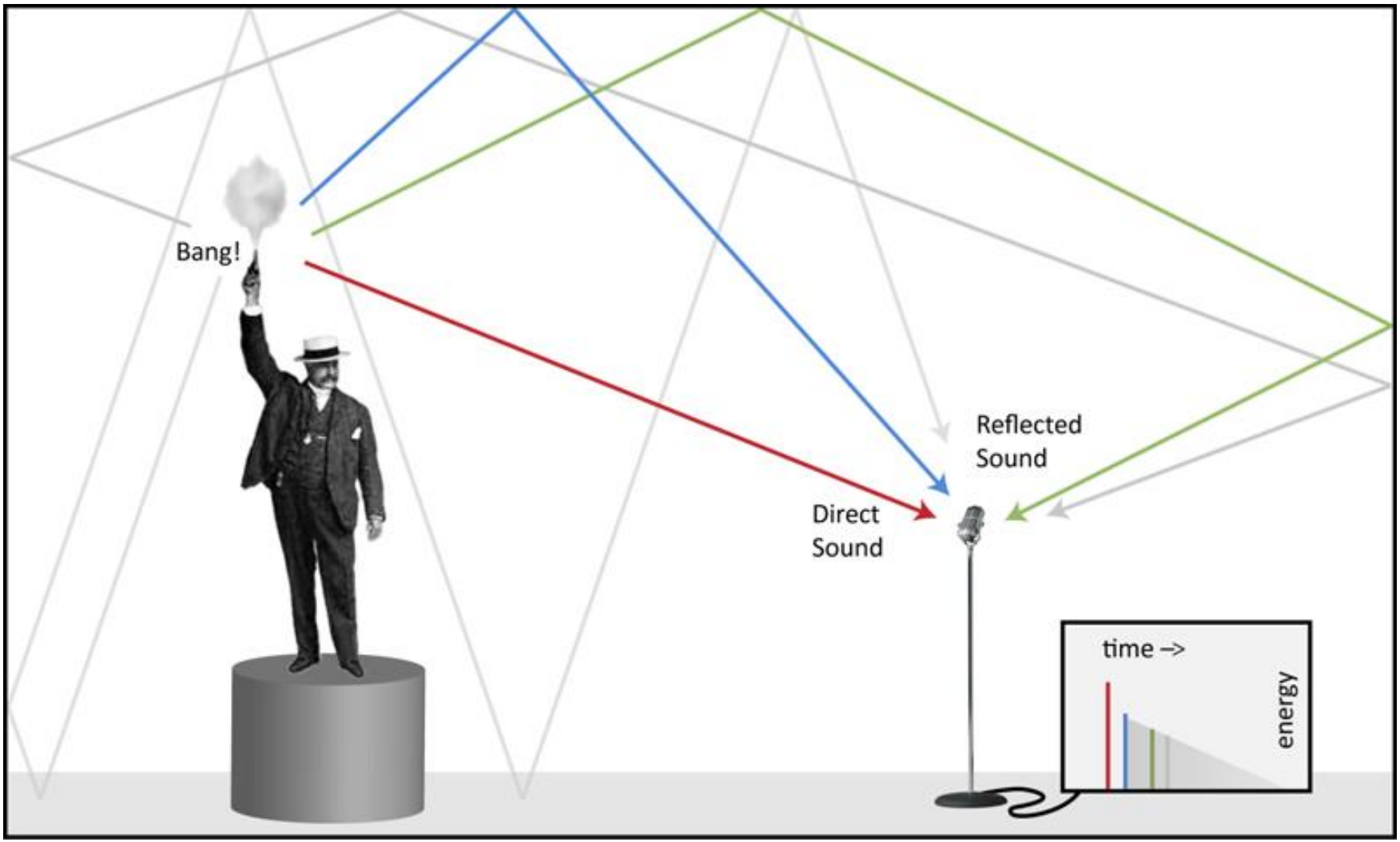
Now put the speaker in a room:

$$y(t) = s(t) * h(t) + n(t)$$

s(t): source h(t): acoustic channel / room y(t): mic signal

The microphone observes the source convolved with the environment: measurement contains information about both source AND environment.

