1 Syllabus

Machine learning and quantum computing have emerged as leading technologies of the twenty-first century and are expected to be increasingly applied to address a wide variety of chemical and materials science challenges. The goal of this course is to introduce fundamental concepts of machine learning and quantum computing to chemists and materials science students through an overview of algorithms, computational methods, and applications. It is intended to empower students to engage with this emerging field and foster the growing field of artificial intelligence for accelerated scientific discoveries in the molecular and physical sciences.

Textbooks. Recommended textbooks for this class are:

R1 "Pattern Recognition and Machine Learning" by Christopher M. Bishop (Springer, 2006). (pdf) (matlab)
R2 "Deep Learning" by Ian Goodfellow, Yoshua Bengio and Aaron Courville. (pdf) (github).
R3 "Deep Learning for Coders with Fastai and PyTorch: AI Applications Without a PhD" by Jeremy Howard and Sylvain Gugger (github)
R4 "Dive into Deep Learning" by Jeremy Howard and Sylvain Gugger. (pdf)
R5 "Quantum Computation and Quantum Information" by Michael A Nielsen and Isaac L. Chuang (Cambridge)
R6 "An Introduction to Quantum Computing" by Phillip Kaye, Raymond Laflamme and Michele Mosca (Oxford University Press), (pdf)
R7 "Learn Quantum Computation Using Qiskit" and notebook.

Our lecture notes will be updated according to the pace of the course and suggestions from the students. References to the textbooks listed above are indicated in the notes as follows: [R1(190)] indicates “from Reference 1, Page 190.”

Pytorch Tutorials and Documentation: Pytorch tutorials and Pytorch documentation will be essential for actual implementations, complemented with tutorials for understanding and implementing sequence-to-sequence (seq2seq) models.

TensorFlow Tutorials and Documentation: The TensorFlow tutorials are written as Jupyter notebooks and run directly in Google Colab.

Scikit-learn Tutorials and Documentation: Scikit-learn is an open source machine learning library that supports supervised and unsupervised learning. It also provides various tools for model fitting, data preprocessing, model selection and evaluation, and many other utilities.

Qutip Tutorials and Documentation: Qutip tutorials will be useful for simulations of light in cavities for quantum information.

Contact Information
Office hours will be held by zoom (Monday, 5:00pm). Zoom ID: 943 8610 8716, Passcode: victor

Grading: There will be no final exam for this class. The final grading evaluation is the same for both undergraduate and graduate students: homework (40%), two mid-terms (40%) on 2/23 and 3/23, and a final project (20 %). Homework will be assigned during lectures and also through Yale canvas. Computer assignments given on Thursdays will be due the following Thursday 9:00 am, and will be assisted for questions during office hours on Monday.
2 Feedforward Neural Networks

Figure (1) shows a simple example of a model neural network (NN), a so-called two-layer feedforward NN based on 3 lists of numbers \(x = [x_0, x_1, \cdots, x_D]\), \(z = [z_0, z_1, \cdots, z_M]\), and \(y = [y_1, \cdots, y_k]\) and adjustable parameters \(\omega_{jk}^{(1)}\) and \(\omega_{jk}^{(2)}\). The \(x\) values are inputs that define the values \(z\) in the so-called hidden layer of neurons, as follows:

\[
z_j = \phi_1 \left( \sum_{k=0}^{D} \omega_{jk}^{(1)} x_k \right).
\]

Here, \(\phi_1(r)\) is the so-called activation function that could be linear (e.g., \(\phi_1(r) = r\)) or non-linear (e.g., \(\phi_1(r) = r, \text{for } r \geq 0\) and \(\phi_1(r) = 0\) for \(r < 0\)), as discussed later in Sec. 2.3. The outputs \(y\) are computed analogously,

\[
y_j = \phi_2 \left( \sum_{k=0}^{M} \omega_{jk}^{(2)} z_k \right).
\]

The resulting NN can then be parametrized by adjusting the values of the weights \(\omega_{jk}^{(1)}\) and \(\omega_{jk}^{(2)}\), with \(x_0 = z_0 = 1\), so that for any given input \([x_1, \cdots, x_D]\) we can predict the corresponding output \([y_1, \cdots, y_k]\), where the weight \(\omega_{j0}^{(1)}\) is often called the bias of layer \(l\).

