STEVENS
INSTITUTE OF TECHNOLOGY
1870

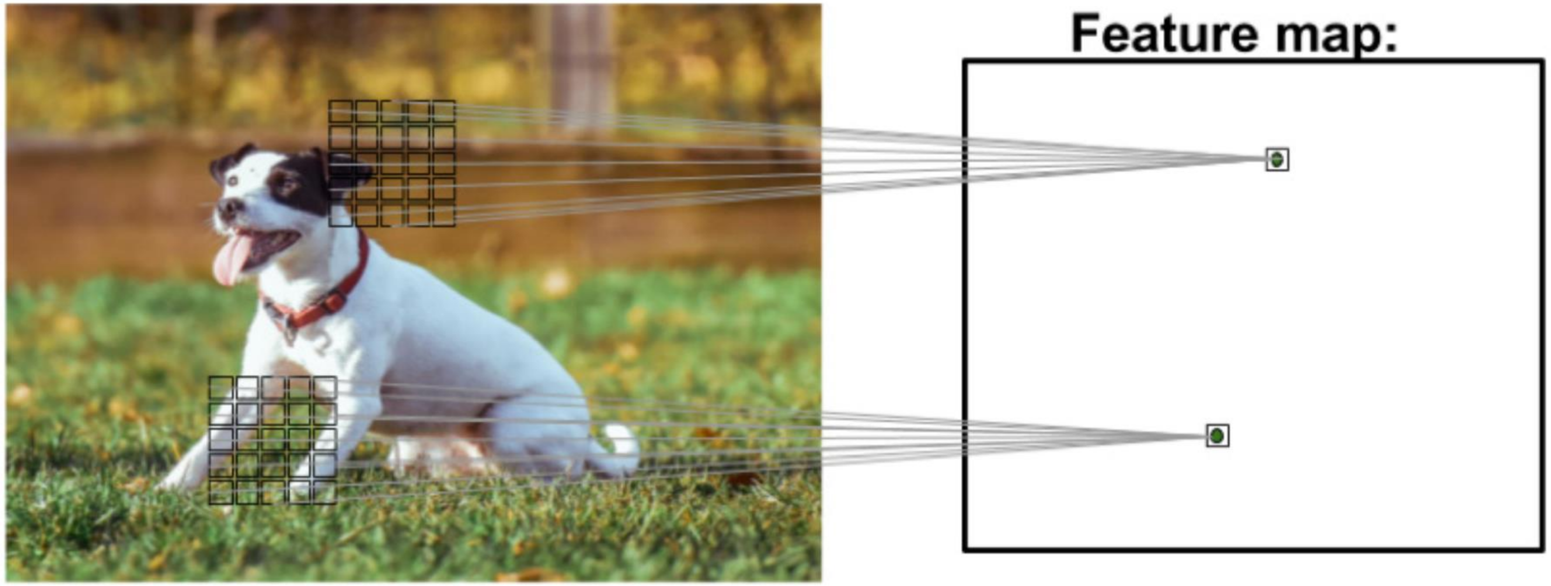
PEP 559 Machine Learning in Quantum Physics

Dr. Chunlei Qu

Spring 2025

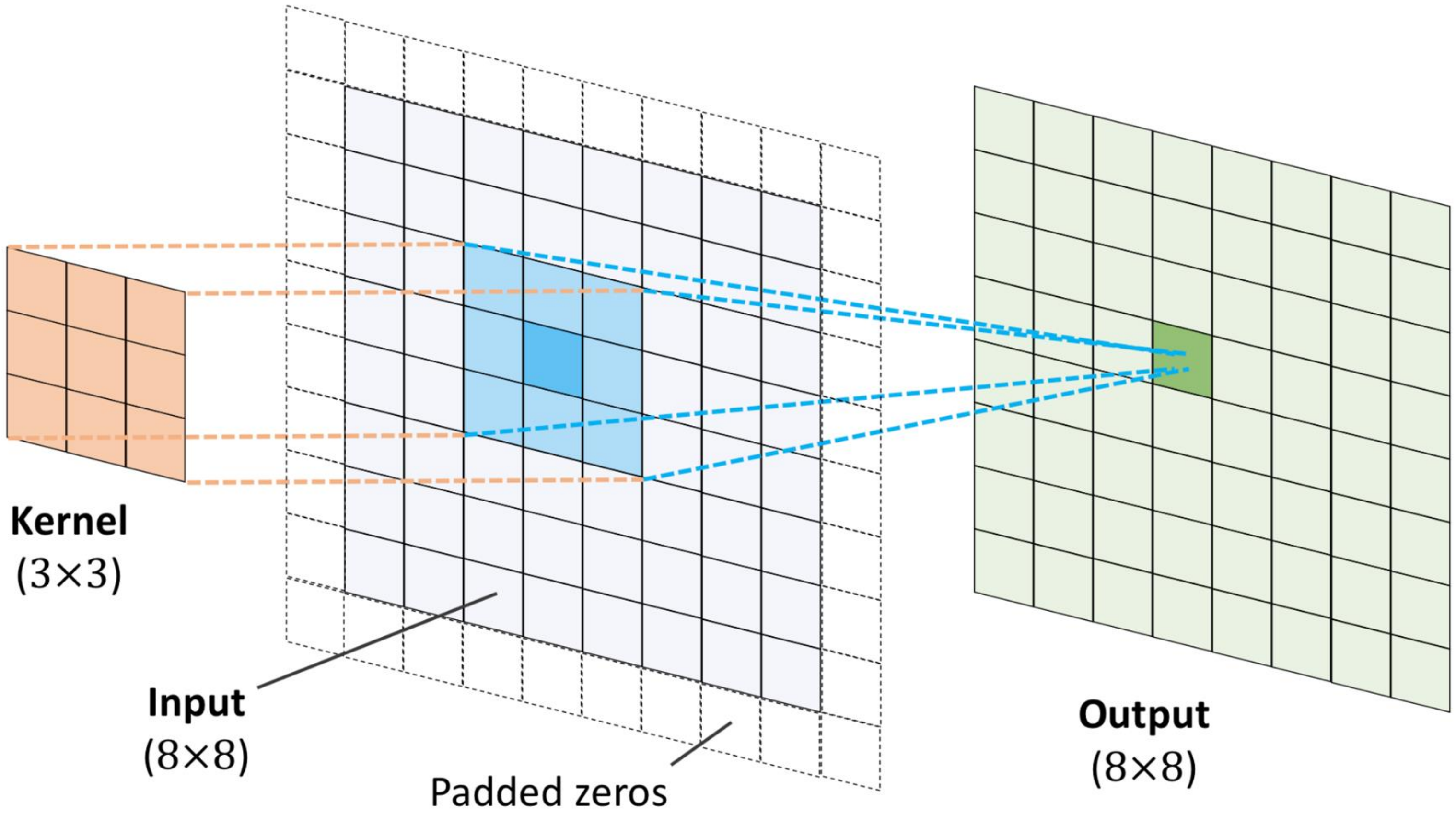
Convolutional Neural Networks

Kernel → Feature map

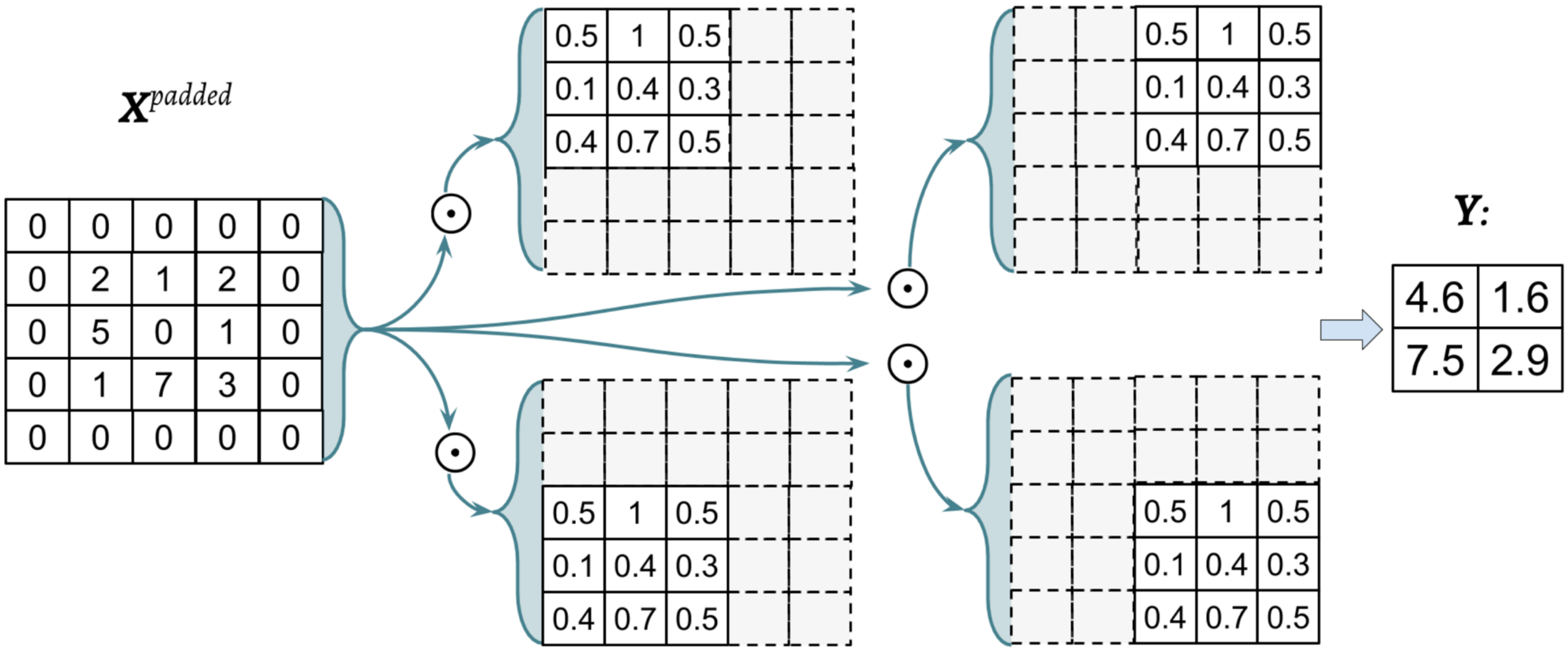


Using a smaller kernel matrix to extract features

Padding



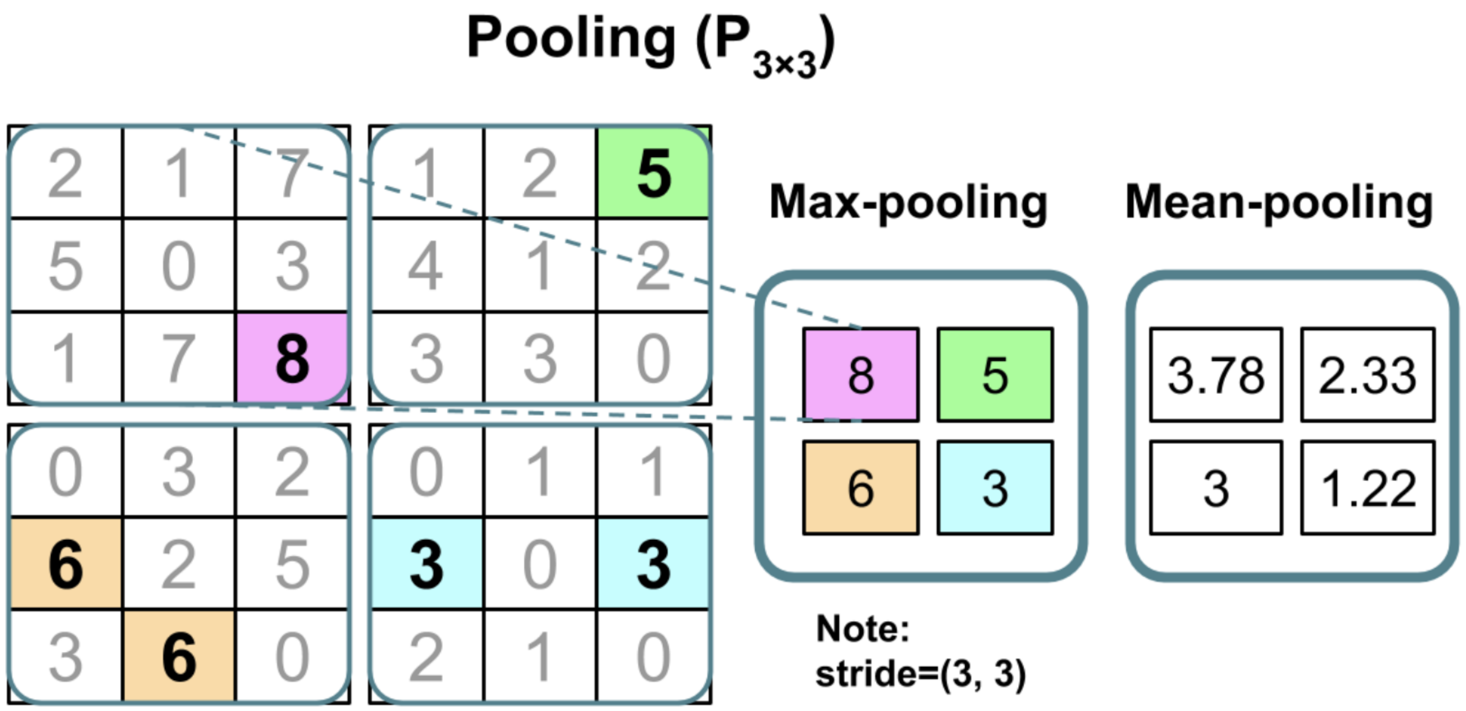
Example of Convolution



Padding p=(1, 1)

Stride s=(2, 2)

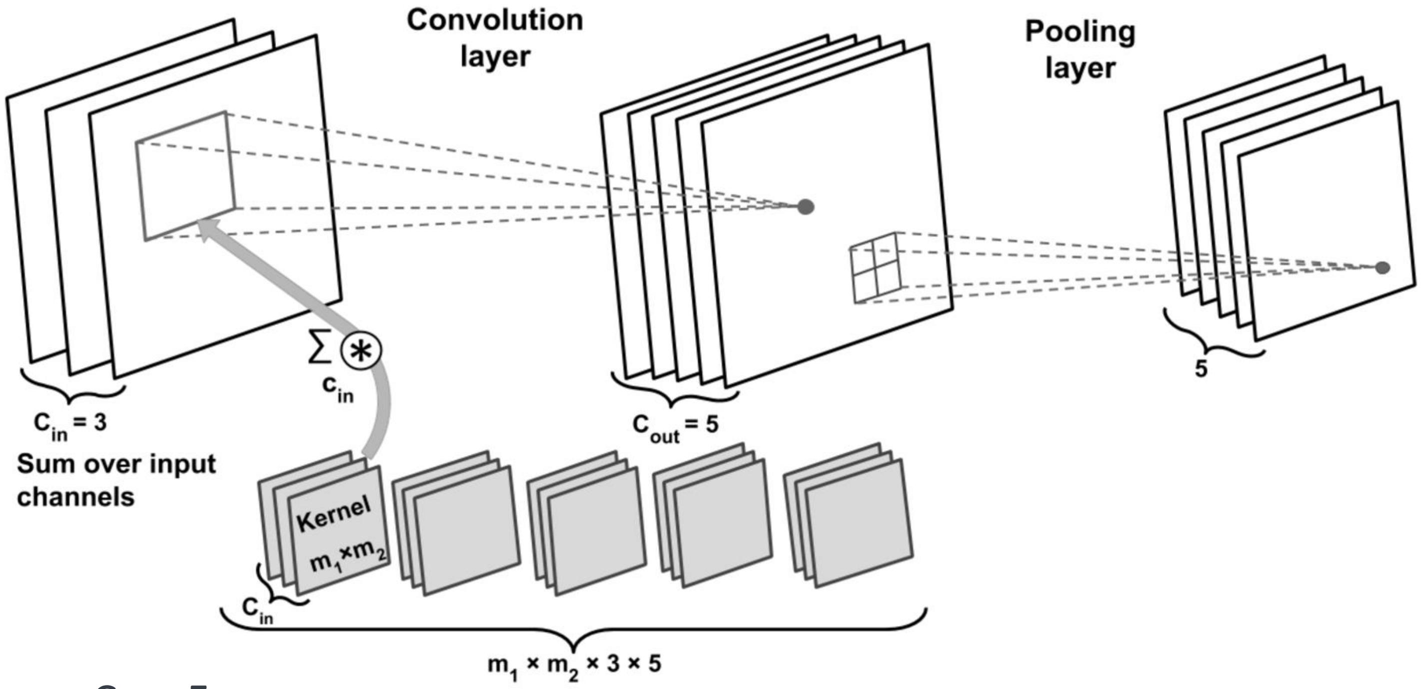
Max-pooling vs. Mean-pooling



Demo of convolution

See Notebook

Implementing CNN



1. Convolution layer

Input channels C_{in}=3
Output channels or feature maps C_{out}=5

Kernel: m_1xm_2

For each input and output channels, there is one such small kernel matrix. Thus, a **four-dimensional tensor**

2. Pooling layer

Training parameters

CNN is efficient because of parameter sharing and sparse connectivity

There are **m_1 x m_2 x 3 x 5** parameters associated with the kernel tensor

There is a bias for each output feature map of the convolutional layer.