B. Sound → microphone: from pressure to a waveform



One equation, five research directions

Source separation | **Localization** | **Room inference** | **Dereverberation** | **Spatial audio**

$$y(t) = s(t) * h(t) + n(t)$$

All five directions follow from one physical fact: this equation mixes what you want (s) with where you are (h).

C. RF / radar → receiver: distance is encoded in time



For one ideal reflector at distance d:

$$y(t) = \alpha\, s(t - \tau) + n(t)$$

$$\tau = 2d / c$$

Distance

is encoded in delay τ

Velocity and direction, too

If the object moves…

$f_D \propto V$

The reflected frequency shifts — the Doppler effect. Velocity is encoded in frequency.

With multiple antennas…

Location / direction becomes encoded in the phase differences between receivers.

Physical properties are transformed by propagation into measurable signal properties.

The Forward Measurement Model

Assembling x, ϑ, F, and n

Forward Measurement Model

$$y = F(x) \longrightarrow y = F(x;\theta) \longrightarrow y = F(x;\theta) + n$$

A deterministic map from state to measurement

+ environment / sensor parameters

+ measurement noise — the full model

Each addition makes the model more physically honest — and each will matter for how we later estimate x.

x: the latent physical state

As a vector of physical quantities

$x = [\,\text{position, velocity, orientation}\,]$

e.g., the state of a target tracked by radar, or a rigid body sensed by an IMU.

As a spatial field

$x(r) = \text{reflectivity at } r$

e.g., a scene's reflectivity map, or a room's acoustic impulse response.

---

x can be a finite-dimensional vector or a continuous field — the abstraction covers both.

What is a forward model?

$$y = F(x; \theta) + n$$

A model of the physical process that maps the state of the world to the measurements produced by a sensor.

F maps the latent physical state x into the expected sensor measurement y, given system/environment parameters θ.

Notice what F does not have to be: it does not need to be the physical state itself, or the measurement itself — it is the mapping between them.

A forward model doesn't have to be exact

F can take many forms:

- Derived from physics
- Simulated numerically
- Approximated
- Partly learned from data
- Completely learned

The test:

"If I gave you the true physical state, could your model predict what the sensor would measure?" If yes — you have a forward model.

The forward question vs. the inverse question

The forward question

"If I know where the target is, what signal should the radar receive?"

The inverse question

"Given the received signal, where is the target?"

A concrete instance — a target at distance d:

$$y(t) = \alpha\, s(t - 2\,x/c) + n(t)$$

$$x \rightarrow y$$

forward model

$$y \rightarrow \hat{x}$$

inverse problem

The forward model, across sensing systems

In every system, the forward model predicts the measurement from the physical state.

System	Physical state x	Forward model predicts
Camera	3D scene, materials, lighting	Pixel intensities
Microphone	Sound source + room	Recorded waveform
Radar	Target position / velocity	RF echo / IQ samples
LiDAR	Surface geometry	Return times / intensities
IMU	Body motion	Acceleration and gyro readings
CT	Internal attenuation map	X-ray projections

θ: what does it actually mean?

$$y = F(x; \theta) + n$$

Parameters describing the sensing system, environment, and propagation conditions that determine the forward process.

θ is not what we are trying to infer — it is what the forward model needs to know to relate x to y.

ϑ is a parameter of the model, not a label — some ϑ are known and chosen, others are unknown and must be estimated alongside x.

θ across five sensing systems

System	x — quantity of interest	θ — forward-model parameters
Camera	Object geometry / pose	Illumination, camera pose, focal length, exposure
Microphone array	Source location	Room geometry/materials, mic locations, speed of sound
Radar	Target position / velocity	Antenna geometry, carrier frequency, transmitted waveform
WiFi localization	User location	AP position, room geometry, furniture, antenna response
IMU	Motion / trajectory	Sensor bias, scale factor, orientation, gravity

The x / θ boundary depends on the task, not the quantity

Task: speaker localization

X = speaker location

θ = room geometry / materials

Task: room reconstruction

X = room geometry / materials

θ = everything else (source signal, mic array geometry, …)

The same physical quantity — room geometry — is ϑ in one task and x in the other. The split is set by what you are trying to infer, not by the physics itself.