Examples for molecular systems could have inputs \(x\) defined in terms of the atomic coordinates of a molecule while the outputs \(y\) could be the potential energy, \(y_1 = V(x)\), and forces \(y_k = F_k(x)\) acting on each atom (as in typical force-fields employed in molecular dynamics, or Monte Carlo simulations). Another NN could have inputs \(x_k\) defined in terms of [Hammett parameters] for the substitutional groups of a molecule, and the outputs \(x_k\) could be the energies of the frontier orbitals \(y_1 = E_{\text{HOMO}}, y_2 = E_{\text{LUMO}}\) of that molecule. Alternatively, the NN could have input numbers \(x_k\) defining the name of a molecule, or the primary sequence of a protein, according to some dictionary or catalog while the outputs \(y_k\) could be the properties of the molecule (e.g., solubility, molecular weight, toxicity, NMR chemical shifts).

A simple way to parametrize the NN shown in Fig. 1 is to adjust the values of \(\omega_{jk}^{(1)}\) and \(\omega_{jk}^{(2)}\) to minimize the error between the predicted and correct values of the outputs (often called the loss) for a so-called training set of inputs for which the outputs are known,

\[
\epsilon = \sum_{j=1}^{K} (y_j - y_j^{(a)})^2,
\]

Figure 1: Network diagram for a two-layer feedforward neural network. The inputs \(x_j\), hidden units \(z_j\), and output variables \(y_j\) are represented by nodes (artificial neurons), and the links between the nodes represent the weight parameters \(\omega_{jk}\). The neurons \(x_0 = z_0 = 1\) are kept constant and linked to other neurons with weights \(\omega_{j0}^{(1)}\) and \(\omega_{j0}^{(2)}\) called bias parameters. Arrows denote the direction of information flow through the network during forward propagation [R1(228)].
where $y_j$ are the calculated values of the outputs, and $y_j^{(a)}$ are the correct values (or labels) corresponding to a given input training set. Note that $\epsilon$ is a function of the parameters weights $\omega_{jk}^{(1)}$ and $\omega_{jk}^{(2)}$ since, according to Eqs. (1) and (2),

$$y_j = \phi_2 \left( \sum_{k'=0}^{M} \omega_{jk'}^{(2)} \phi_1 \left( \sum_{i=0}^{D} \omega_{k'i}^{(1)} x_i \right) \right). \quad (4)$$

So, we can minimize $\epsilon$ by adjusting the parameters $\omega_{jk}^{(l)}$, as described below in Sec. 2.1.

### 2.1 Gradient Descent

The parameters of NNs are typically adjusted by using the gradient descent method, illustrated in Fig. 2. To visualize the minimization process, we consider that the weights $\omega_{jk}$ are functions of time $t$ along the optimization process that advances with small integration time steps $\tau$. Therefore, the loss $\epsilon$ decreases along the optimization time by evolving with values $\epsilon(t)$, $\epsilon(\tau)$, $\epsilon(2\tau)$, $\cdots$, $\epsilon(N\tau)$, with $N\tau$ the total optimization time.

Given an initial set of randomly assigned values, the adjustable parameters $\omega_{jk}^{(l)}$ are updated along the direction of minus the gradient of the loss with respect to the parameters,

$$\frac{\partial \omega_{jk}^{(l)}}{\partial t} = -\frac{\partial \epsilon}{\partial \omega_{jk}^{(l)}}, \quad (5)$$

where $s$ parametrizes the evolution of parameters, as follows:

$$\omega_{jk}^{(l)}(t + \tau) = \omega_{jk}^{(l)}(t) - \frac{\partial \epsilon}{\partial \omega_{jk}^{(l)}} \tau, \quad (6)$$

with the learning rate parameter $\tau$ defined as a small positive number.

Such a choice of gradients ensures that $\epsilon$ decreases monotonically since

$$\epsilon(t + \tau) = \epsilon(s) + \sum_{jkl} \frac{\partial \epsilon}{\partial \omega_{jk}^{(l)}} \frac{\partial \omega_{jk}^{(l)}}{\partial t} \tau,$$

or

$$\epsilon(t + \tau) = \epsilon(t) - \sum_{jkl} \left| \frac{\partial t}{\partial \omega_{jk}^{(l)}} \right|^2 \tau. \quad (7)$$
2.1.1 Stochastic Gradient Descent

The minimization of the error between the predicted and correct values of the outputs, for a training set of \( K \) inputs requires the evolution of the weights \( \omega_{ik}^{(l)} \), according to Eqs. (6) and (3), as follows:

\[
\omega_{ik}^{(l)}(s + \Delta s) = \omega_{ik}^{(l)}(s) - \frac{\partial e}{\partial \omega_{ik}^{(l)}} \Delta s,
\]

which is the so-called standard (or 'batch') gradient descent, corresponding to the batch of \( K \) outputs. However, when \( K \) is very large, evaluating the sums of gradients becomes very expensive.