The pooling layer does not have any trainable parameters;

In total, **m_1 x m_2 x 3 x 5 + 5**

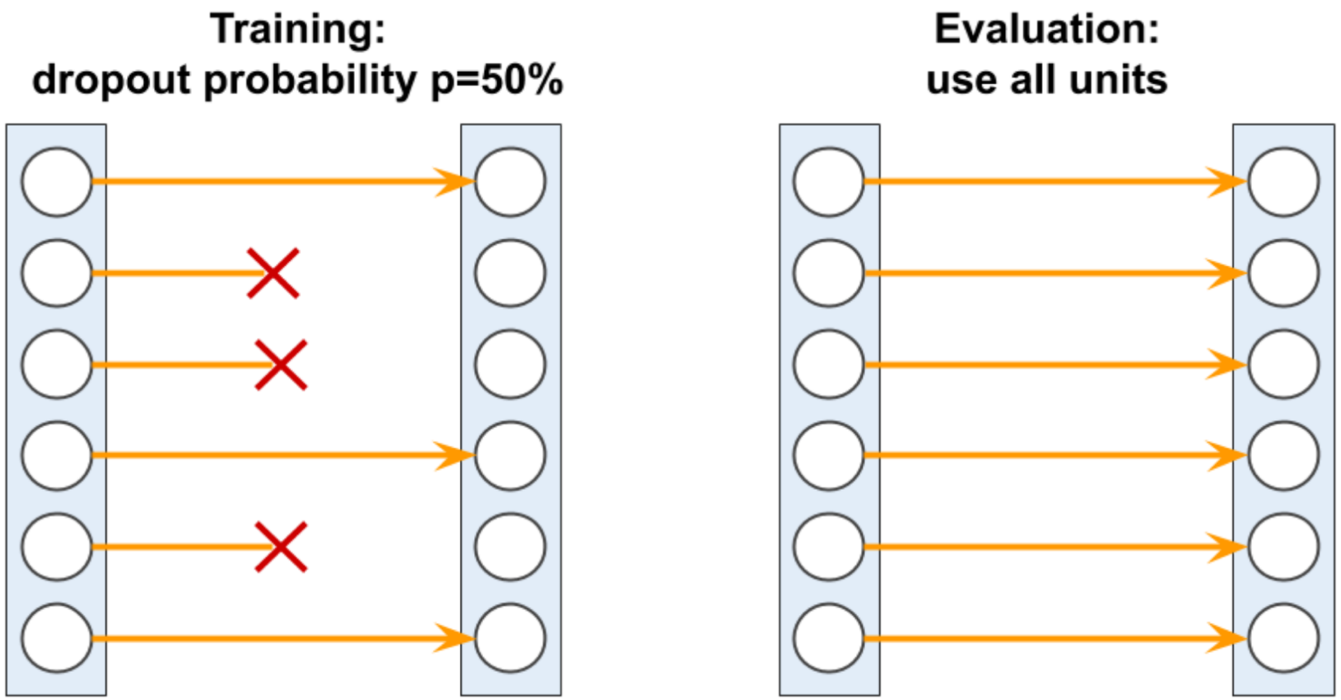
Input image size n_1 x n_2 x 3 (channels), output feature map size (if padding properly), n_1 x n_2 x 5.

If we use fully-connected layer instead of CNN, the total number will be

$$(n_1 \times n_2 \times 3) \times (n_1 \times n_2 \times 5) = (n_1 \times n_2)^2 \times 3 \times 5$$

plus bias

Dropout during training

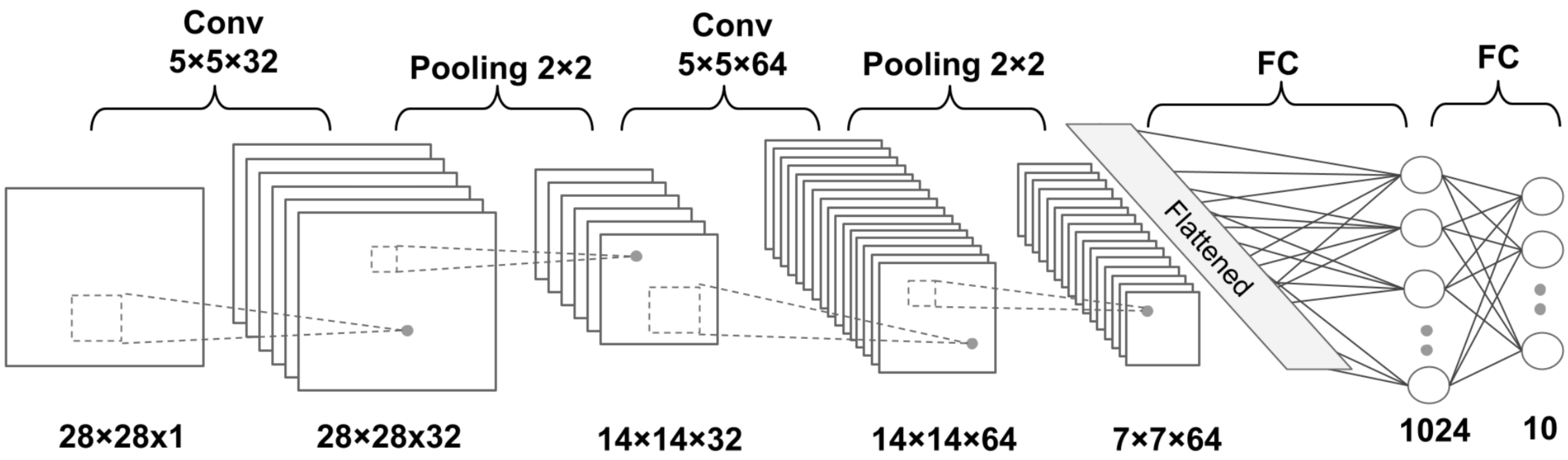


A popular technique for regularizing NNs to avoid overfitting, thus improving the generalization performance

During the training, a fraction of the hidden units is randomly dropped at every iteration with probability p.

That is, we randomly turn off some hidden units, so that the network learning will not rely on the activation of any set of hidden neurons.

Illustration of a deep CNN



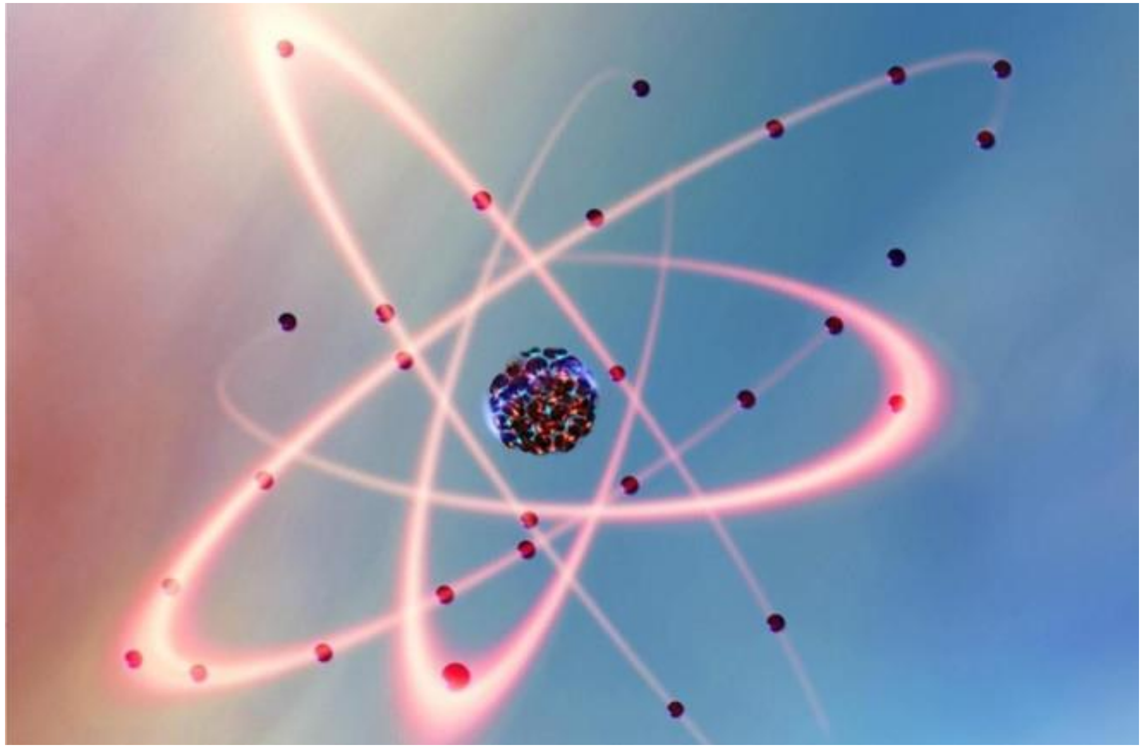
Demo of CNN for MNIST

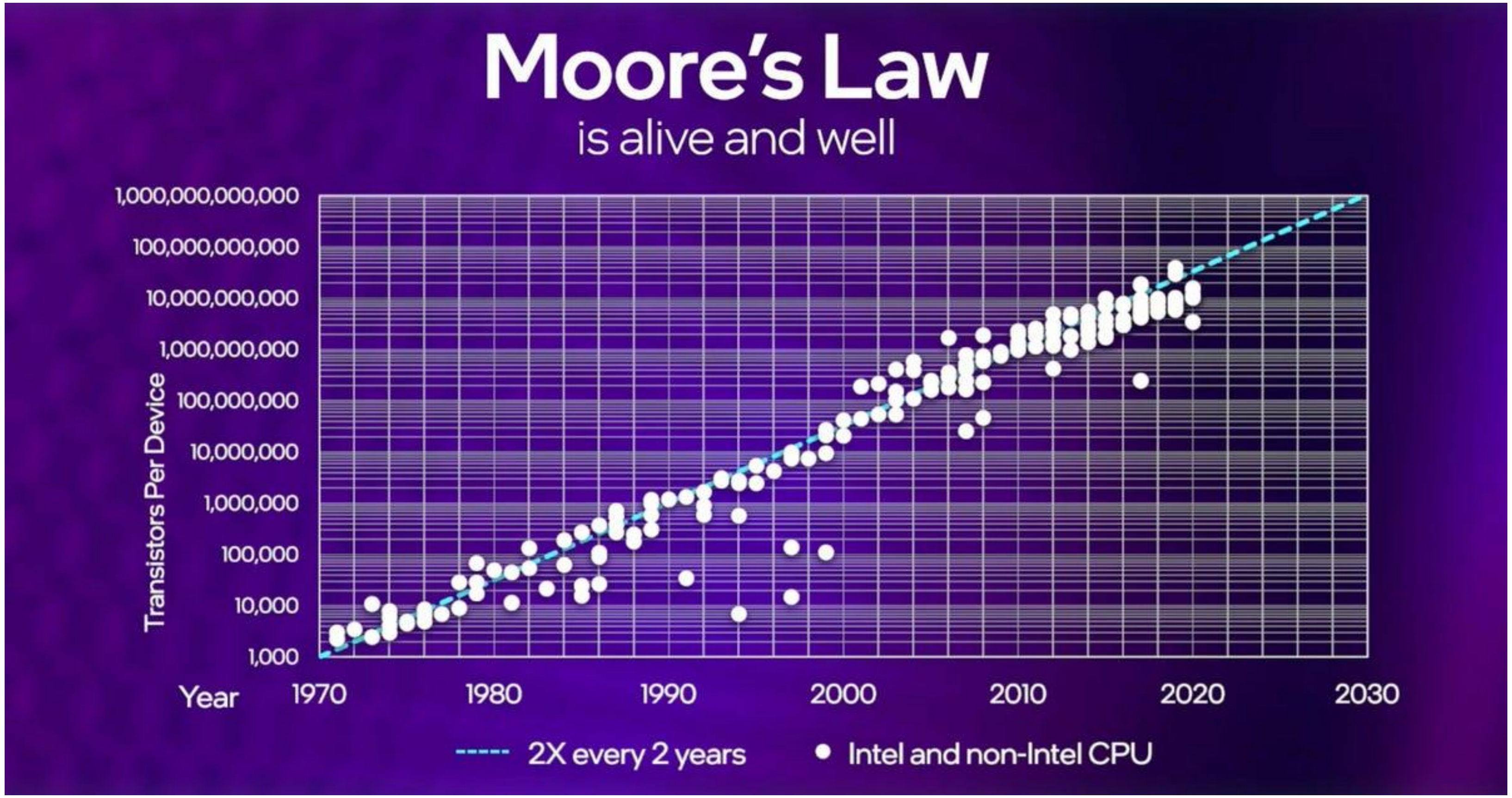
See notebook

Application of ML in quantum physics

Why Quantum?

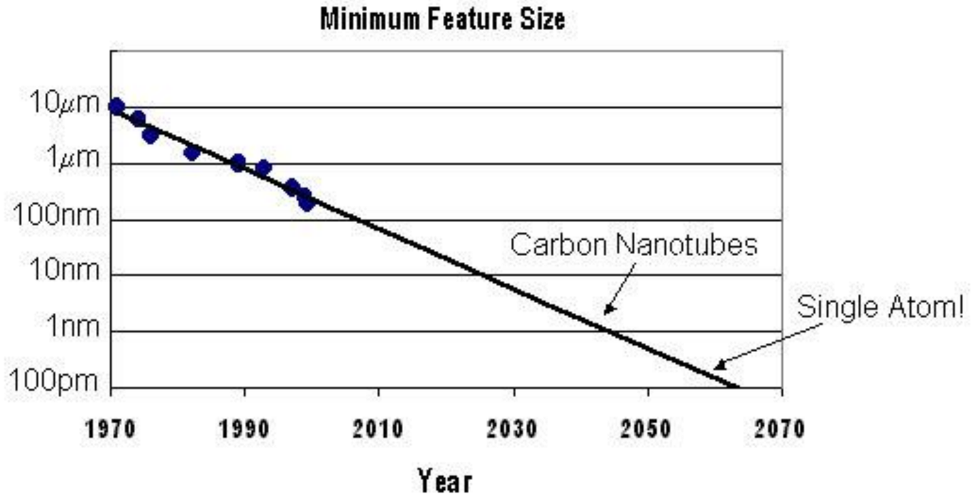
Because **conventional technologies** become incapable to address some **big challenges**

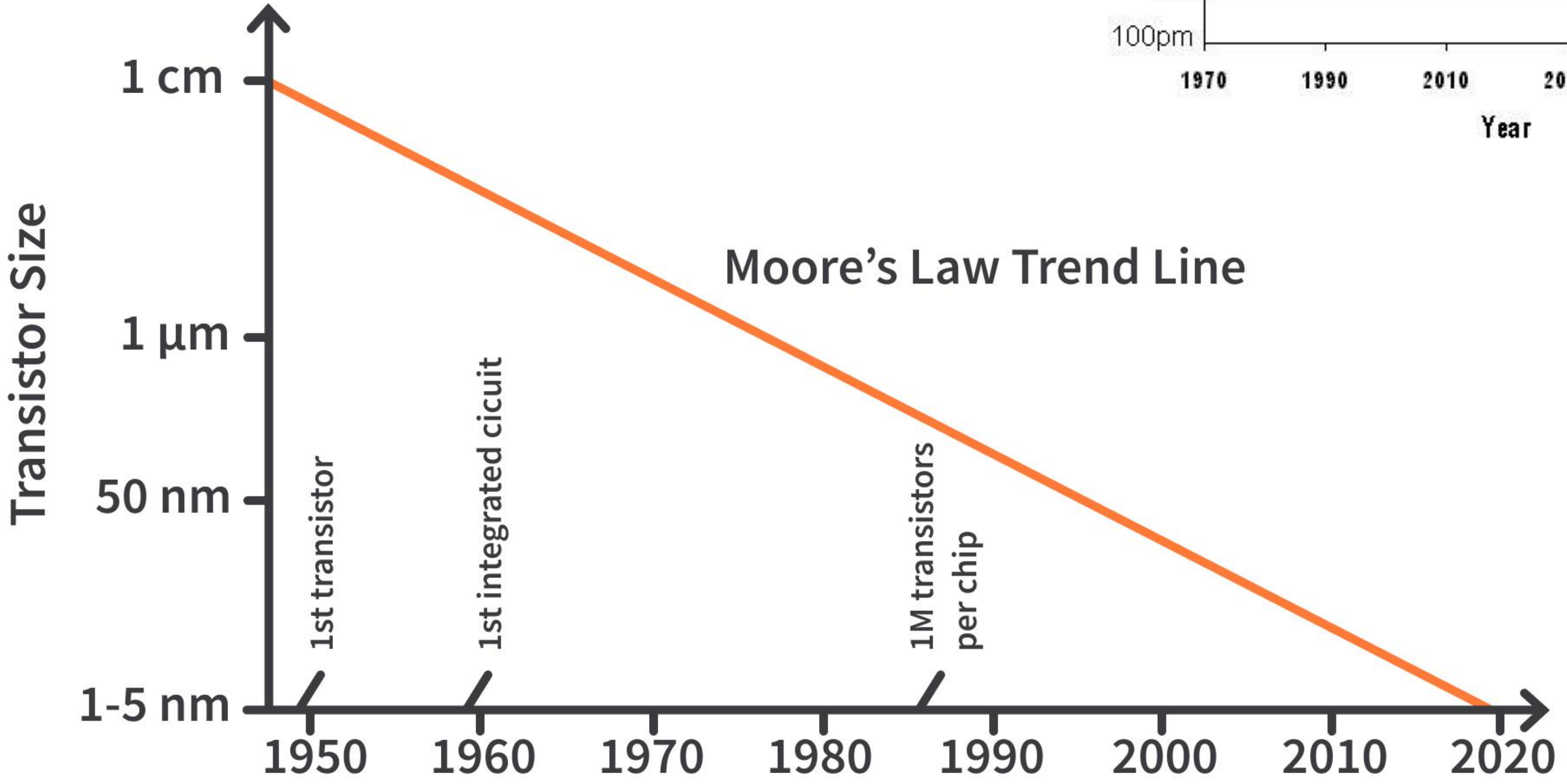




- The number of transistors on a chip doubles every two years
- Will this continue? What happens if the size of the transistor is too small?