F vs. θ: model structure vs. instantiation

F specifies the structure of the forward model; θ specifies the parameters that instantiate it for a particular system or environment.

$$y(t) = (1/r)\ s(t - r/c) + n(t)$$

$x = r$ (distance) $\quad \theta = c$ (propagation speed)

For a richer room-acoustic model:

θ = { room geometry, wall reflection coefficients, microphone positions, c, … }

F	θ	x
model structure / physical laws	parameters that instantiate that model	physical state we want to infer

Why not just call θ "nuisance parameters"?

Some θ are known, deliberately chosen design parameters:

- Antenna spacing
- Radar carrier frequency
- Camera focal length

Others are genuinely unknown nuisance parameters:

- Room materials — and other unmodeled environmental conditions

---

θ mixes both known design choices and unknown nuisances — so "nuisance parameters" is too narrow a name for what it actually captures.

F and n: process and uncertainty

F: the forward model

The known — or partially known — physical process transforming x into measurements. Sometimes physics-based, sometimes learned, often both.

n: measurement noise

The measurement uncertainty not captured by the deterministic part of the model — sensor electronics, thermal noise, unmodeled effects.

From a deterministic map to a probabilistic model

Once we admit noise, y is no longer a deterministic function of x — it is a random variable given x.

$$p(y \mid x)$$

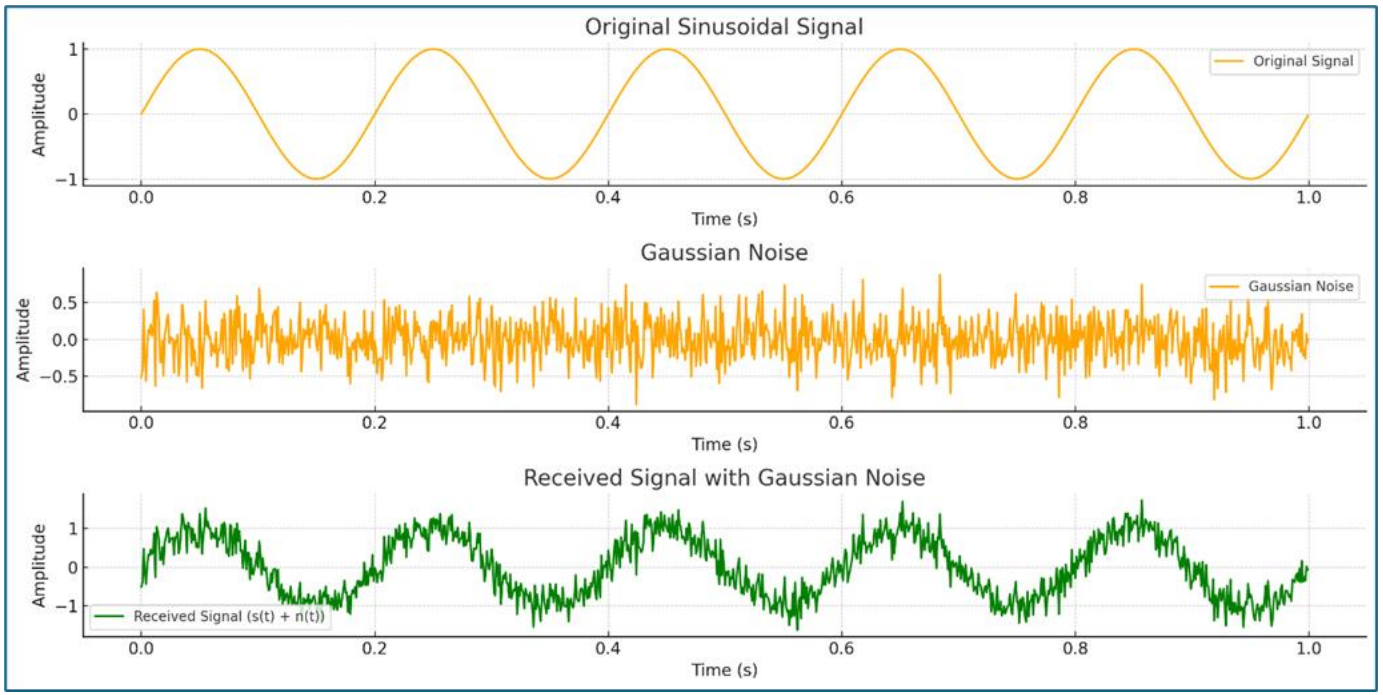
the probabilistic measurement model

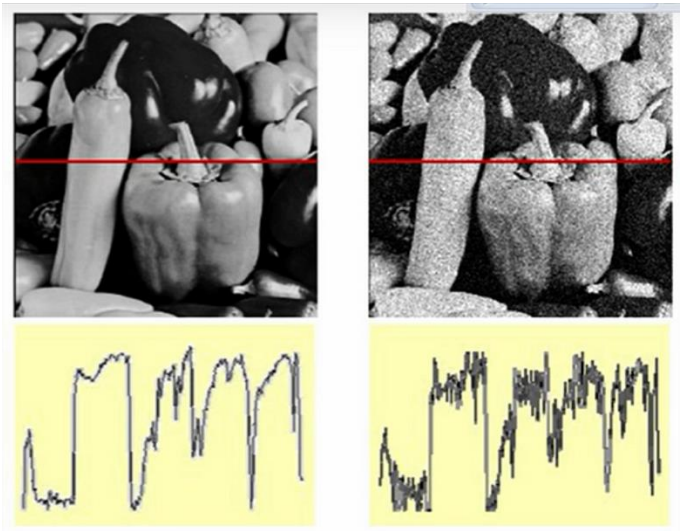
Bayesian inference (later lectures) are built on this model.

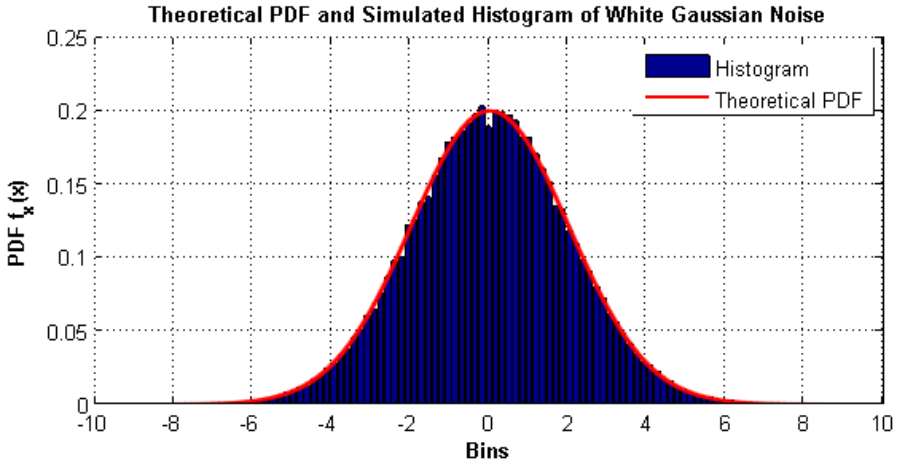
Deterministic Model vs. Stochastic Measurement

Why y = F(x) is almost never literally true

Gaussian noise







Why y = F(x) is almost never literally true

Instead, we write y = F(x) + n. For Gaussian noise:

$$n \sim \mathcal{N}(0, \sigma^2) \longrightarrow p(y|x) \propto \exp\left[-\|y - F(x)\|^2 / 2\sigma^2\right]$$

- This is the origin of a very common pattern in estimation: under Gaussian measurement noise, maximizing p(y|x) becomes minimizing squared measurement error.
- We will not derive it in detail today — just recognize the shape when it returns in Lecture 3.