The stochastic gradient descent (SGD) reduces the computational cost at every iteration by randomly sampling a subset of \( j \) output values at every step (effectively applying dropout on the outputs). Therefore, the key difference compared to standard gradient descent is that only a portion of the output data is used to approximate the gradient of the loss, and that portion is picked randomly at each step.

2.1.2 Exercise: Gradient Descent

To illustrate the gradient descent method as applied to optimize the parameters of a model, optimize the linear coefficients of a third-order polynomial

\[
y_{pred}(x) = \sum_{j=0}^{3} c_j x^j,
\]

to approximate the output \( y(x) = \sin(x) \) for an input of 2000 equally spaced values of \( x \) in the range \( x = [-\pi, \pi] \).

Go through the Pytorch tutorials [learning with examples] and see how to formulate the linear regression problem in terms of the parametrization of a 2-layer feedforward linear NN.

Solution: Download the Jupyter notebook for the solution from [Exgradientdescent.ipynb] where you can have a first exposure of the concepts that we will discuss in subsequent sections and the basic aspects of Pytorch, including:

- The concept of a PyTorch tensor
- How to define a neural network with learnable parameters (or weights).
- Forward propagation: How to process the input through the network
- How to compute the loss (how far the output is from being correct)
- Backward propagation: How to compute the gradients of the loss w.r.t. the parameters of the network.
- How to update the weights of the network, typically using a simple update rule: weight = weight - learning_rate.
2.2 Colab, Python, Tensorflow, Keras and Pytorch

Several software libraries and computational packages are currently available for high level implementations of machine learning methods, including [Pytorch], [torch], [tensorflow], and [keras]. A comparison of these three popular deep learning frameworks can be found [here]. Pytorch is the newer framework (based on Torch and open-sourced on GitHub in 2017 by Facebook’s AI research group). Pytorch’s popularity is rapidly growing among AI researchers due to simplicity, flexibility, efficient memory usage, speed, and dynamic computational graphs.

For this class, I recommend working in the [Google Colab] environment, so you can run your codes in Google’s computers. In Colab, everything you need is already installed, or you can upload by mounting your Google drive as shown in the [Navigating tutorial.ipynb] Jupyter notebook for which you need to have the following [csv file] in the same folder. A brief (and fun) tutorial on how to work with RDKit at Google Colab is available at the [RDKit Mendeleev.ipynb] notebook.

These notebooks also introduce the SMILES notation for representation of molecules in terms of strings of characters, which can be converted into lists of numbers to input molecules into neural networks.

The SMILES notation follows the following rules (examples shown in Fig. 3):

1. Atoms are represented by their atomic symbols.

2. Hydrogen atoms are omitted (are implicit).

3. Neighboring atoms are represented next to each other.

4. Double bonds are represented by ‘=’, triple bonds by ‘#’.

5. Branches are represented by parentheses.

6. Rings are represented by allocating digits to the two connecting ring atoms.

7. Aromatic rings are indicated by lowercase letters.

Molecules can also be represented by features defined according to the type of atoms and their corresponding neighborhoods in the molecular structure, using the extended-connectivity fingerprints (ECFPs), also known as circular fingerprints available at DeepChem. The main properties of ECFPs are that (i) they are defined by considering circular atom neighborhoods (Fig. 15, middle panel, Sec. 5); (ii) they are rapidly calculated; (iii) they represent substructures; (iv) they can account for a huge number of different molecular features (including stereochemical information); and (v) they represent both the presence and absence of functionality.

I also recommend brushing up your python knowledge with the excellent tutorial provided in Sec. 21, adapted by Kevin Zakka for the Spring 2020 edition of cs231n and available at [Python_tt.ipynb].
2.3 Activation Functions

Figure 4 shows some of the most commonly used activation functions which are essential building blocks of neural networks.

2.3.1 Linear Activation

The so-called ‘Adaline’ activation, or linear regression function, produces only linear models (i.e., model NN without hidden layers), since the output is always a simple linear combination of the input data,

\[
y_j = \sum_{i=0}^{M} \omega_{ji}^{(2)} \sum_{i=0}^{D} \omega_{ik}^{(1)} x_k
\]

so even models with multiple hidden layers with linear activation functions are equivalent to models without hidden layers. Therefore, to go beyond simple linear regression models, it is imperative to include non-linear activation functions (i.e., functions \( \phi(z) \) with non-constant derivatives).