Need new materials at the nano or even single-atom level which are governed by Quantum Mechanics



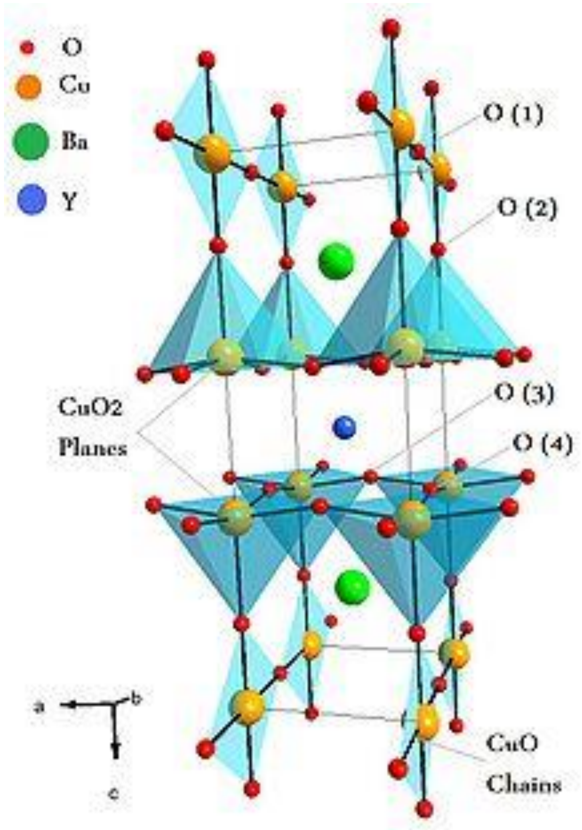


Traveling salesman problem

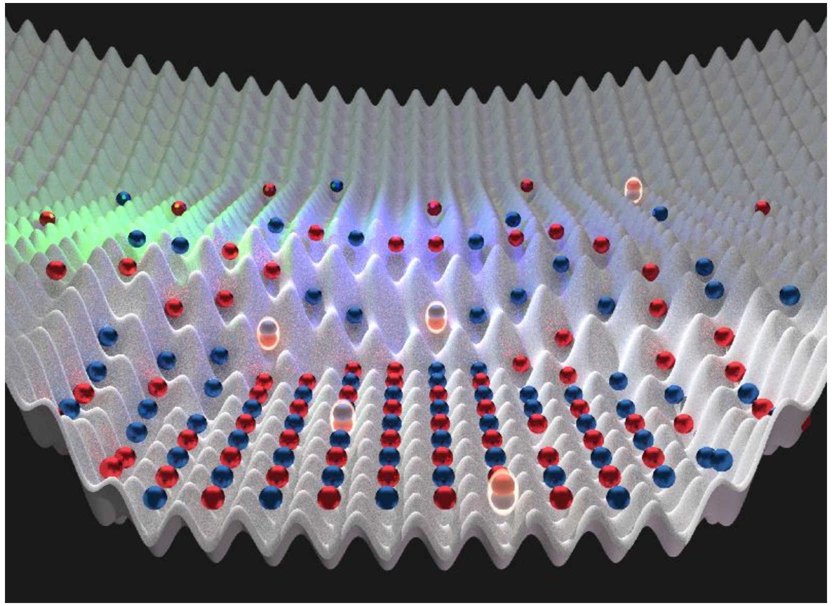
- The salesman must travel to all cities once before returning home
- The distance between each city is given, and is assumed to be the same in both directions
- Objective – minimize the total distance to be traveled

Quantum computer may solve problems that are hard or impossible on conventional classical computers

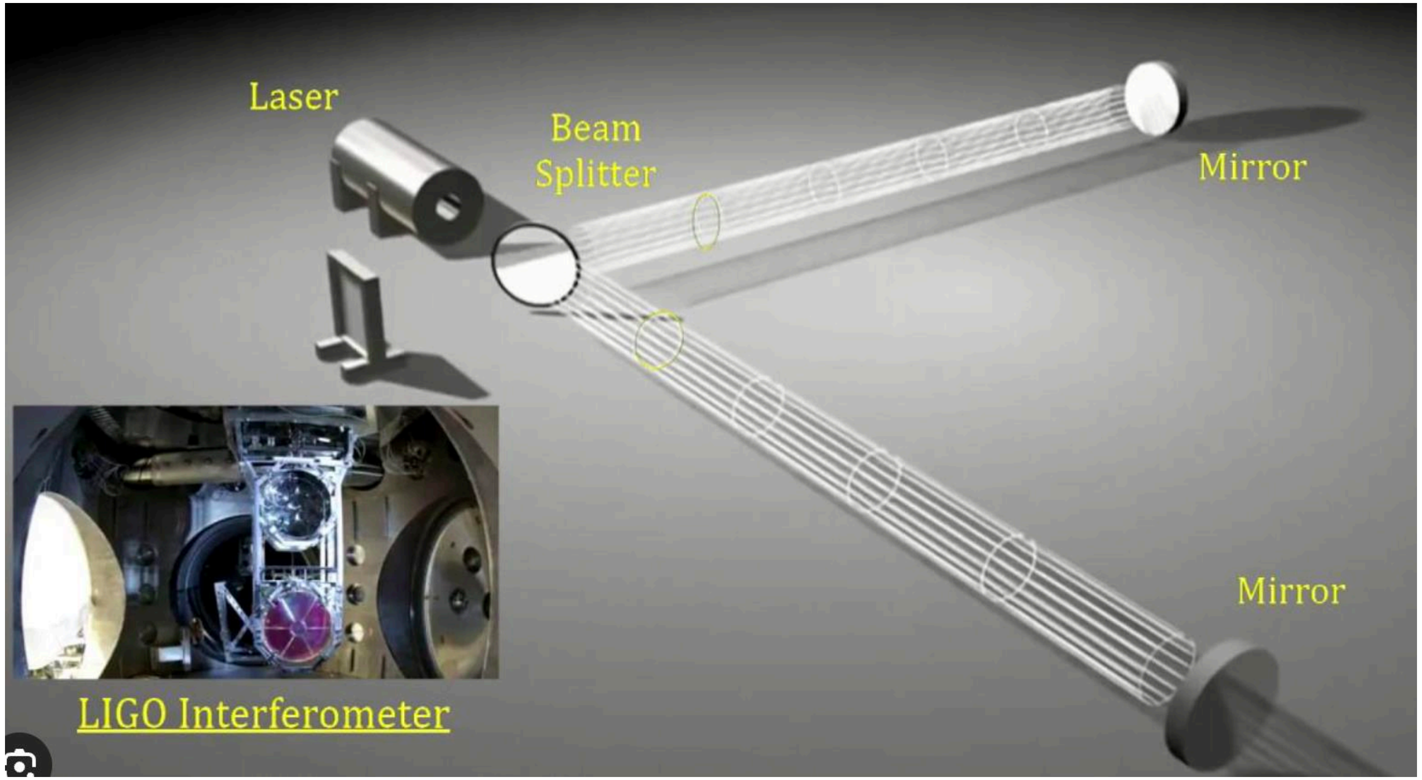
High-temperature superconductivity

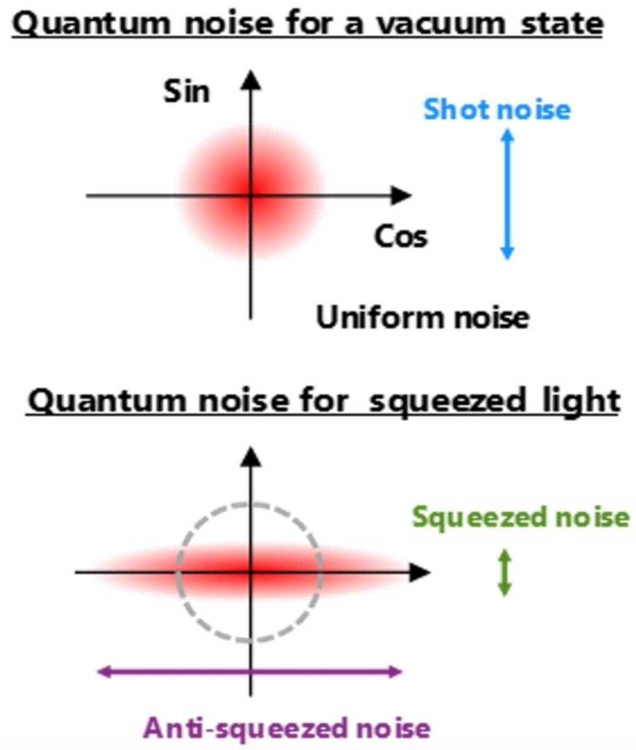


- Many-body problem – difficult to solve with classical computers
- Can make a quantum simulator: a controllable synthetic quantum system that mimics the behavior of real complex quantum material