Why y = F(x) is almost never literally true

$$n \sim \mathrm{N}(0, \sigma^2 I).$$

$$y = \mathrm{F}(x) + n$$

$$\boxed{y \mid x \sim \mathrm{N}(\mathrm{F}(x), \sigma^2 I)}$$

$$p(y \mid x) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{\|y - \mathrm{F}(x)\|_2^2}{2\sigma^2}\right).$$

$$\boxed{\|y - \mathrm{F}(x)\|_2^2}.$$

Why does least-squares appear everywhere?

$$\hat{x}_{\mathrm{ML}} = \arg\max_x p(y \mid x).$$

$$\hat{x}_{\mathrm{ML}} = \arg\max_x \exp\left[-\frac{\|y - \mathrm{F}(x)\|_2^2}{2\sigma^2}\right].$$

$$\boxed{\hat{x} = \arg\min_x \|y - \mathrm{F}(x)\|_2^2}$$

Why does least-squares appear everywhere?

$$\text{argmax}_x \; p(y \mid x)$$

$$\equiv$$

$$\text{argmin}_x \; \|y - F(x)\|^2$$

Because under Gaussian measurement noise, maximizing the likelihood of y given x is the same optimization as minimizing measurement error. That is why least-squares estimation appears everywhere.

What Information Is Lost During Sensing?

From forward models to inverse problems

What information is lost during sensing?

A sensor does not capture the complete physical state of the world.

The forward process transforms, mixes, compresses, and often irreversibly discards information before producing the measurement.

$$y = F(x; \theta) + n$$

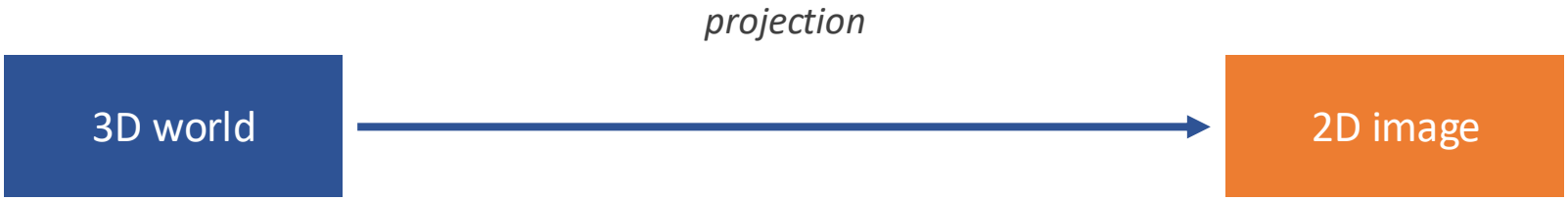
even with no noise:

$$y = F(x; \theta)$$

We generally cannot uniquely recover x from y.

1. Dimensionality reduction

The physical world often has many more degrees of freedom than the measurement — the clearest example is a camera.



A pixel tells us where something projects onto the image plane, but not its depth.

$$u = fX/Z \qquad v = fY/Z$$

Many 3D points can produce the same (u, v) — e.g. (X, Y, Z) and (2X, 2Y, 2Z) project to the same image location.

Depth information has been partially lost through projection.

2. Multiple causes, one measurement

Suppose a camera measures a dark pixel. Why is it dark?