2.3.2 Non-linear Activation

Two representative examples of non-linear activation functions, listed in Fig. 4 are the Logistic sigmoid and the ReLU functions. The main difference between these two activation functions is that the derivative of the sigmoid \( \phi(z) = 1/(1 + e^{-z}) \) is smaller than 1, in fact \( \phi'(z) \leq 1/4 \) since \( \phi'(z) = e^{-z}/(1 + e^{-z})^2 \), while the derivative of the ReLU is always equal to 1 when \( z > 0 \). As we discuss later in Sec. 2.3.3, activation functions with \( \phi'(z) < 1 \) can be problematic since they lead to the so-called vanishing gradient problem. As shown in Sec. 2.3.3, the gradients of the loss with respect to the weights are proportional to the product of the derivatives of the activation functions, and multiplying derivatives of activation functions that are \( < 1 \) leads to gradients of the loss that approximately equal to zero. So, the parameters of the NN cannot be adjusted and the NN cannot be trained. A solution to that problem has been the implementation of the ReLU, \( \phi(z) = \max(0, z) \), and Leaky ReLU, \( \phi(z) = \max(0.01, z) \), activation functions that overcome the vanishing gradient problem since their derivatives are always equal to 1 when \( z > 0 \). So, the ReLU function has enabled training of NNs with many layers and thus the emergence of the field of deep-learning briefly described for image recognition in the following [YouTube]. The softplus function is essentially a smooth ReLU activation (i.e., with continuous gradients).
An important observation is that non-linear activation functions, such as ReLU, introduce correlations between different inputs since all of the inputs of a neuron combined determine whether its argument is $z > 0$. So, the activation correlates the various inputs.

### 2.3.2.1 Classification and Non-linear Regression Problems

Supervised learning problems involved training data sets that include both the inputs and the corresponding labels. Pattern recognition of a class or type (e.g., type of amino acid, class of organic molecule, etc.) in which the aim is to assign each input one of a finite number of discrete categories (e.g., one of the 20 types of natural amino acids, or one of the various types of organic molecules, etc.) are called classification problems. In contrast when the desired output consists of continuous variables (e.g., the solubility value, the NMR chemical shift value, etc., based on the input molecular structure), then the task is called regression. Other examples of regression could be the prediction of the yield of a chemical reaction based on inputs corresponding to the concentrations of reactants, the temperature, and the pressure [R1(3)]. Non-linear activation functions are essential for both non-linear regression and classification problems.

#### Logistic Sigmoid:
Nonlinear activation functions are typically used for output layers of classification problems. For example, the logistic sigmoid function $\phi(z_k) = 1/(1 + e^{-z_k})$ is typically used for binary classification (e.g., is the molecule toxic or not, is it flammable or not, etc.), predicting one class when $\phi > 0.5$ and the other when $\phi \leq 0.5$.

#### Softmax:
The softmax activation function $\phi(z_k) = \exp(z_k) / \sum_j \exp(z_j)$ is a multiclass generalization of the logistic sigmoid that outputs a normalized probability distribution over the predicted output classes (e.g., 97% probability that the molecule is an aldehyde, 2% probability that is a ketone, 1% that is an alcohol). The softmax function is typically used in conjunction with loss functions such as the KL divergence or cross-entropy, described below. For classification problems, the loss defined by the cross-entropy or the KL divergence allow for faster training than the sum-of-square differences as well as improved generalization [R1(235)].

#### KL Divergence and Cross-Entropy:
The Kullback-Leibler (KL) divergence is defined, as follows:

$$KL(p\|\phi) = - \sum_k p(z_k) \log_2(\phi(z_k) / p(z_k)) = H(p, \phi) - H(p),$$

where $H(p) = - \sum_k p(z_k) \log_2(p(z_k))$ is the entropy corresponding to the target probabilities $p(z_k)$, and

$$H(p, \phi) = - \sum_k p(z_k) \log_2(\phi(z_k)).$$

is the cross entropy. Both functions allow for comparisons of two discrete probability distributions and are commonly used for training models to produce outputs corresponding to a target distribution. Note that minimizing the KL divergence corresponds exactly to minimizing the cross-entropy since the entropy $H(p)$ of the target distribution does not depend on the adjustable weights.