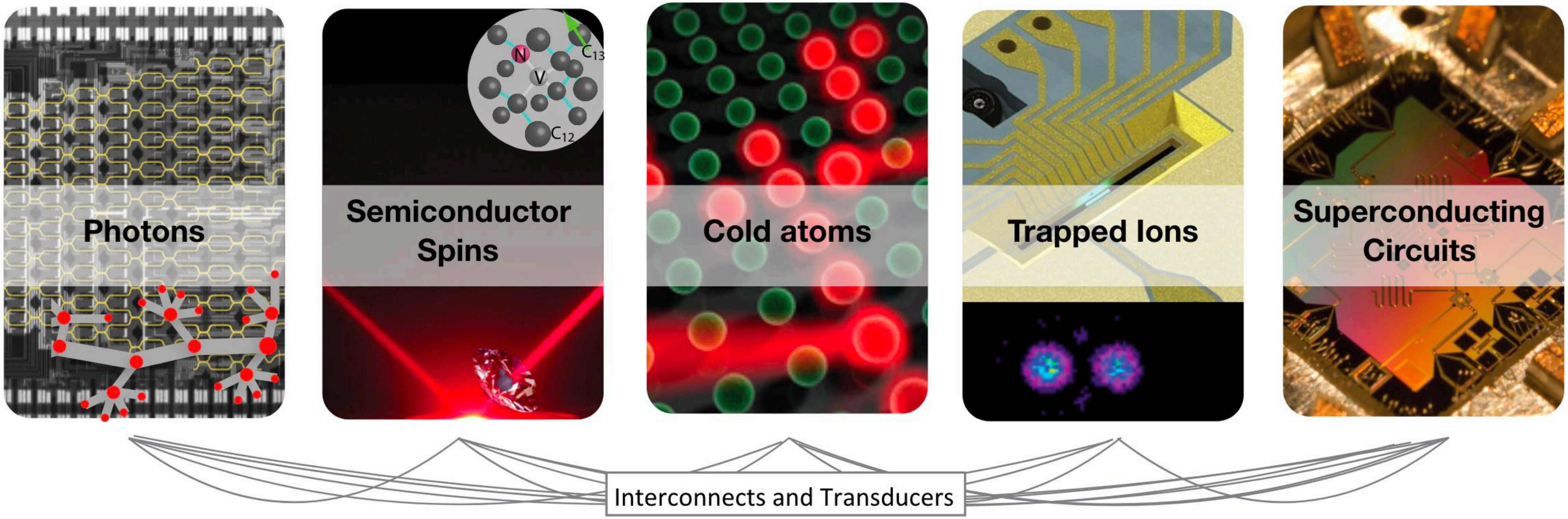


Quantum sensing



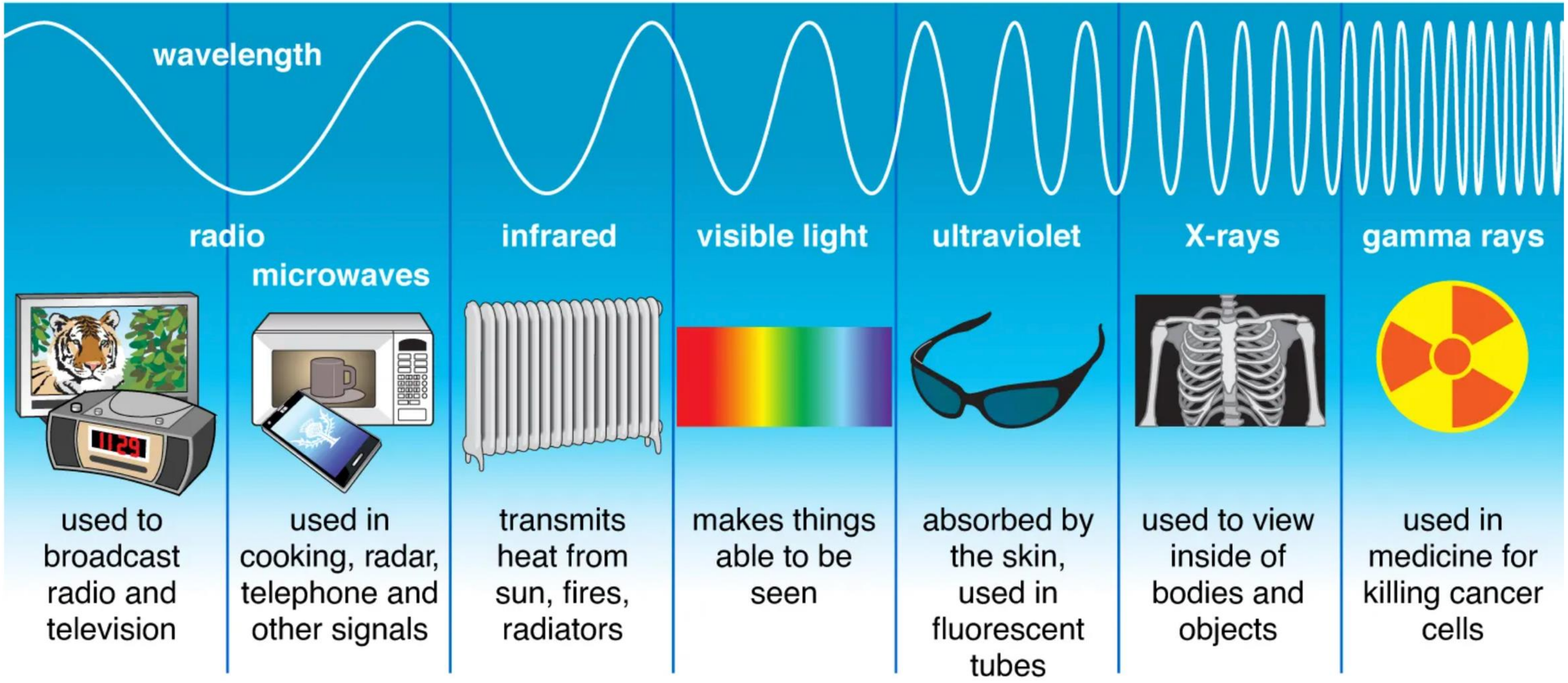


Leading physical platforms

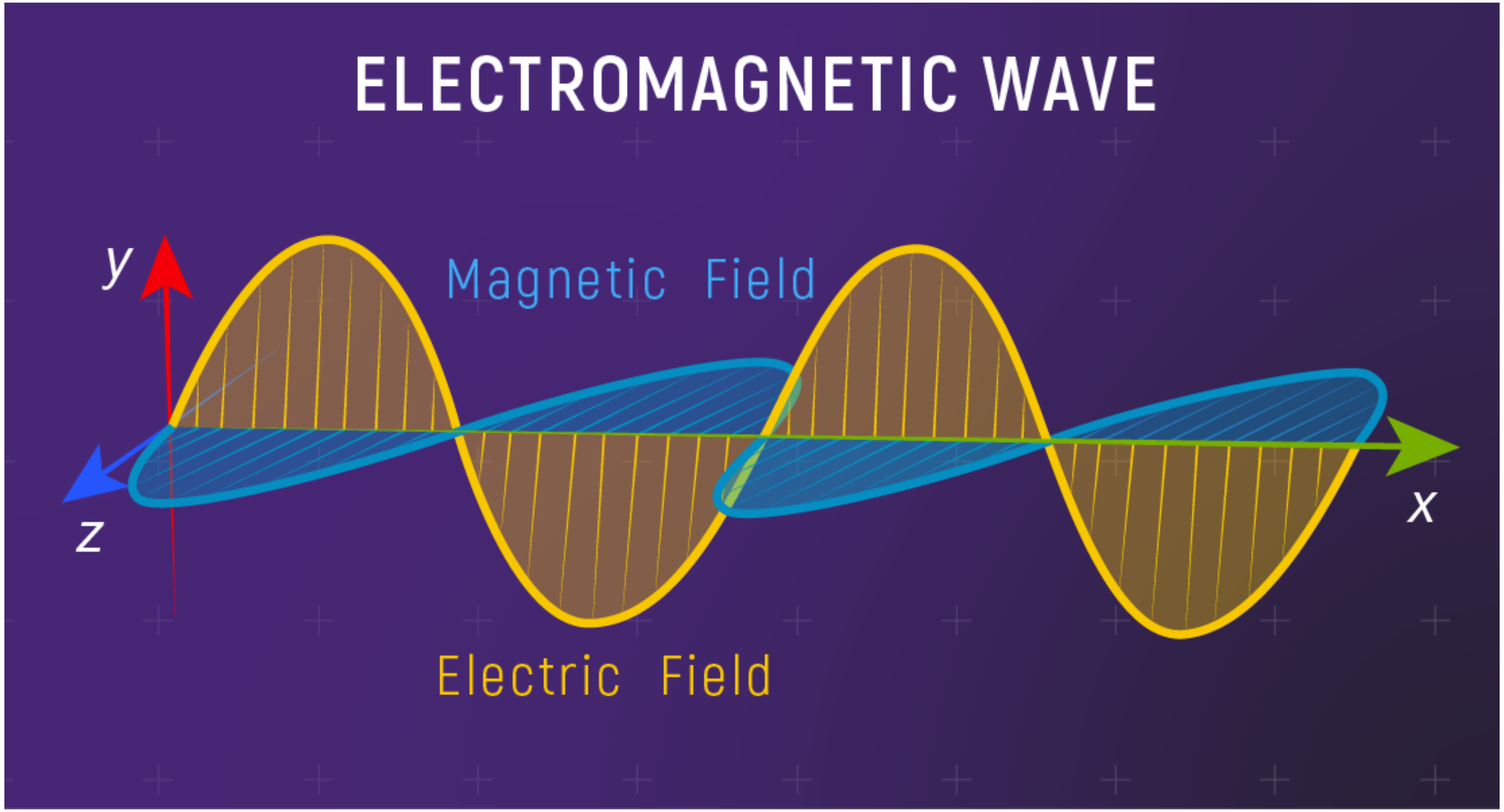


To connect these platforms often requires the development of transducers to photonic states, which can travel long distance with little decoherence

Quantum Photonics



Light is an electromagnetic wave, 400-700nm



Plane wave: the field vectors (**E** and **B**) are perpendicular to the propagation direction (**k**)

Wave nature of light

$$\mathbf{E}(\mathbf{x}, t) = Ae^{i\phi}e^{ikz - i\omega t}\hat{\mathbf{e}}$$

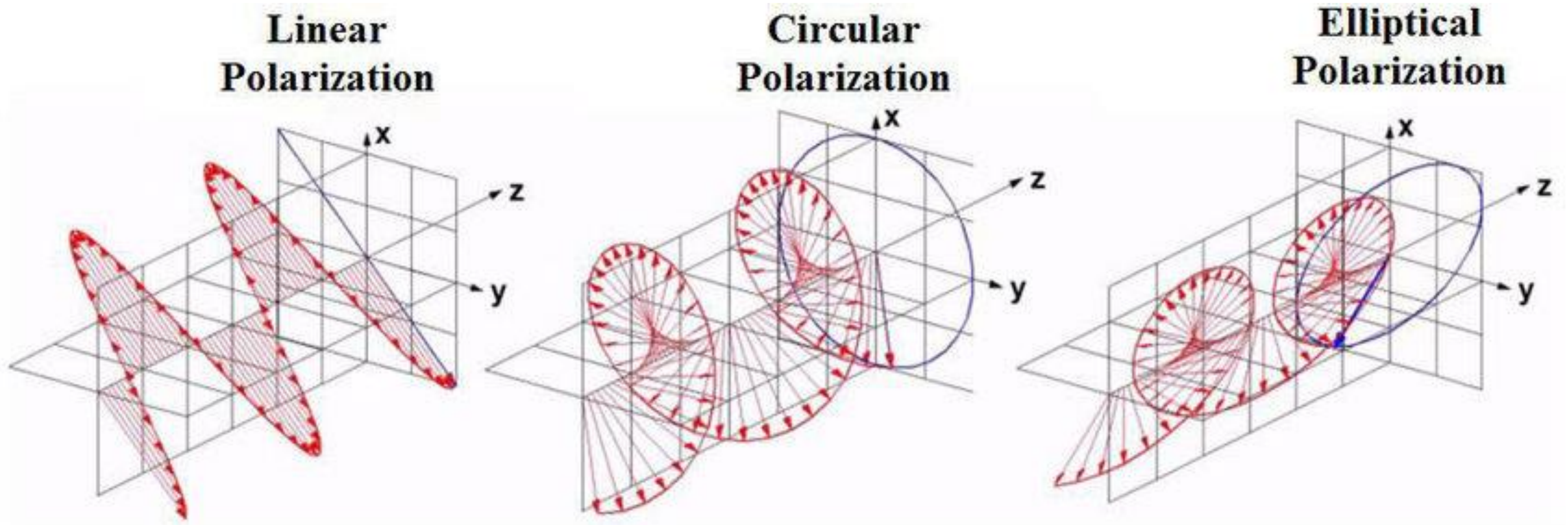
Wavevector, wavelength, frequency, velocity

$$k = \frac{2\pi}{\lambda} \qquad \lambda f = c$$

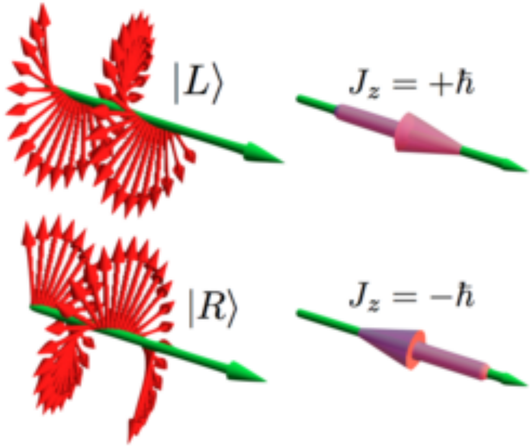
Polarization direction

$$\hat{\mathbf{e}} = \hat{\mathbf{e}}_x + e^{i\delta}\hat{\mathbf{e}}_y$$

Polarization



For circular polarized light,
it carries ***spin angular momentum***



$$|L\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\\ 0\end{pmatrix} e^{i(kz-\omega t)}$$

$$|R\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -i\\ 0\end{pmatrix} e^{i(kz-\omega t)}$$

Planck's Quantum Theory (1900)

$$\boxed{E = h\nu}$$

which is essentially a photon

where, **E** = energy of a quantum of radiation.

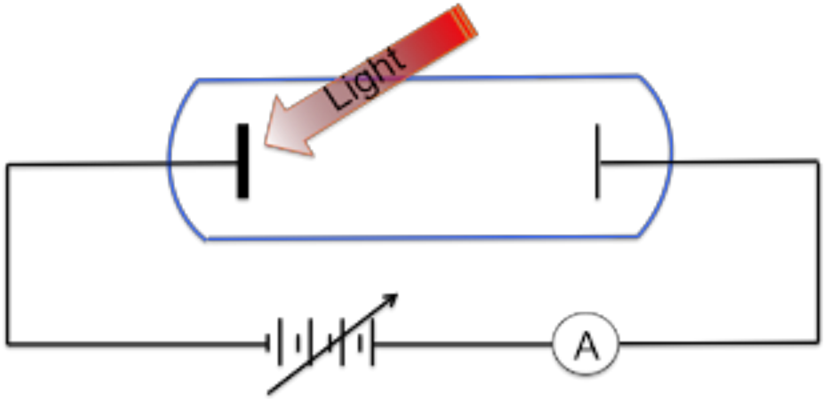
h = Plank constant = 6.627×10^{-27} erg sec

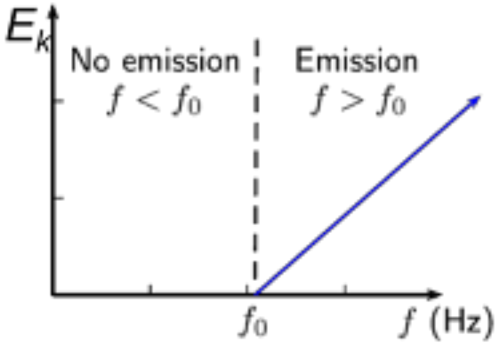
or 6.627×10^{-27} Joule sec,

ν = frequency of radiation

Photoelectric effect

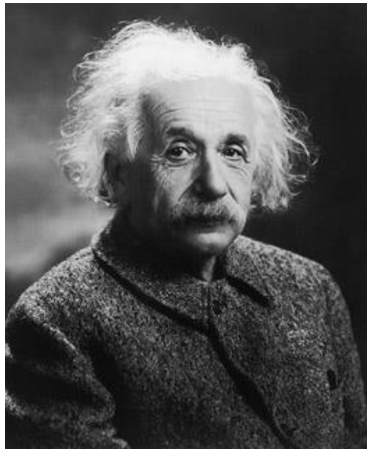
Photoelectric effect happens after light hits and excites electrons on the surface of a metal, and the electron ejects itself in the form of photoelectrons





- The kinetic energy of the photoelectrons depends solely on the wavelength of the incident light, not on its intensity
- The number of ejected photoelectrons is proportional to the light intensity
- There is no measurable time delay between irradiation and electron ejection

Photoelectric effect



1905 – Albert Einstein explained photoelectric effect
1921 – Nobel Prize in Physics

- Light consists of photons, each with a particular amount of energy
- Upon collision, each photon can transfer its energy to a single electron
- The more photons strike the surface of the metal, the more electrons are liberated and the higher is the current

$$hf = W + K.E_{max}$$

EM wave: a harmonic oscillator

Energy of the single-mode EM wave

$$
\begin{aligned}
H &= \frac{1}{2}\int dV\left[\varepsilon_0 \mathbf{E}^2(\mathbf{r},t) + \frac{1}{\mu_0}\mathbf{B}^2(\mathbf{r},t)\right] \\
&= \frac{1}{2}\int dV\left[\varepsilon_0 E_x^2(z,t) + \frac{1}{\mu_0}B_y^2(z,t)\right] \\
&= \frac{1}{2}(p^2 + \omega^2 q^2).
\end{aligned}
$$

which is the same as that for the 1D harmonic oscillator, with q and p interpreted as the canonical position and momentum

Quantization of EM wave

Correspondence principle

$$q \longrightarrow \hat{q}$$

$$p \longrightarrow \hat{p}$$

c-number   quantum operators

with commutation relation $[\hat{q}, \hat{p}] = i\hbar$

Quantum harmonic oscillator

The Hamiltonian

$$\hat{H} = \frac{1}{2}\left(\hat{p}^2 + \omega^2 \hat{q}^2\right)$$

- $\hat{p}^+ = \hat{p}$, $\hat{q}^+ = \hat{q}$ $\Rightarrow$ Hermitian operators $\Rightarrow$ observables

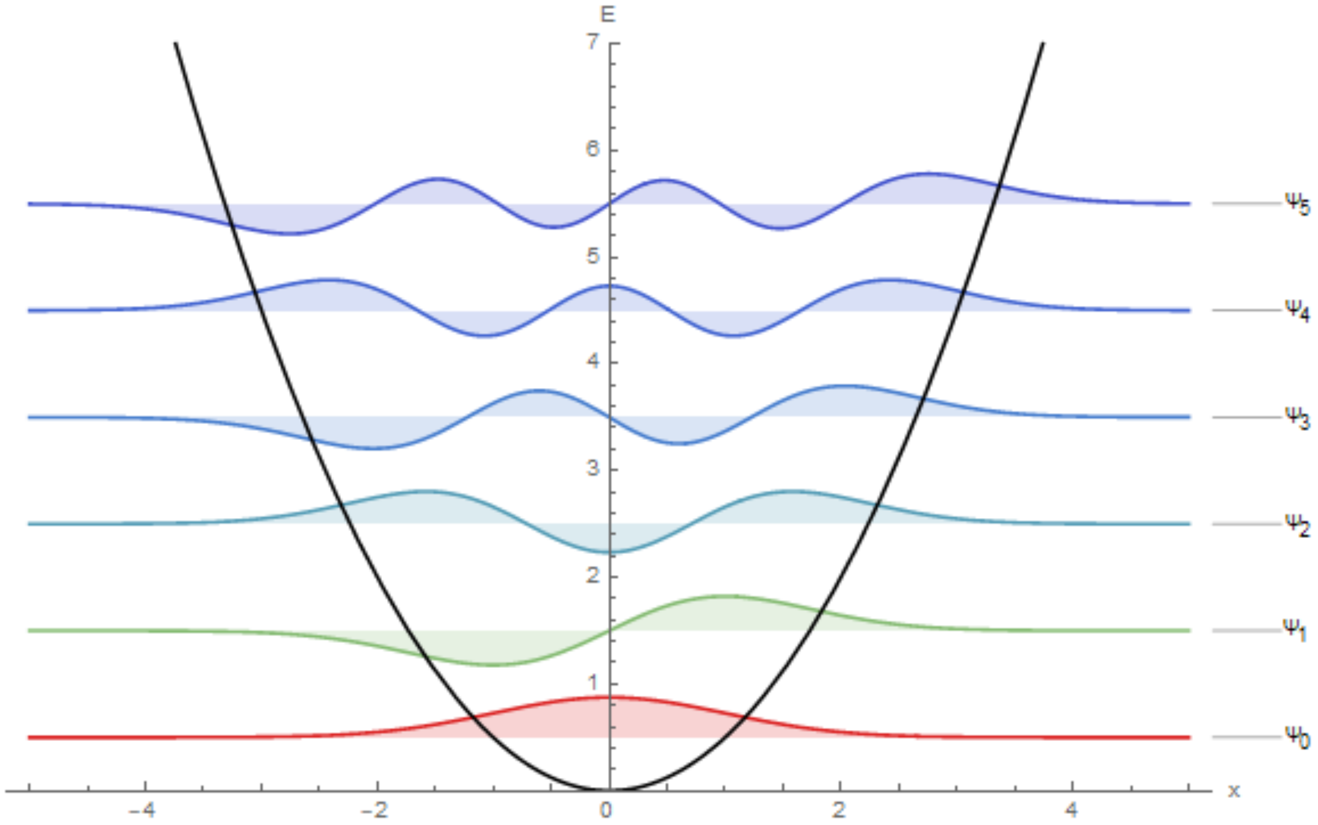
quadratures

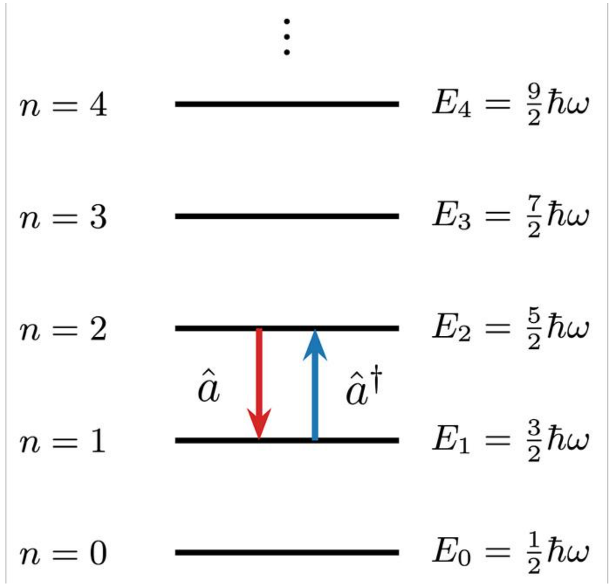
- It is convinient to introduce

$$\left.\begin{aligned} \hat{a} &= \frac{1}{\sqrt{2\hbar\omega}}\left(\omega\hat{q} + i\hat{p}\right) \\ \hat{a}^+ &= \frac{1}{\sqrt{2\hbar\omega}}\left(\omega\hat{q} - i\hat{p}\right) \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \hat{q} &= \sqrt{\frac{\hbar}{2\omega}}\left(\hat{a} + \hat{a}^+\right) \equiv \hat{X}_1 \\ \hat{p} &= -i\sqrt{\frac{\hbar\omega}{2}}\left(\hat{a} - \hat{a}^+\right) \equiv \hat{X}_2 \end{aligned}\right.$$