- Dark material
- Weak illumination
- Surface orientation
- Shadow

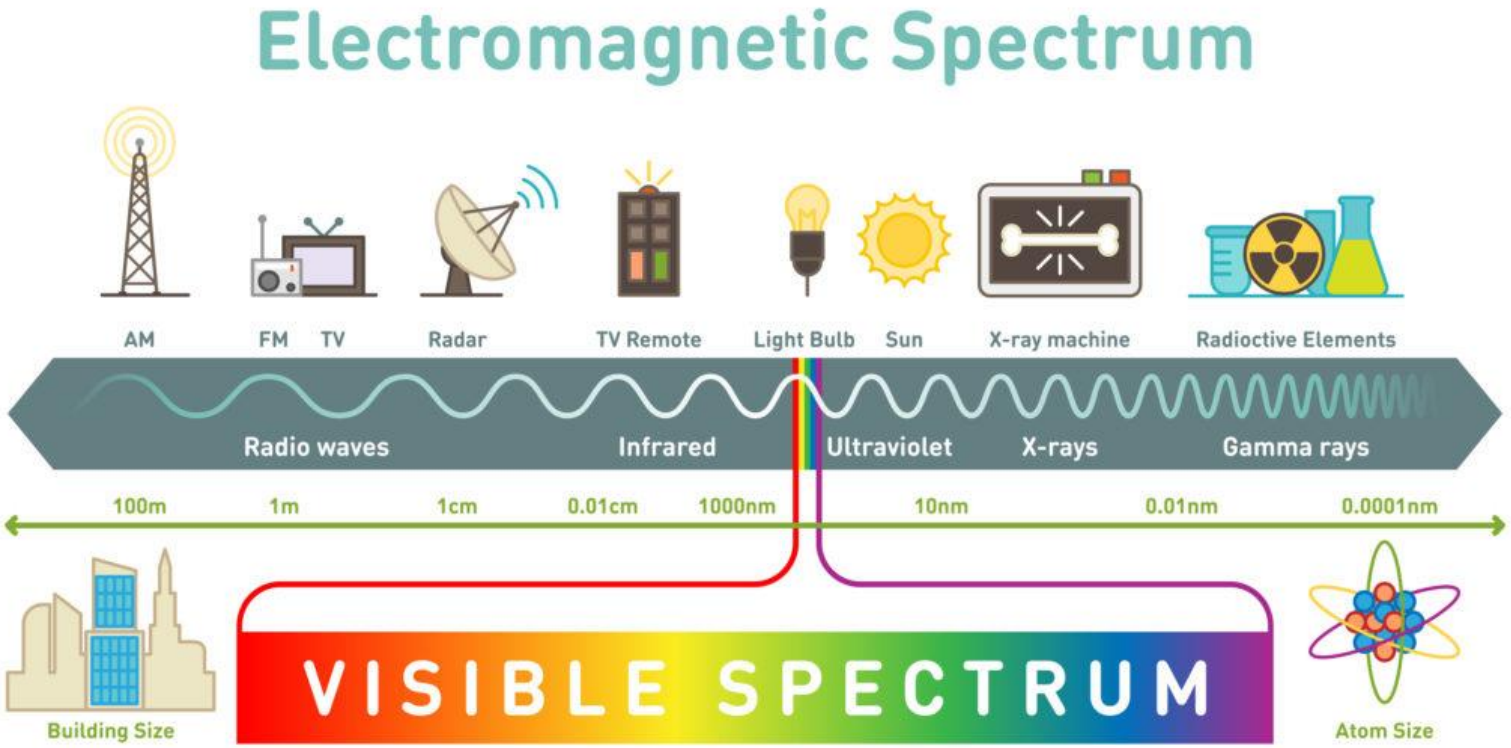
$$y \propto \text{illumination} \times \text{reflectance}$$

Illumination is governed by ϑ; reflectance depends on x (and sometimes ϑ too).

$$x_1 \neq x_2 \quad \text{but} \quad F(x_1; \theta_1) = F(x_2; \theta_2)$$

Different combinations of physical quantities can give the same y — the fundamental many-to-one nature of many forward models.

3. The sensor is an information bottleneck



3. The sensor is an information bottleneck

Every sensor has limited sensitivity.

$$L(\lambda) \longrightarrow (R, G, B)$$

A rich electromagnetic spectrum is reduced to approximately three channels.

Two different spectral distributions can therefore produce the same RGB values.

Similarly, a microphone measures pressure variation over a limited frequency range — it does not capture every property of the acoustic field.

The sensor itself acts as an information bottleneck.