#### 2.3.2.1.1 Optional Exercise:
Use the bound $\log(x) \leq x - 1$ to show that the KL divergence is always $\geq 0$ (Gibbs Inequality). Therefore, the cross-entropy is always larger or equal than $H(p)$, and equal to $H(p)$ when the two distributions are the same.
2.3.3 Vanishing Gradient Problem

The gradients of the loss $\epsilon$ with respect to the adjustable parameters $\omega_{jk}^{(l)}$, introduced in Sec. 2.1, are computed according to the chain rule:

$$\frac{\partial \epsilon}{\partial \omega_{jk}^{(l)}} = \sum_{i=1}^{K} \frac{\partial \epsilon}{\partial y_i} \frac{\partial y_i}{\partial \omega_{jk}^{(l)}}.$$  \hfill (12)

For example, according to Eq. (4),

$$\frac{\partial y_i}{\partial \omega_{jk}^{(1)}} = \frac{\partial}{\partial \omega_{jk}^{(1)}} \phi_2 \left( \sum_{k'=0}^{M} \omega_{k'k}^{(2)} \phi_1 \left( \sum_{l=0}^{D} \omega_{k'l}^{(1)} x_l \right) \right),$$

$$= \frac{\partial \phi_2(s)}{\partial s} \frac{\partial}{\partial \omega_{jk}^{(1)}} \sum_{k'=0}^{M} \omega_{k'k}^{(2)} \phi_1 \left( \sum_{l=0}^{D} \omega_{k'l}^{(1)} x_l \right),$$

$$= \frac{\partial \phi_2(s)}{\partial s} \frac{\partial \phi_1(s)}{\partial s} \omega_{ij}^{(2)} x_k.$$  \hfill (13)

Therefore, the gradient of the loss with respect to $\omega_{jk}^{(1)}$ is obtained, as follows:

$$\frac{\partial \epsilon}{\partial \omega_{jk}^{(1)}} = \frac{\partial \phi_2(s)}{\partial s} \frac{\partial \phi_1(s)}{\partial s} \sum_{i=1}^{K} \frac{\partial \epsilon}{\partial y_i} \omega_{ij}^{(2)} x_k,$$  \hfill (14)

which is proportional to the product of two gradients of activation functions. Analogously, for a NN with 3 layers (i.e., 2 hidden layers) we would have gradients defined, as follows:

$$\frac{\partial \epsilon}{\partial \omega_{jk}^{(1)}} = \sum_{i=1}^{K} \frac{\partial \epsilon}{\partial y_i} \frac{\partial y_i}{\partial \omega_{jk}^{(1)}},$$

$$= \frac{\partial h_3(s)}{\partial s} \frac{\partial \phi_2(s)}{\partial s} \frac{\partial \phi_1(s)}{\partial s} \sum_{i=1}^{K} \frac{\partial \epsilon}{\partial y_i} \sum_{k'=0}^{M} \omega_{k'k}^{(2)} \sum_{l=0}^{D} \omega_{k'l}^{(1)} x_k,$$  \hfill (15)

which are clearly proportional to the product of 3 gradients of activation functions and make the gradient of the loss very small when the gradients of the activation functions are $< 1$. Therefore, deep neural networks with multiple hidden layers (Fig. 5) typically rely on ReLU or Leaky ReLU activation functions that enable efficient training since they do not suffer from the vanishing gradient problem. Fig. 5 also shows that a 2-dimensional array of data, corresponding to the 28 × 28 intensities of pixels of hand written numbers can be vectorized to reshape it as a 1-dimensional array of 784 neurons.

Figure 5: Dense neural network with multiple hidden layers, with input neurons corresponding to the pixels intensities of hand written numbers reshaped by vectorization into a 1-dimensional array.
2.3.4 Validation, Cross Validation and Bootstrapping

Validation of the model is essential to ensure that the NN works well on data that has not been used during the training process. Therefore, it is important to keep aside a portion of the data that is not used for training, and use it for testing and validation to ensure that the loss of the training and testing sets are comparable.

A simple approach is to split it into 70:30, as shown in Fig. 6 (left), which is a fine procedure when there is enough data so long as the split ensures that the training and testing samples have the same distribution (e.g., randomized).

When we have limited data, the simple splitting procedure described above might introduce bias in the parametrization of the NN since the training set might miss some key sample points. So, neither the training set distribution nor the testing set might be representative of the original data set. Therefore, with limited data, it might be necessary to implement cross-validation.

The k-fold cross validation method, represented in Fig. 6 (right panel), is a popular resampling procedure that generates a less biased model even when having limited data since it ensures that every data point from the original dataset is included in the training and testing sets. Rather than splitting the data into 70:30, the data is split randomly into k folds (k is typically between 5-10). At each iteration, k − 1 folds are used for training while the remaining k − th fold is used for validation. The process is repeated until every fold has been included in the testing set. Finally, all k results are averaged (a process often called bagging).