Note, $\hat{a}$ and $\hat{a}^+$ are not Hermitian operators $\Rightarrow$ not observables

Quantum harmonic oscillator





- Discrete energy levels
- The neighboring eigenstates are connected with the so-called creation and annihilation operators

Photon number states (Fock states)

$$H = \hbar\omega(a^+a + \tfrac{1}{2})$$

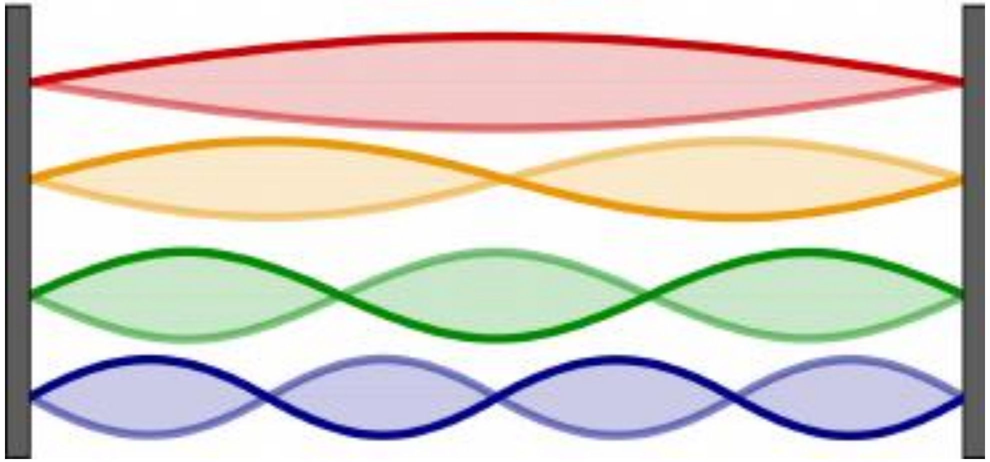
number operator $\hat{n} = \hat{a}^+\hat{a}$

$$\hat{n}|n\rangle = n|n\rangle$$

$$\boxed{a^+|n\rangle = \sqrt{n+1}\,|n+1\rangle}$$

$$\boxed{a|n\rangle = \sqrt{n}\,|n-1\rangle}$$

$$\boxed{|n\rangle = \frac{1}{\sqrt{n!}}(a^+)^n|0\rangle}$$



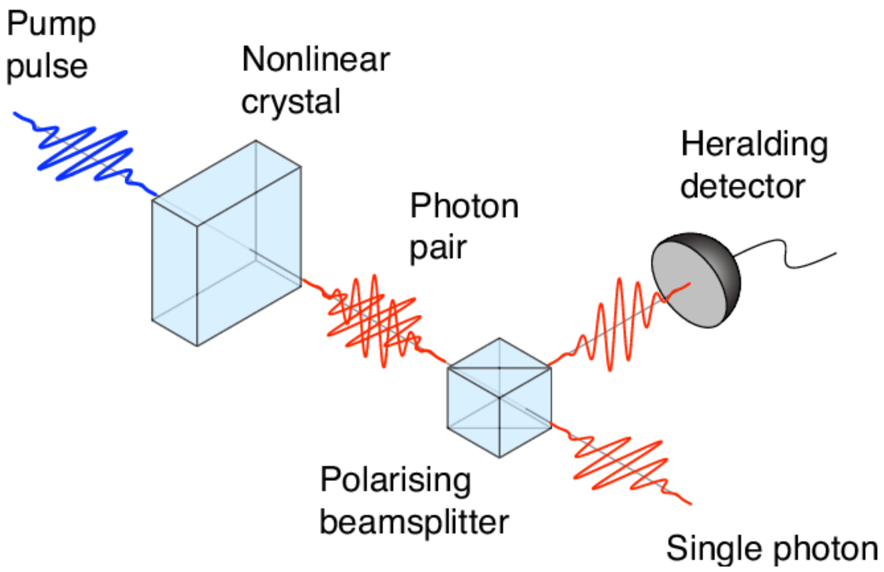
For each of these spatial modes, it may be occupied by 0, 1, 2, …, photons

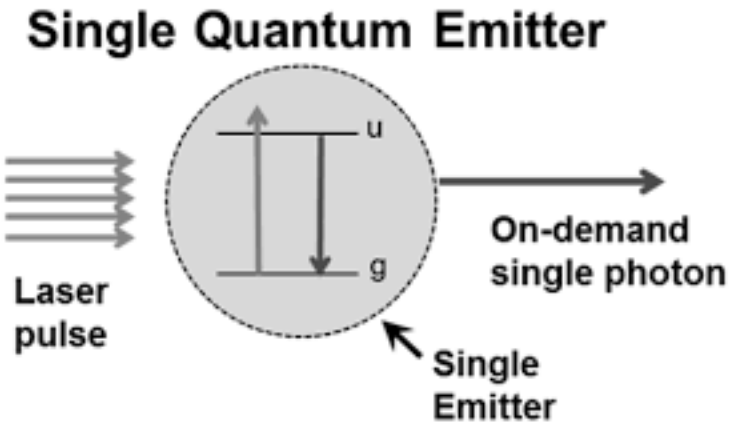
If n=0, it is vacuum state

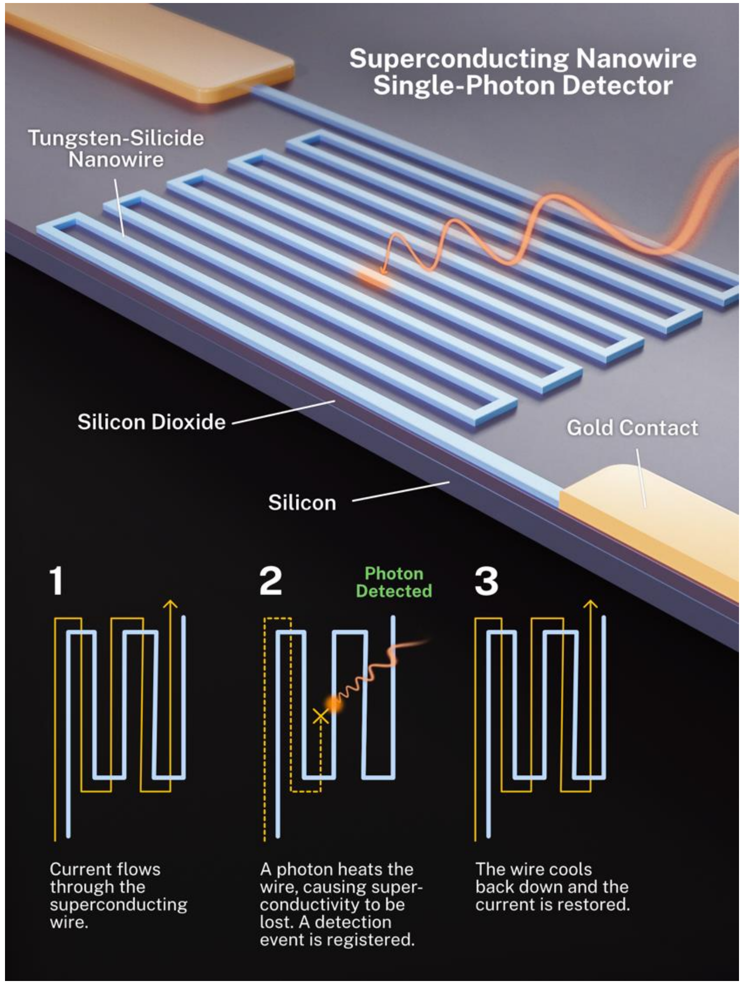
If n=1, it is a single-photon state

Single-photon source and detector

are crucial for quantum technology



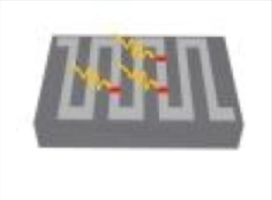




Photon number resolving detectors

Useful for many-photon experiments

Optica Vol. 4, Issue 12, pp. 1534-1535 (2017) • https://doi.org/10.1364/OPTICA.4.001534



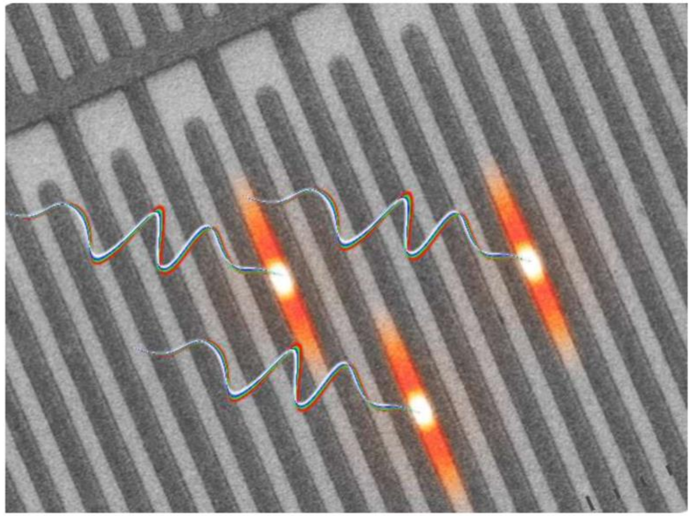
Multi-photon detection using a conventional superconducting nanowire single-photon detector

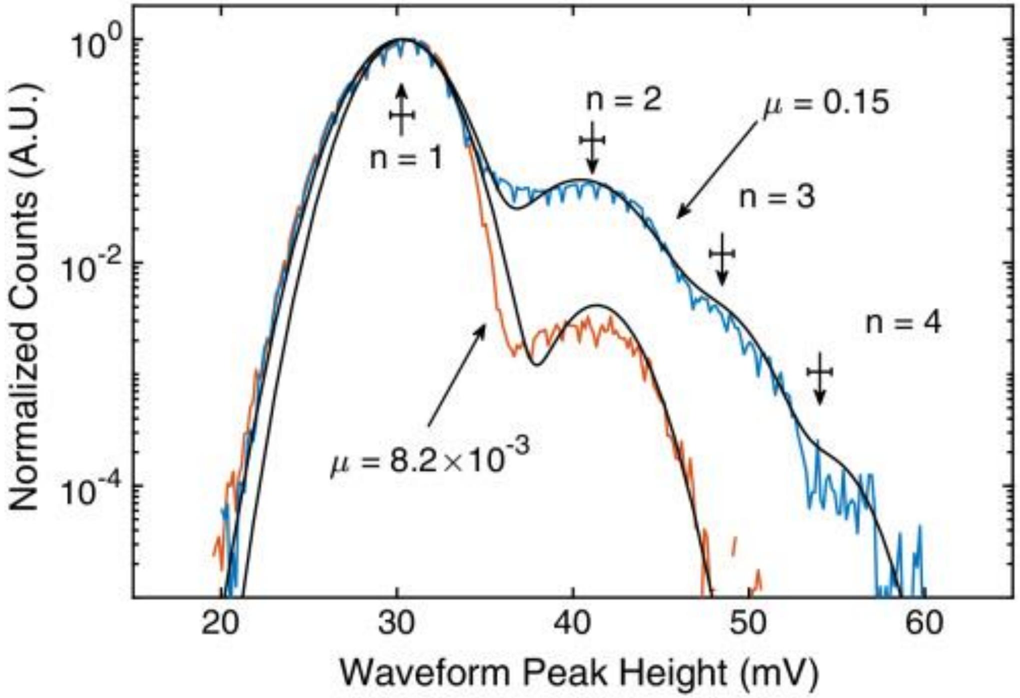
Clinton Cahall, Kathryn L. Nicolich, Nurul T. Islam, Gregory P. Lafyatis, Aaron J. Miller, Daniel J. Gauthier, and Jungsang Kim

DECEMBER 15, 2017

Single-photon detector can count to four

by Ken Kingery, Duke University





Quantum fluctuation of a single-mode EM field

- As a physical **observable**, E field can be directly measured.

- Since photon number states (Fock states) are NOT the eigenstates of E field, **we cannot measure E field accurately** when the wave is in such a quantum state

$$\hat{E}_x(z,t) = \mathcal{E}_0 (a + a^+) \sin(kz) \quad \text{where} \quad \mathcal{E}_0 = \sqrt{\frac{\hbar\omega}{\varepsilon_0 V}}$$

a and a^+ are not Hermitian operators $\rightarrow$ not directly observable

$\hat{E}_x(z,t) \propto (a + a^+)$ is Hermitian $\rightarrow$ directly observable

Quantum fluctuation of a single-mode EM field

In number states

$$\langle n| E_x(z,t) |n\rangle = \mathcal{E}_0 \sin(kz) \left[\langle n| a |n\rangle + \langle n| a^+ |n\rangle \right] = 0$$

$$\langle n| E_x^2(z,t) |n\rangle = \langle n| \mathcal{E}_0^2 \sin^2(kz) (a+a^+)^2 |n\rangle$$

$$= \mathcal{E}_0^2 \sin^2(kz) \langle n| a^2 + aa^+ + a^+a + (a^+)^2 |n\rangle$$

$$= \mathcal{E}_0^2 \sin^2(kz) \langle n| aa^+ + a^+a |n\rangle$$

$$= \mathcal{E}_0^2 \sin^2(kz) \langle n| 2a^+a + 1 |n\rangle$$

$$= \mathcal{E}_0^2 \sin^2(kz) (2n+1)$$

$$= 2\mathcal{E}_0^2 \sin^2(kz) \left(n + \tfrac{1}{2}\right)$$

Quantum fluctuation of a single-mode EM field

Variance of E_x

$$\langle \Delta E_x^2 \rangle = \langle E_x^2 \rangle - \langle E_x \rangle^2$$

$$= 2\,\mathcal{E}_0^2 \sin^2(kz)\left(n + \tfrac{1}{2}\right)$$

Standard deviation of E_x

$$\Delta E_x = \sqrt{\langle \Delta E_x^2 \rangle} = \sqrt{2}\,\mathcal{E}_0 \sin(kz)\sqrt{n + \tfrac{1}{2}}$$

Even for $n=0$ state, there is fluctuation $\longrightarrow$ vacuum fluctuation

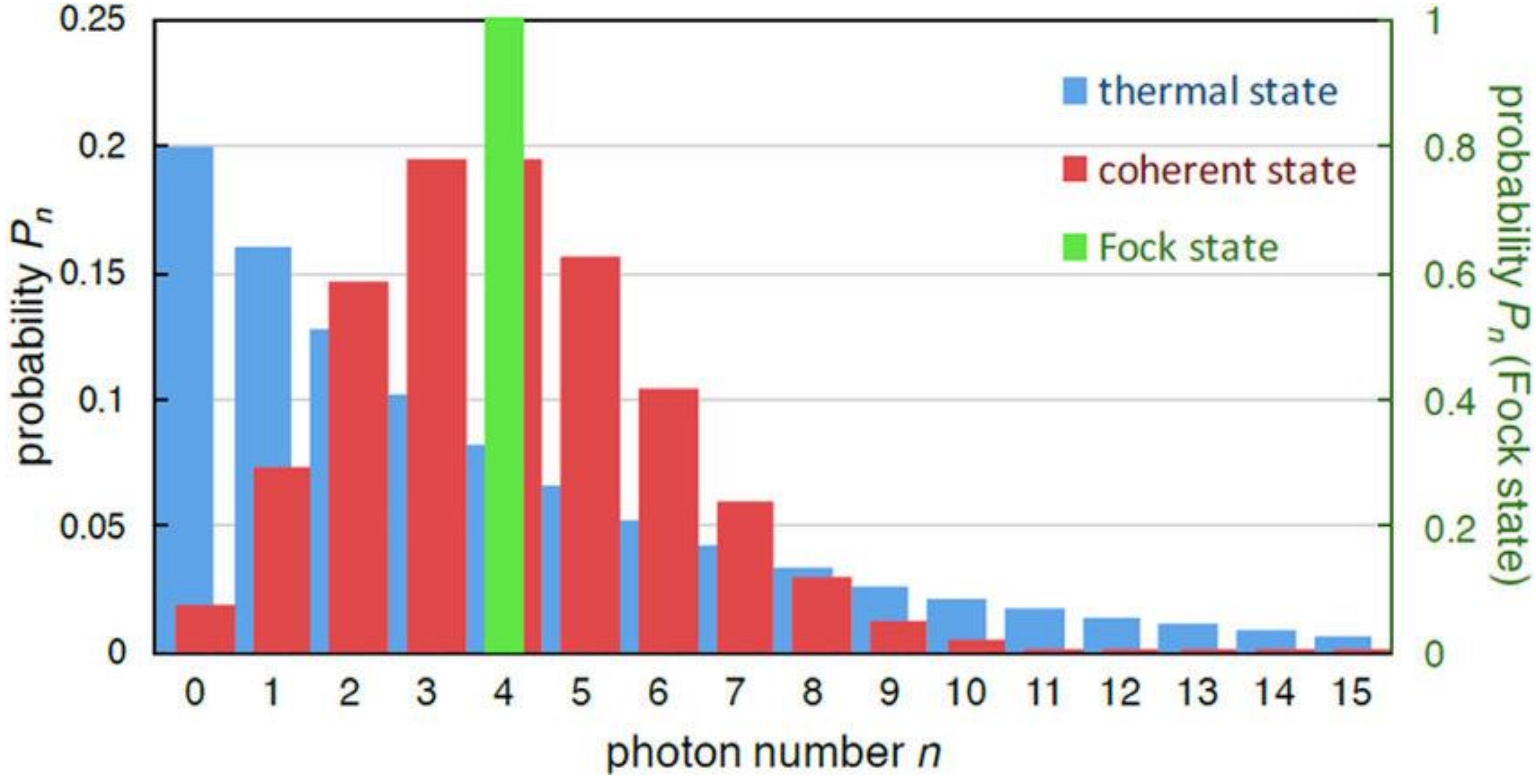
Other quantum states of light

In many cases, light is not in the Fock state and there is no definite photon number. For example, laser is not in the photon number state.