4. Finite resolution destroys detail

If two objects are closer than the radar's range resolution, their reflections may merge into one return.

$$\text{continuous scene} \rightarrow \text{finite pixels}$$

$$x(t) \rightarrow x[n]$$

Fine-scale spatial or temporal information can disappear.

PREVIEW — LECTURE 2

Sampling can destroy information.

PREVIEW — LECTURE 6

Measurement geometry determines resolution.

5. Occlusion: some states are never measured

A camera observes the front surface of a box, but not the back.

$$\text{complete 3D object} \rightarrow \text{visible-surface projection}$$

No amount of processing of a single image can directly recover measurements that were never acquired. Any reconstruction of the back must come from assumptions or prior knowledge.

The same issue occurs in radar shadowing, ultrasound, LiDAR, and many other sensing systems.

6. Mixing: contributions become entangled

$$y(t) = s(t) * h(t) + n(t)$$

A microphone measurement contains both the source s(t) and the environment h(t) — they are not separately observed.

Likewise, wireless CSI may contain the superposition of many propagation paths:

$$H(f) = \Sigma_k \ \alpha_k \exp(-j2\pi f \tau_k)$$

The receiver measures the combined channel, not necessarily each individual reflector/path.

Information may not be completely absent — it may be entangled.

lost information ≠ mixed/latent information — a sophisticated inverse method can sometimes disentangle the latter.

7. Noise limits distinguishability

$$y = F(x; \theta) + n$$

Suppose two physical states produce slightly different ideal measurements:

$$F(x_1) \approx F(x_2)$$

If the difference is smaller than the measurement noise, the sensor may not be able to distinguish them.

Even a theoretically invertible mapping may be practically unrecoverable.

Three concepts this motivates

Ambiguity

Can different physical states explain the same measurement?

Identifiability

Can the physical state of interest be uniquely determined from the available measurements?

Observability

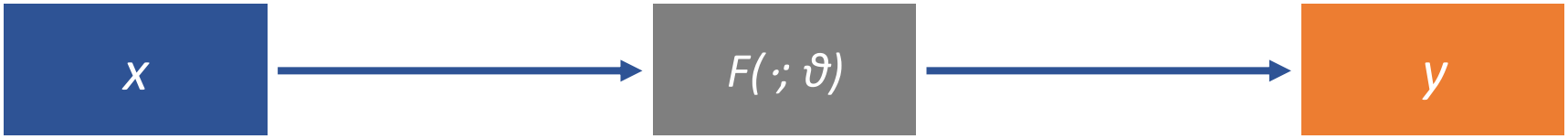
Do the measurements contain enough information about the state to recover it?

Especially important for dynamical systems, later in the course.

$$x_1 \neq x_2, \quad F(x_1) \approx F(x_2)$$

For Lecture 1 we emphasize ambiguity and identifiability, and only briefly mention observability — formal observability fits better with state estimation.

Where does the lost information go?



3D → 2D	**projection loss**
continuous → sampled	**resolution loss**
many wavelengths → RGB	**spectral loss**
multiple paths → one signal	**mixing**
hidden surfaces → no measurement	**occlusion**
ideal signal → noisy signal	**uncertainty**

If the forward process loses information, how can we recover the physical state?

What information was lost?

The sensor gives us *y*. We actually want *x*.

How do we get from y back to x?

The bridge to inverse problems

Measurements + Physics + Priors + Learning

$\longrightarrow$ Physical inference

Much of physical sensing ML exists because the measurement alone is insufficient.

We use physical structure, multiple measurements, priors, and learned regularities to resolve what the forward process made ambiguous.

Eight concepts to leave with

- Latent physical state x
- Observable measurement y
- Forward / measurement operator F
- Physical propagation & interaction
- Transduction
- Measurement noise & uncertainty
- Nuisance / environmental parameters θ
- Identifiability / ambiguity, intuitively

The target: comfortably interpret $y = F(x; \theta) + n$ for any new sensing problem.

In-class exercise: decompose four sensing systems

For each system, fill in: x, physical phenomenon, F, y, ϑ, n.