Bootstrap sampling is another method commonly used when there is a limited amount of data to generate models with less bias and less overfitting, a problem discussed in Sec. 2.6.2. Bootstrap sampling also enables quantification of uncertainty of the predictions. Similarly to cross-validation, bootstrap sampling generates multiple data sets as follows [R1 (23)]. Suppose our original big data set consists of n data points \( X = x_1, ..., x_n \) (Fig. 7). We can create a new data set \( X_B \) by drawing n points at random from \( X \), with replacement, so that some points in \( X \) may be replicated in \( X_B \), whereas other points in \( X \) may be absent from \( X_B \). The process is repeated N times to generate N data sets, each of size n and each obtained by sampling from the original data set \( X \). Predictions from all data sets are averaged (a process called bootstrap aggregation, or bagging) and the uncertainty of the prediction is estimated with the standard deviation.
2.3.4.1 Exercise: Bootstrapping  To see how bootstrapping can reduce the variance of a prediction by simply averaging the results of individual estimators, compute the distribution of means obtained by rolling a fair dice (i.e., with uniform distribution for the six possible outcomes) 100 times (i.e., N=100).

- Show that the distribution of outcomes is uniform, since the dice is fair. Obtain the mean of outcomes, which is close to 3.5 and standard deviation σ which is about 1.7.

- Compute the distribution of averages obtained by repeating that calculation 10,000 times and show that the distribution of means is a Gaussian with the same mean and a standard deviation σ = σ / √N = 0.17.

This is a demonstration of the Central Limit Theorem at work. The theorem states that the distribution of a sufficiently large number of means obtained with N samples drawn with replacement from any arbitrary distribution with variance σ², is a Gaussian with variance σ² / N.

Solution: Download the Jupyter notebook with the solution from CLT_dice.ipynb.

2.4 Tutorial Assignment on Hammett Neural Networks with Keras/TensorFlow

[This tutorial assignment has been designed and developed by Jessica Freeze]

The tutorial on how to build and execute a Hammett neural network with Keras/TensorFlow for prediction of frontier orbital energies of tungsten-benzylidyne catalysts using Hammett parameters as input descriptors can be downloaded as a notebook:

- Assignment_1_NeuralNetworks.ipynb
- Assignment_1_NeuralNetworks.pdf

2.5 Tutorial Regressive Models for Chemical Predictions with Scikit-Learn

[This tutorial assignment has been designed and developed by Jessica Freeze]

The tutorial on how to implement regressive models with Scikit-Learn for chemical properties predictions can be downloaded as a notebook:

- Assignment_2_RegressiveLinearModelsForChemistryPrediction.ipynb
- Assignment_2_RegressiveLinearModelsForChemistryPrediction.pdf

2.6 Prediction of Molecular Toxicity by Linear Classification with DeepChem

A turn-key tutorial on how to make predictions of molecular toxicity with respect to 12 different assays, using a classification neural networks with DeepChem, can be downloaded as a notebook:

- 04_Molecular_Fingerprints.ipynb
- 04_Molecular_Fingerprints.pdf

The 2-layer NN is trained with Tox21, a database that contains information about the toxicity of molecules with respect to 12 different assays. The input molecular features are defined according to the type of atoms and their corresponding neighborhoods in the molecular structure, using the extended-connectivity fingerprints (ECFPs), also known as circular fingerprints available at DeepChem. The main properties of ECFPs are that (i) they are defined by considering circular atom neighborhoods (Fig. 15, middle panel, Sec. 5); (ii) they are rapidly calculated; (iii) they represent substructures; (iv) they can account for a huge number of different molecular features (including stereochemical information); and (v) they represent both the presence and absence of functionality.
2.6.1 Training

Figure 8 illustrates the typical iterative procedure implemented for training of neural networks. The training data is typically divided into $N$ batches. The training process is initiated with the first batch of data that is input for forward propagation, a process that computes the values of neurons in the hidden and output layers, using the initially given set of weights (often randomly chosen weights). The resulting output values are then compared to the actual labels of the input data to compute the loss, introduced by Eq. (3), and the gradient of the loss with respect to the weights, according to Eq. (12). The gradients are then used for back propagation, the process that updates the weights by the gradient descent optimizer. Having updated the weights, the second batch is input and the forward and backward propagation steps are applied. The process is repeated for each batch until the first epoch is completed. Then, the loss is computed and if it is not sufficiently low the weights are further optimized by processing more epochs.