Below are some typical states of light:

- **Thermal state (natural light)**
- **Coherent state (laser light)**
- **Squeezed state**
  - **One-mode squeezed state**
  - **Two-mode squeezed state**
- **N00N state**

Photon number distribution of different quantum states of light



Thermal state: natural light
(exponentially decaying)

Coherent state: laser light

Fock state: photon number state

Coherent state (laser)

It is defined as the eigenstate of the annihilation operator

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$$

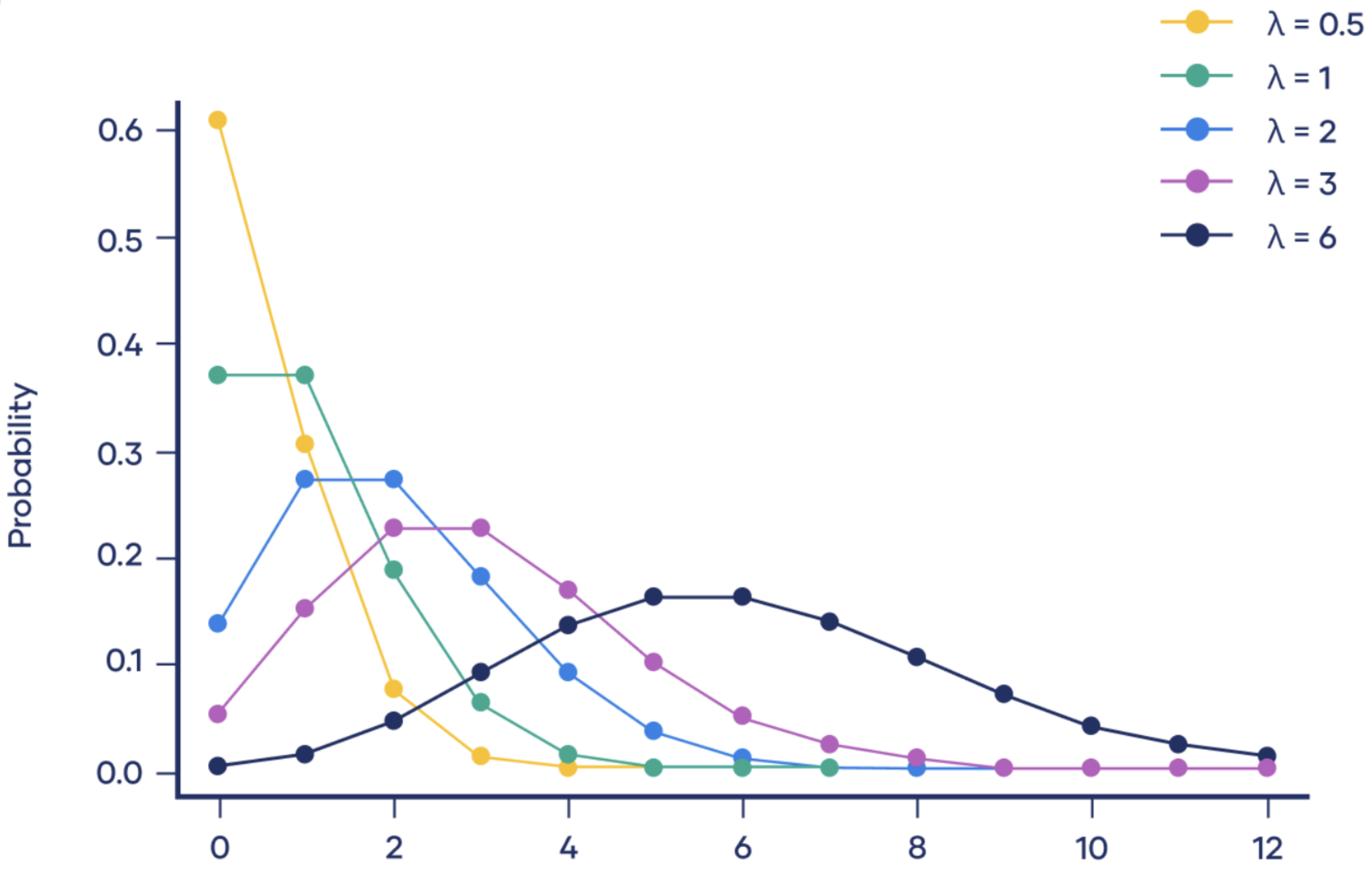
It can be expressed as a superposition of photon number states (Fock states)

$$|\alpha\rangle = \sum_{n=0} \alpha^n \frac{e^{\frac{-|\alpha|^2}{2}}}{\sqrt{n!}} |n\rangle$$

The probability of finding n photons in the coherent state is [a Poisson distribution]

$$P_n = \frac{\left(|\alpha|^2\right)^n}{n!} e^{-|\alpha|^2}$$

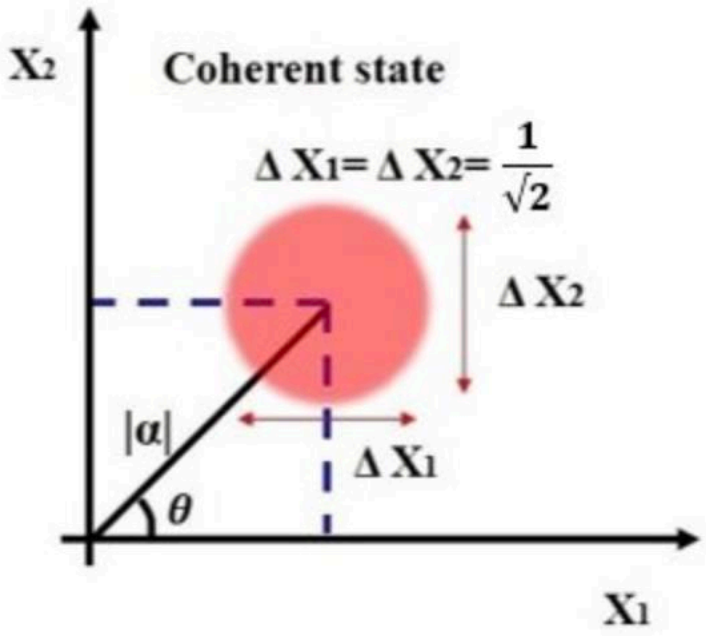
Poisson distribution



x (photon number)

$$f(x) = \frac{\lambda^x}{x!}e^{-\lambda}$$

Phase space representation of coherent state



For coherent state, $|\alpha\rangle$, we know

$$\langle\alpha|(\Delta\hat{X}_1)^2|\alpha\rangle = \frac{1}{4}$$

$$\langle\alpha|(\Delta\hat{X}_2)^2|\alpha\rangle = \frac{1}{4}$$

For vacuum state $|0\rangle$, we also have

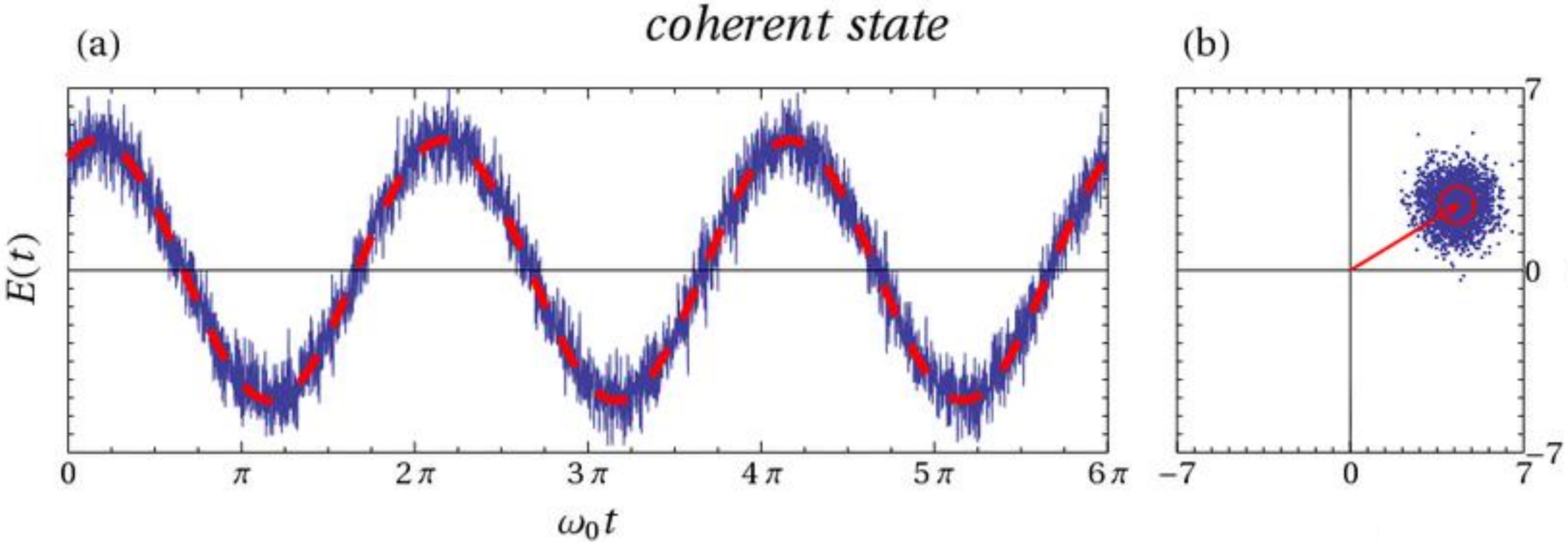
$$\langle 0|(\Delta\hat{X}_1)^2|0\rangle = \langle 0|(\Delta\hat{X}_2)^2|0\rangle = \frac{1}{4}$$

$$\bar{n} = \langle\alpha|\hat{n}|\alpha\rangle = \langle\alpha|a^\dagger a|\alpha\rangle = |\alpha|^2$$

For laser with higher intensity, the corresponding average photon number is larger

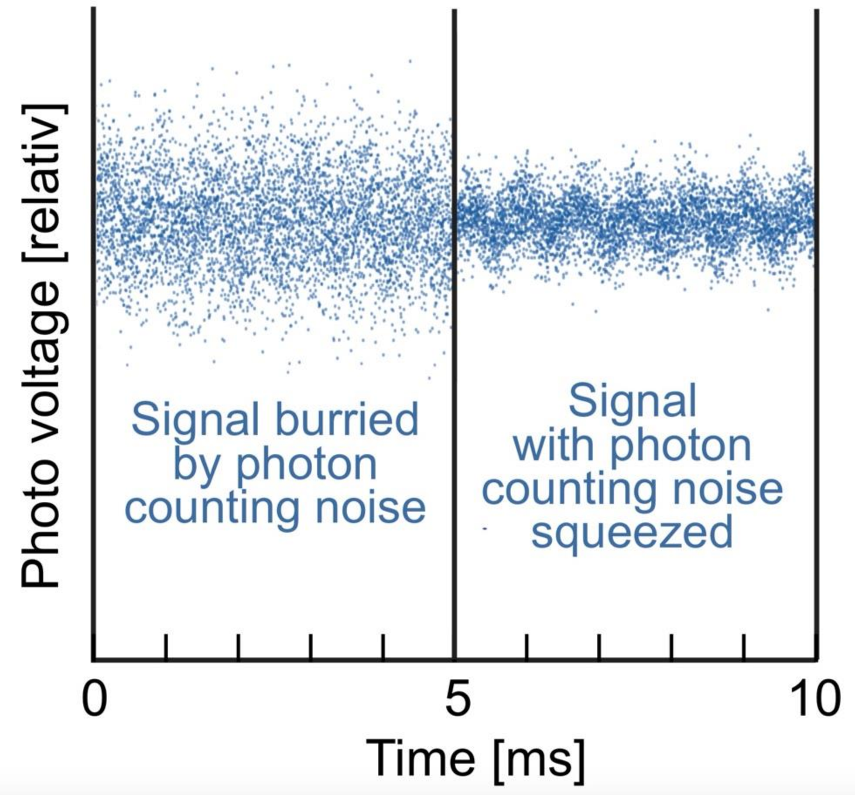
Coherent state

Photon number in a coherent state (laser beam) is indefinite
Thus, the measurement of electric field always show some noise (due to the photon number fluctuation)



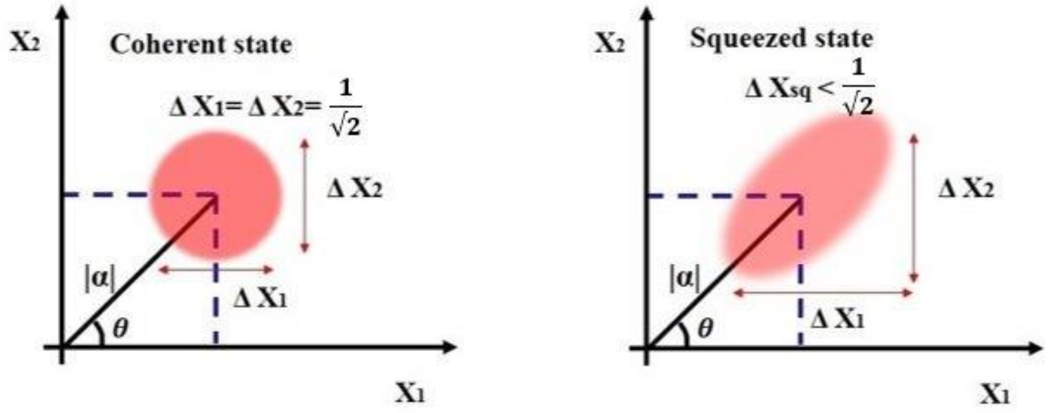
Phase space

Squeezed light

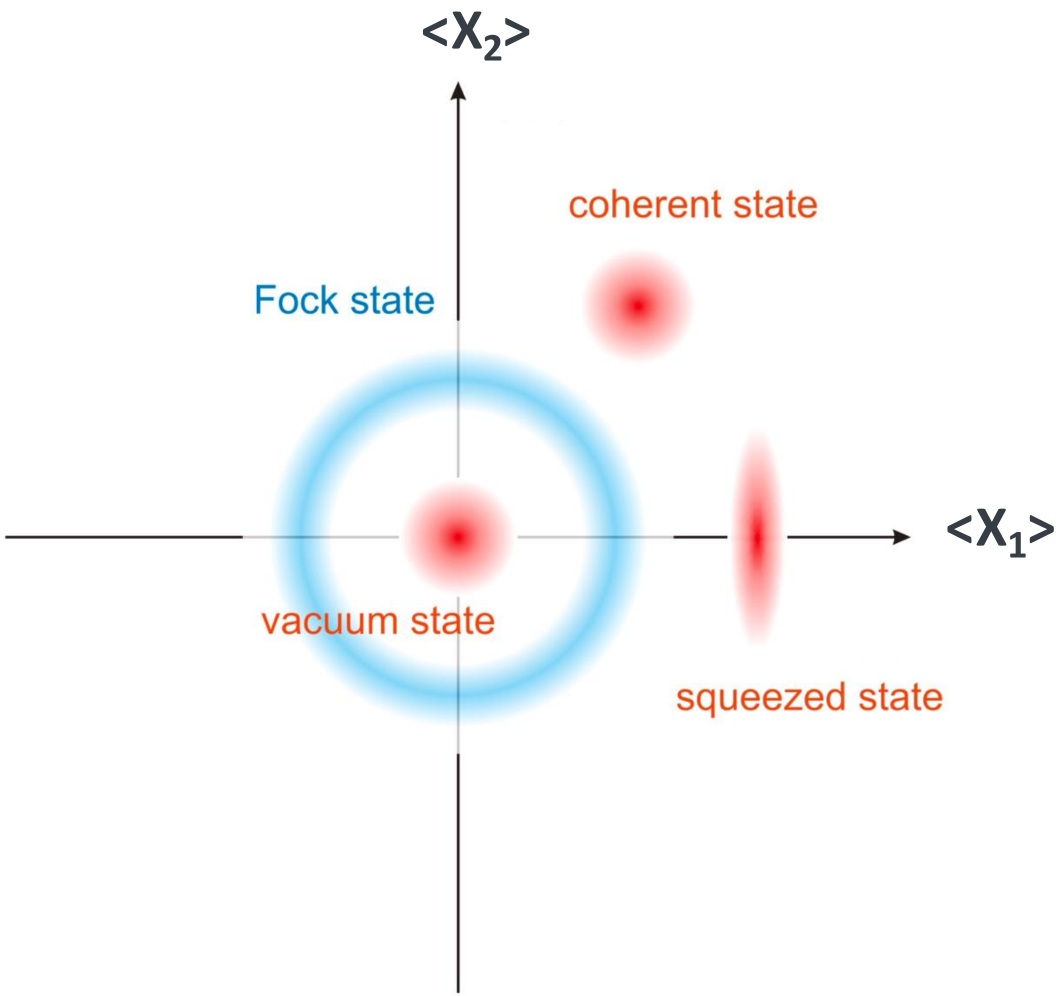


How to reduce the fluctuation?