Camera | **Microphone array** | **Automotive radar** | **Smartphone IMU**

Worked example — Automotive radar

x = { target position, velocity, reflectivity }

F = { EM propagation + reflection + antenna + receiver }

y = { complex IQ samples }

θ = { antenna geometry, carrier frequency, environment }

If you can perform this decomposition for an unfamiliar sensing system, Lecture 1 has succeeded.

A recurring analytical language

1. What is x?
2. What is y?
3. What is F?
4. Which parts of F are known? Which are learned?
5. What information about x actually survives into y?

The forward model, decomposed across eight sensing systems

For every system: $y = F(x; \vartheta) + n$ — the same five terms, instantiated differently.

Sensor / System	x Latent physical state	F Forward process	y Measurement	θ Nuisance parameters	Information encoded	Ambiguity
Camera	3D position, geometry, pose, material/appearance	Illumination → surface interaction/reflection → propagation → optics → image sensor	Pixel intensities / RGB image	Illumination, camera pose, focal length, exposure, reflectance, background	Projection, texture, edges, apparent size, motion, perspective	3D→2D projection loses depth; different geometry/material/illumination combinations can yield similar images
Microphone Array	Source position, source signal	Sound generation → acoustic propagation → reflections/reverberation → microphones	Multichannel audio waveforms	Room geometry, wall materials, microphone geometry, speed of sound, microphone response	Inter-microphone delay, phase difference, amplitude, reverberation structure	Multipath can mimic direct paths; symmetric locations or poor array geometry can produce similar measurements
Automotive Radar	Target range, velocity, angle, reflectivity	RF transmission → propagation → target reflection/scattering → return propagation → antenna/receiver	Complex IQ samples; range-Doppler data	Carrier frequency, waveform, antenna geometry, environment, target reflectivity	Delay → range; Doppler → velocity; inter-antenna phase → angle; amplitude → reflectivity	Multiple targets may overlap; multipath can create ghost targets; finite bandwidth/aperture limits resolution
Smartphone IMU	Position, velocity, orientation	Physical motion → inertial forces/rotation → MEMS transduction → electrical signal	3-axis acceleration and angular velocity	Sensor orientation, gravity, scale factors, bias, temperature, calibration	Translational acceleration, angular motion, short-term dynamics	Absolute position is not directly observed; bias/noise integration causes severe drift
LiDAR	3D surface geometry, reflectivity	Laser emission → propagation → surface reflection → return propagation → photodetector	Return time, intensity, point cloud	Sensor pose, beam direction, surface reflectivity, atmosphere	Delay → depth; scan angle + range → 3D location; intensity → reflectivity	Occlusion; weak returns from dark surfaces; glass, fog, and multiple reflections can confuse measurements
WiFi CSI Localization	Device/person location, orientation, environment state	RF transmission → multipath propagation/reflection/diffraction → receiving antenna	Complex CSI across antennas/subcarriers	AP/receiver geometry, room layout, walls, furniture, people, frequency, antenna response	Amplitude, phase, multipath delay structure, frequency-selective fading	Different locations can yield similar CSI; environmental changes can alter CSI without changing location
Medical Ultrasound	Tissue geometry/properties, acoustic impedance/scattering distribution	Ultrasound transmission → propagation through tissue → reflection/scattering → return propagation → transducer	Echo RF waveforms across transducer elements	Speed of sound, attenuation, transducer geometry, transmitted pulse	Delay → depth; amplitude → tissue/interface properties; array differences → direction	Multiple scattering, variable sound speed, and similar tissue responses can create ambiguous reconstructions
Thermal Camera	Surface temperature, emissivity, geometry	Thermal emission → atmospheric propagation → IR optics → detector	Infrared intensity image	Emissivity, ambient temperature, atmospheric absorption, camera calibration	Thermal radiation intensity related to temperature/material	Temperature and emissivity are entangled; different combinations can produce similar measured intensity

Questions?

Next: Lecture 2 — Signal Representations