2.6.2 Overfitting Problem

The design of neural networks is usually based on previously developed successful neural networks or the result of work based on trial and error to tune the hyperparameters that define number of layers, number of neurons per layer, and the types of layers. The neural network is designed to achieve minimum loss for both the training and validation/testing data. The loss of the training set can always be minimized by increasing the complexity of the network, increasing the number of layers and number of neurons per layer. However, increasing too much the number of adjustable parameters can lead to the problem of overfitting where increasing the complexity reduces the loss of the training set but increases the loss of the validation data set, as shown in Fig. 9. By trial an error, one can increase the complexity and reach the ideal complexity range where the error for the validation/testing data is minimum, making sure not to fall into overfitting range where the error for the testing data set increases.
2.6.3 Regularization

[Dropout] is a simple [regularization] method, commonly used for reducing the number of adjustable parameters in artificial neural networks. Randomly chosen weights are simply omitted or ‘dropped out’ by zeroing them during the training process, as shown in Fig. 10. The technique is also called random pruning and belongs to the family of dilution methods (i.e., methods based on adding damping noise to the parameters of the model).

Other regularization techniques are typically included in most optimization methods used for parametrization of neural networks, including $L_2$ regularization (called Ridge Regression) where the loss includes an additional term to penalize according to the sum of the squares of the weights. For example, for the 2-layer NN introduced in Sec. 2, the loss for Ridge regression would be:

$$
\epsilon = \sum_{j=1}^{K}(y_j - y_j^{(a)})^2 + \lambda \left( \sum_{i,k}(\omega_{ik}^{(1)})^2 + \sum_{i,k}(\omega_{ik}^{(2)})^2 \right).
$$

$L_1$ regularization (also called Lasso Regression), by adding a term to the loss proportional to the sum of absolute values of the weights:

$$
\epsilon = \sum_{j=1}^{K}(y_j - y_j^{(a)})^2 + \lambda \left( \sum_{i,k} |\omega_{ik}^{(1)}| + \sum_{i,k} |\omega_{ik}^{(2)}| \right).
$$

The main difference between Ridge and Lasso regressions is that Lasso leads to sparse representations (with more coefficients equal to zero), since the regularization term introduces an aggressive gradient of the loss with respect to small coefficients, as defined by $\lambda$. In contrast, the corresponding gradient in Ridge regression is $2\lambda\omega_{jk} - i.e.,$ proportional to the small value of the coefficients $\omega_{jk}$.

Note that the regularization terms added to the loss, transform the problem into a more constrained problem with fewer possible solutions, eliminating for example solutions with large oscillations, as shown in Fig. 9, allowing us to find a simple solution. In general, we look for simple solutions since data collected from systems is usually very simple and the experiments correspond to weak perturbations leading to an observable response that is very simple either linear in the perturbational field or of very low order.

![Figure 10: Top: Dilution by ‘dropout’, a technique that reduces the number of adjustable parameters by simply omitting randomly chosen weights (i.e., links), dropping them out of the training process. Bottom: Representation of complex solutions, obtained by overfitting with dense neural networks (left), as compared to simple solutions obtained by regularization.](image)
3 Clustering and Regression Algorithms

3.1 Random Forest

Random forest is a recursive algorithm for classification of data into subgroups, according to selected features of the data, using a 'forest' of decision binary trees (Fig. 11). Therefore, the algorithm involves an unsupervised process that finds out the subgroups of data samples that share common properties, and creates decision trees to find out which subgroup of samples and corresponding properties can be assigned to an unknown sample.

A simple example is a data set of molecules that we might want to classify into subgroups of certain toxicity, solubility, etc., by analyzing their features, and at the same time exploit the classification process to build decision trees that would enable classification of a molecule with unknown properties by running it through the decision trees and finding out in which leaf of the tree the molecule ends up branching out.

Each tree is created by using a subset of the training data set (sampled with replacement), according to an iterative procedure defined by the following steps:

1. Randomly choose \( n \) features of the data to be analyzed as potentially splitting features, according to steps 2 and 3.

2. For each feature, find the splitting point that minimizes the entropy (or, linearized entropy, called Gini impurity as discussed in Sec. 3.1.1).

3. Split the data into 2 subgroups, using the feature with best performance.

4. Stop, if stopping criterion is met (e.g., the entropy decrease is smaller than a given threshold, or some other stopping criteria is met such as the maximum depth of the tree has been reached, or the number of samples in the node is smaller than a minimum value, etc., to ensure that the tree does not classify noise). Otherwise, goto 1 and repeat the process for the generated subgroups.