We need entanglement!



Phase space representation of quantum states of light



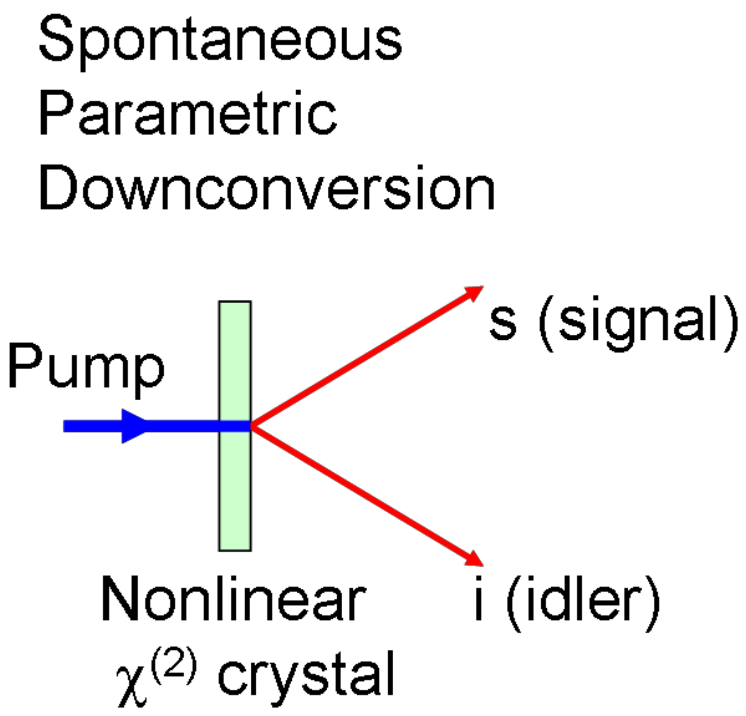
$$\hat{q} = \sqrt{\frac{\hbar}{2\omega}}\left(\hat{a} + \hat{a}^{+}\right) \equiv \hat{X}_1$$

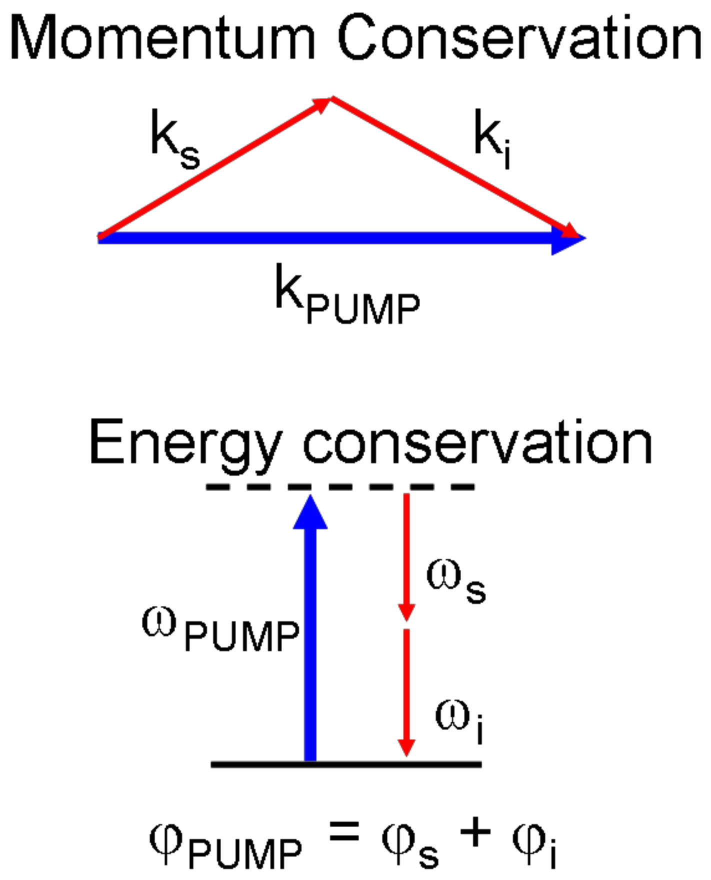
$$\hat{p} = -i\sqrt{\frac{\hbar\omega}{2}}\left(\hat{a} - \hat{a}^{+}\right) \equiv \hat{X}_2$$

The average values of the two quadratures and their fluctuations are very different in different quantum states of lights

Spontaneous parametric down-conversion (SPDC)

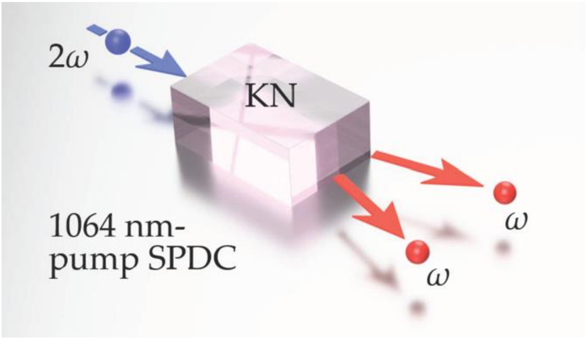
- **Pump light is very weak**
- **To create entangled photon pairs from vacuum fluctuation**





Degenerate case: single-mode squeezing

$$H = \hbar\omega a^\dagger a + \hbar\omega_p b^\dagger b + i\hbar\chi^{(2)}\left[a^2 b^\dagger - (a^\dagger)^2 b\right]$$



Under this Hamiltonian, the wave function evolves from vacuum state to

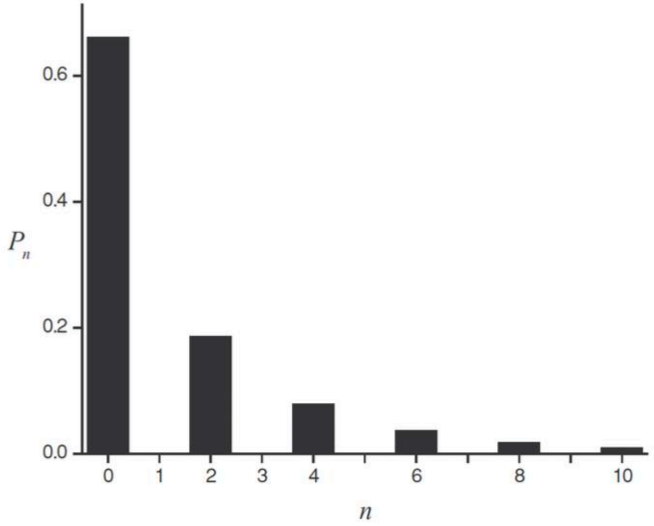
$$|\xi\rangle = \frac{1}{\sqrt{\cosh r}} \sum_{m=0}^{\infty} (-1)^m \frac{\sqrt{(2m)!}}{2^m m!} e^{im\theta} (\tanh r)^m |2m\rangle$$

The probability of finding 2m photons

$$P_{2m} = \frac{(2m)!}{2^{2m}(m!)^2} \frac{(\tanh r)^{2m}}{\cosh r}$$

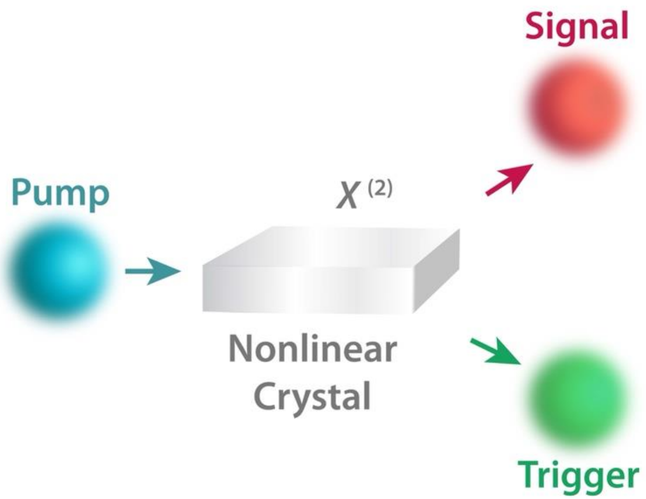
and

$$P_{2m+1} = 0$$

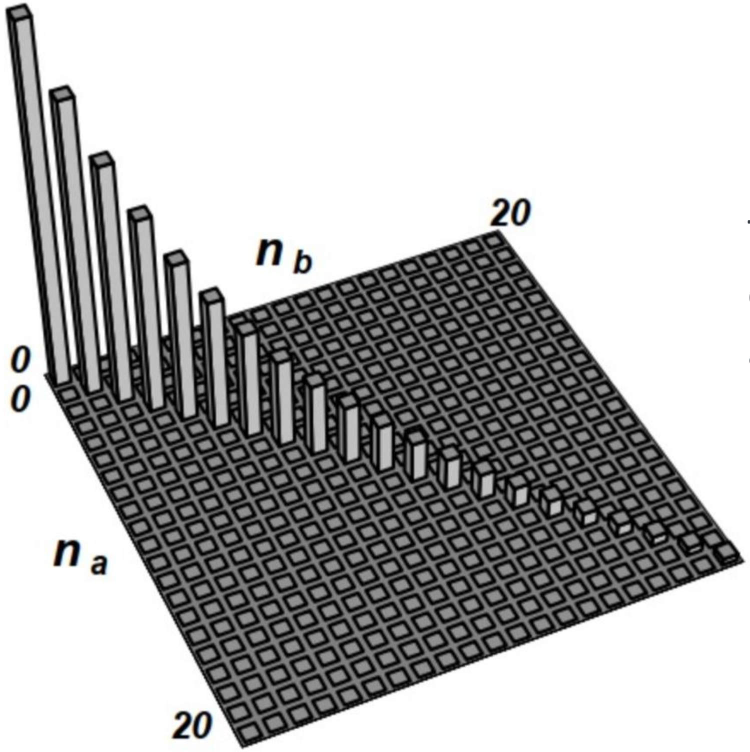


Since photon pairs are created, when expanded as a superposition of Fock states, we only have terms corresponding to even number of photons

Non-degenerate case: two-mode squeezed state



$$|\psi(0)\rangle = \frac{1}{\cosh r}\sum_{n=0}^{\infty}(\tanh r)^n |n, n\rangle$$



Exponentially decaying

The photon number distribution is the same as thermal light

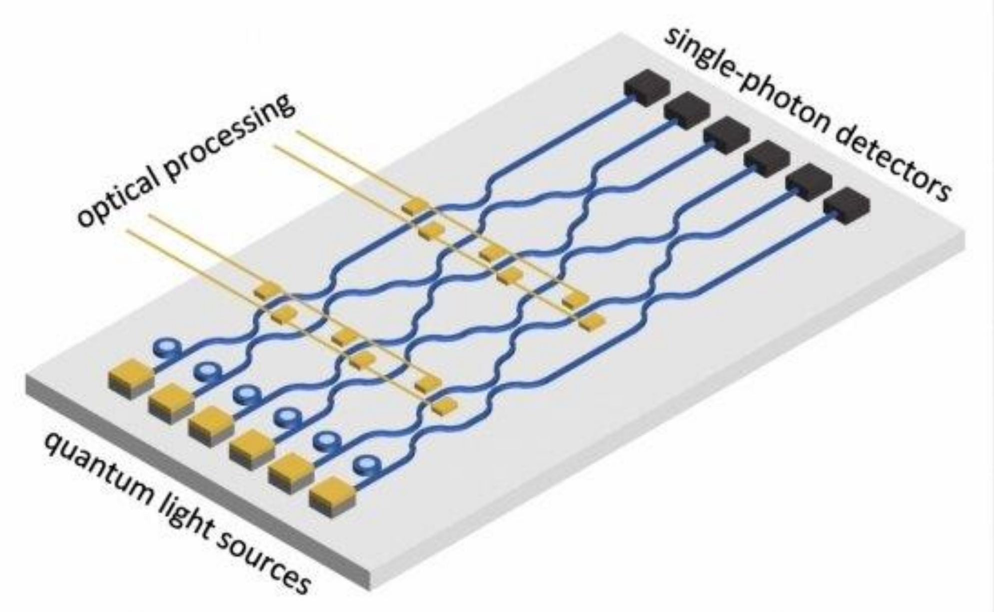
The generated photon pair do not have the same frequency – they are distinguishable
When expanded as a superposition of Fock states, each terms include the same number of idler photons and signal photons

Quantum photonics

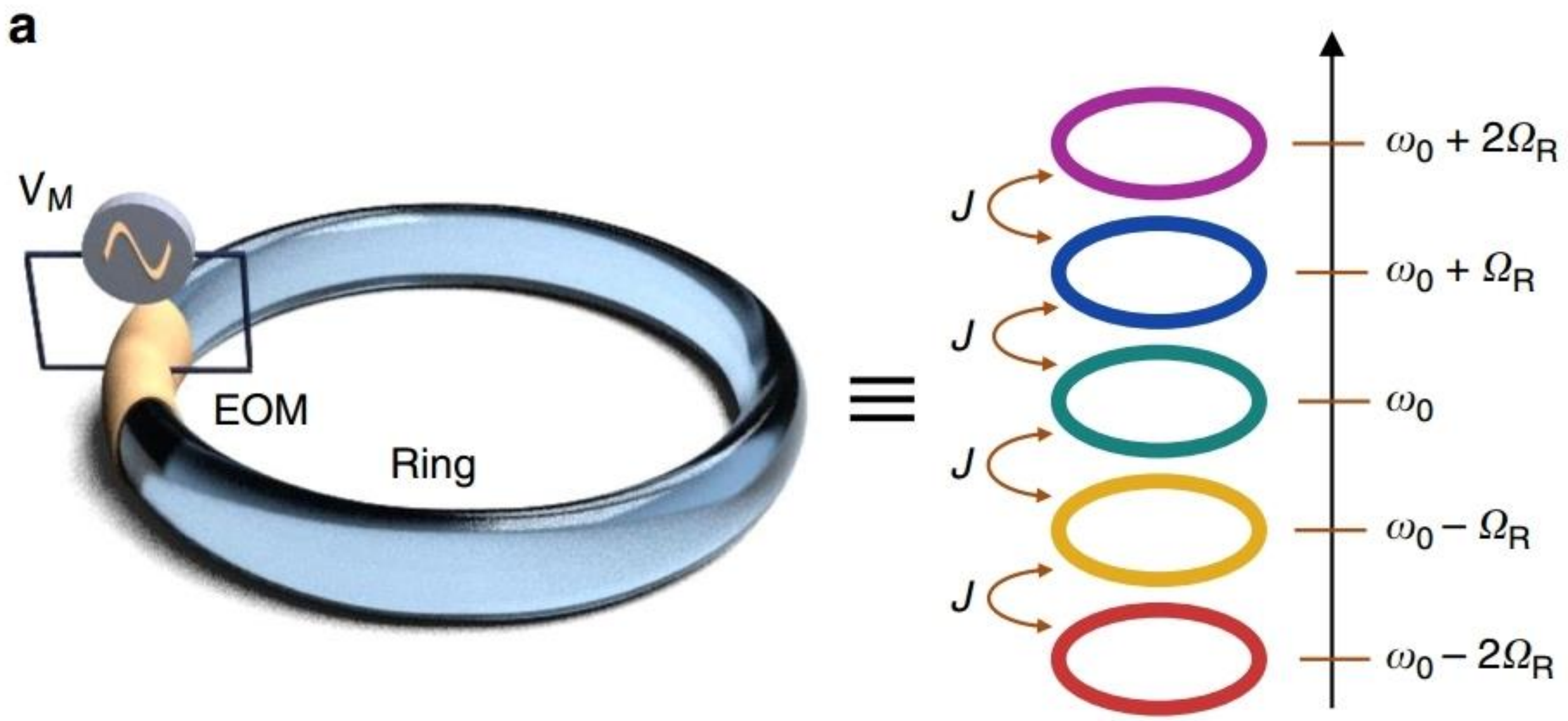
Linear optical quantum computing (LOQC)

- Use spatial modes

However, there are limitations with path encoding → scalability of the device in physical dimension



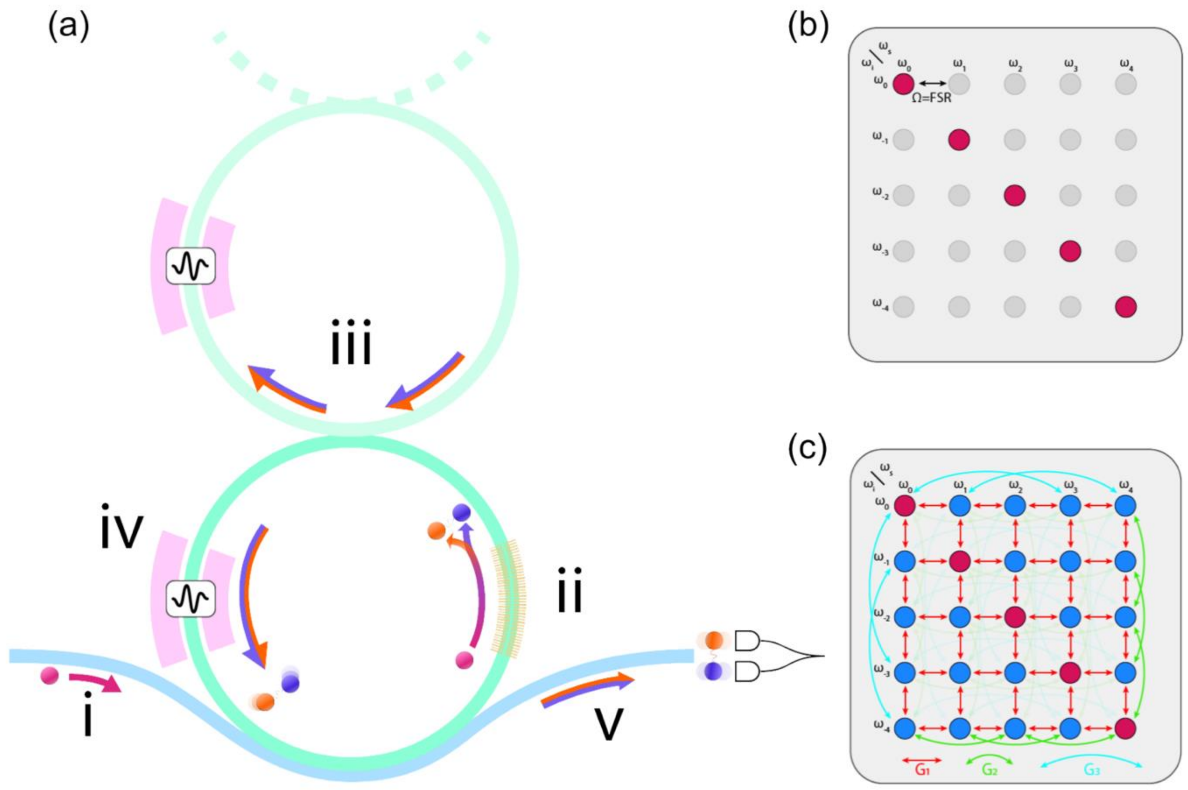
Synthetic dimension



Cavity photons frequency can take many possible values

These frequency modes provide additional degree of freedom for quantum information encoding and manipulation

Correlation photon pairs in a ring cavity



Joint-spectral-intensity
(density-density correlation)

$$[\mathrm{JSI}]_{mn} = \langle \hat{c}^{\dagger}_{s,n}(\omega_1)\hat{c}^{\dagger}_{i,m}(\omega_2)\hat{c}_{i,m}(\omega_3)\hat{c}_{s,n}(\omega_4)\rangle$$

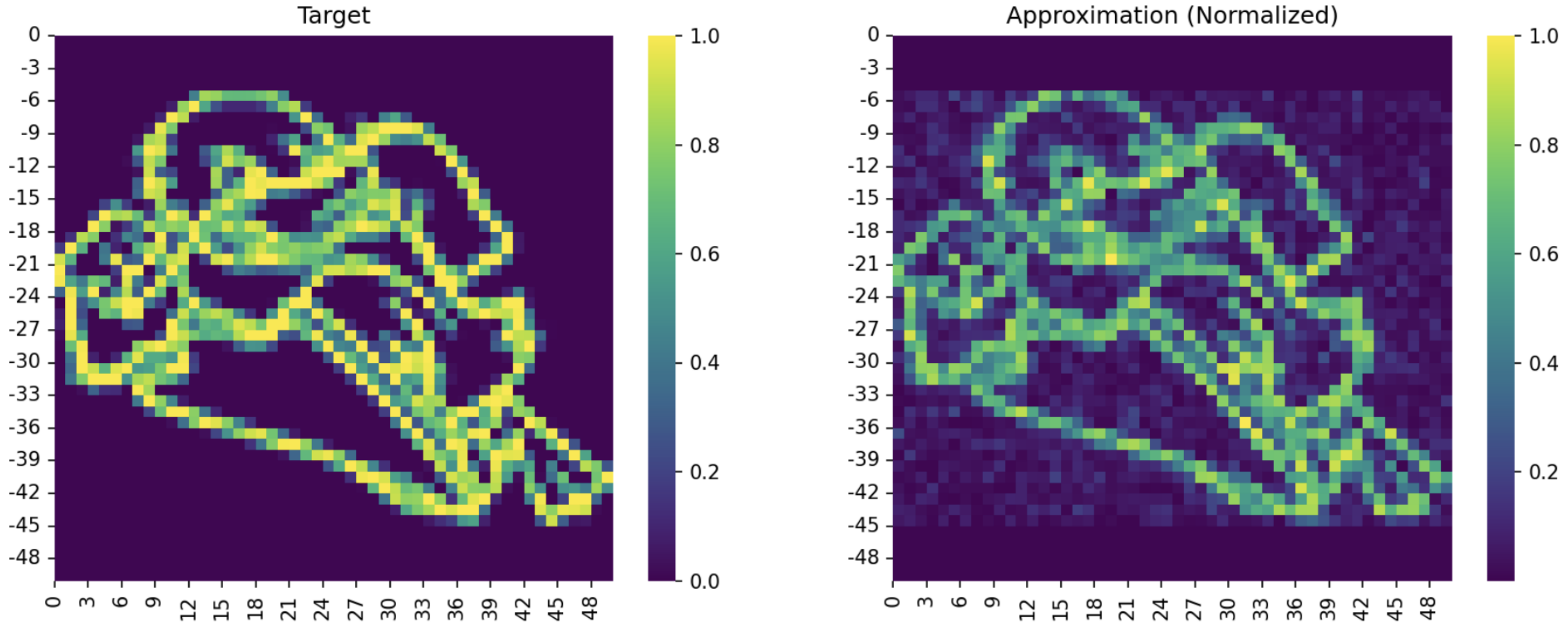
JSI matrix

Depends on

- Number of rings
- Coupling between rings
- Coupling between frequency modes due to EO modulation
- Losses

$$\overleftrightarrow{\mathrm{JSI}} = 4\pi^2 diag(\gamma_{\vec{ex};S})[(\overleftrightarrow{T_0}^* - \overleftrightarrow{T_1}^* \overleftrightarrow{T_2}^T\ diag(\vec{\gamma_t})) \odot (\overleftrightarrow{T_0} - \overleftrightarrow{T_1}\overleftrightarrow{T_2}^{*T}\ diag(\vec{\gamma_t}))]diag(\gamma_{\vec{ex};I})$$

Generating arbitrary correlation matrix?



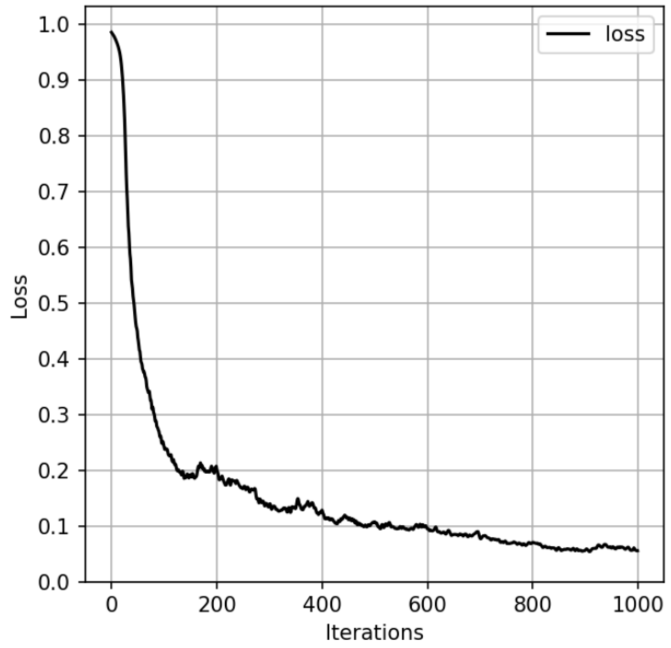
Gradient descent optimization

- 5 rings
- 149 modulation lengths
- 50 padded dimensions on each side of the pattern

Loss vs. gradient descent optimizer iteration

Frobenius norm

$$\text{loss} = 1 - \frac{\mathrm{tr}^2(\overleftrightarrow{\mathrm{JSI}}_{pred}\,\overleftrightarrow{\mathrm{JSI}}_{target})}{\mathrm{tr}(\overleftrightarrow{\mathrm{JSI}}^2_{pred})\mathrm{tr}(\overleftrightarrow{\mathrm{JSI}}^2_{target})}$$



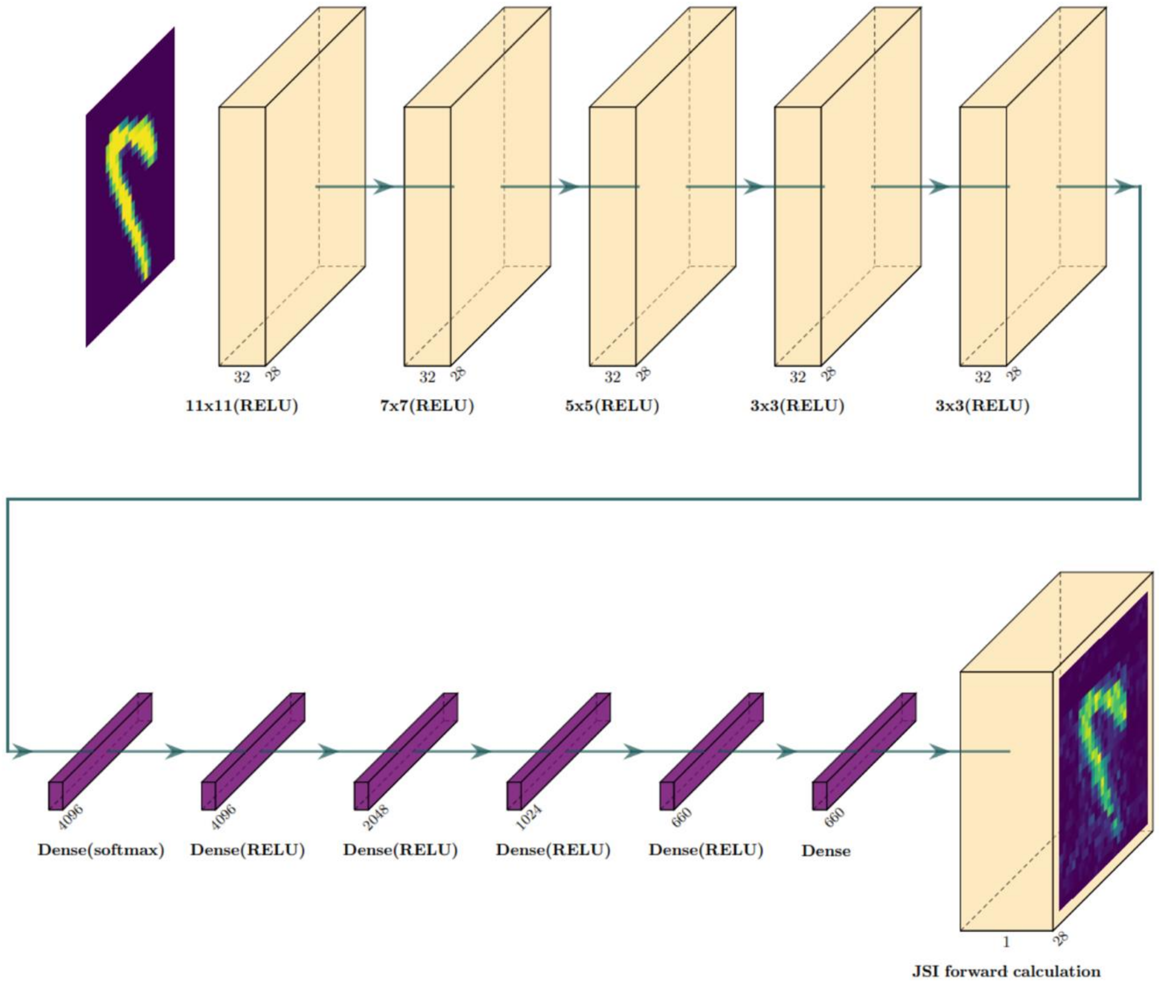
The gradient descent optimization is flexible and accurate

It is computationally expensive! It takes a few hours to generate a set of optimized parameters

Another concern is the dimension of JSI matrix

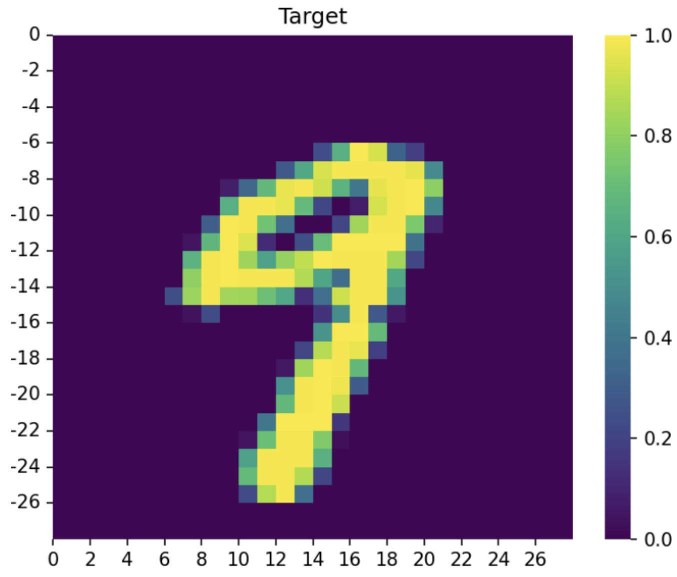
CNNs provide a lighter and more efficient approach

Architecture of the optimization CNN

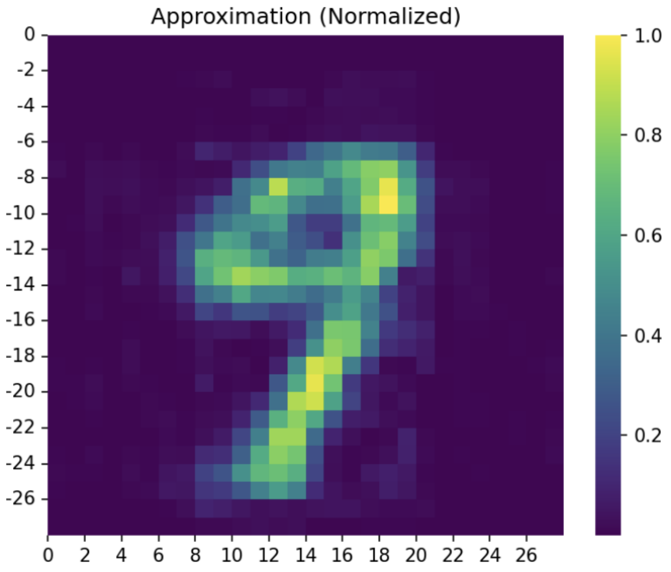


- Starts with AlexNet
- No max-pooling
- Used softmax for the first dense layer (RELU for the other layers)
- Using forward algorithm → generate 100,000 (parameter, JSI) pairs as dataset to train the network
- Then integrate our forward JSI generation algorithm into the loss function. The inferred parameters are fed into the forward algorithm to compute the loss function

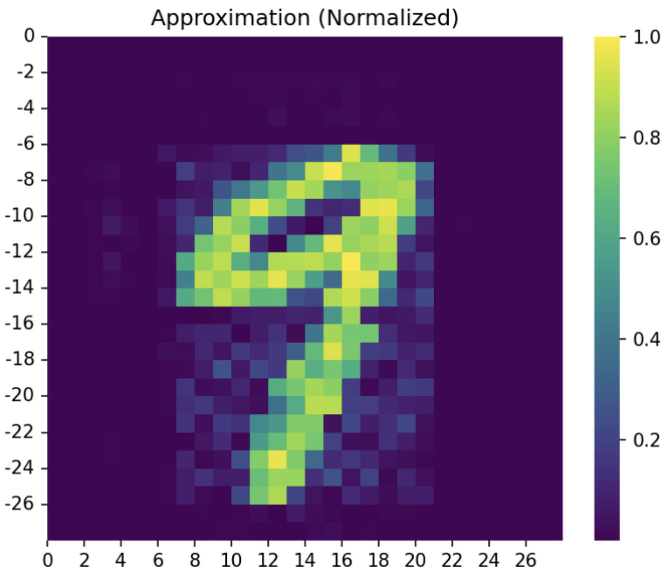
Comparison



(a) Target image



(b) Neural network



(c) Gradient Descent