Clearly, the algorithm subdivides the data into subgroups, each of which with more uniform population (i.e., more pure in a certain type of feature) than the complete subset of data used for construction of the tree. The purity of the subgroups can then be exploited to classify a sample of unknown type by running it through the decision tree and assigning to its type, the type of data corresponding to the subgroup where the sample branched out. Running the unknown sample through all of the trees of the forest and averaging the results from all trees gives the ensemble average sample classification either as a majority vote, or otherwise a weighted vote based on the level of confidence for the prediction from each tree.

Figure 11: Schematic representation of the forest of binary decision trees generated by the random forest algorithm, and the outcome prediction for a testing sample (blue).

\[ ^1 \text{Alternatively, one can simply choose a random splitting point (a variation of the method called extremely randomized trees).} \]
As described above, the random forest algorithm is a random subspace method for ensemble learning where correlations between prediction estimators (trees) are suppressed by training the trees on random samples of data and random selection of features instead of using entire feature set. The bootstrap aggregation, or bagging, combines the predictions produced by several learners into an ensemble that performs better than the original learners.

3.1.1 Entropy and Gini

The entropy of a tree node $k$ with $N_k$ elements of $N_t$ possible types is defined, as follows:

$$S_k = - \sum_{j=1}^{N_t} p_{jk} \log_2(p_{jk}),$$

where $p_{jk} = n(j,k)/N_k$ is the likelihood that an element of node $k$ is of type $j$, as defined by the number $n(j,k)$ of elements of type $j$ in node $k$ over the total number $N_k$ of elements of that node.

The so-called Gini impurity measure $G_k$ of node $k$ is defined as the linearized entropy, obtained by approximating $\log_2(p_{jk})$ by its first order expansion around $p_{jk} = 1$ (i.e., using the expansion $\log_2(p_{jk}) \approx \log_2(1) + (p_{jk} - 1) + \cdots = p_{jk} - 1$): $G_k = -\sum_{j=1}^{N_t} p_{jk}(p_{jk} - 1) = 1 - \sum_{j=1}^{N_t} p_{jk}^2$. Note that $G_k = S_k = 0$ when all of the elements of node $k$ are of a single type $i$ (i.e., the composition is pure in type $i$). In addition, both $S_k$ and $G_k$ are maximum when all the possible types have equally probable, (i.e., with probability $p(j|k) = 1/N_t$), and the composition is maximally impure. Therefore, both $S_k$ and $G_k$ are equally valuable as metrics of impurity. In practice, however, $G_k$ is more popular since it is less expensive to compute.

3.2 K-means Algorithm

The k-means clustering algorithm classifies the samples of a data set into $K$ subgroups of elements that happen to be closer to the centroid of their cluster than to the centroids of any of the other $K - 1$ clusters, as defined by a given measure of distance.

The k-means clustering procedure can be described, as follows. Starting with the complete data set, $K$ centroids are defined at random in the space of features,

1. Assign samples to their corresponding nearest centroid.
2. Recompute the centroid of each cluster according to their own samples.
3. Stop, if stopping criterion is met (e.g., the composition of each cluster has not changed). Otherwise, goto 1.

The number of clusters that defines the most unbiased classification, as determined by the composition of the clusters, is typically the number of clusters that maximizes the entropy $S_K = -\sum_{j=1}^{K} \sum_{i=1}^{N_t} p_{ij} \log_2(p_{ij})$, where $p_{ij} = n(i,j)/N_j$ is the likelihood that an element of cluster $j$ is of type $i$, as defined by the number $n(i,j)$ of elements of type $i$ in cluster $j$ over the total number of elements $N_j$ in that cluster. The maximum entropy principle gives the most unbiased distribution since an inference made on the basis of incomplete information should be drawn from the probability distribution that maximizes the entropy, subject to the constraints on the distribution.
3.3 **K-Nearest Neighbors Algorithm**

The [K-nearest neighbors](#) (KNN) algorithm is a classification method based on the plurality vote of the $K$ nearest neighbors, in the space of features (K-NN classification). It can also be used as a regression method based on the average of the properties of the $K$ nearest neighbors (K-NN regression). The average, or the vote, can also be weighted inversely proportionally to the distances of the sample to each of its $K$ nearest neighbors. The underlying assumption of the method is that samples that are close together in the space of features have similar properties.

3.4 **Unsupervised Classification Assignment: K-means and Random Forest**

*[This tutorial assignment has been designed and developed by Jessica Freeze]* The assignment for unsupervised classification can be downloaded as a notebook [UnsupervisedClassification.ipynb] or [UnsupervisedClassification.pdf]
Regularization can also be accomplished by convolution as shown in Fig. 12 for the convolution of a $7 \times 7$ neural layer, using a $3 \times 3$ convolutional kernel of weights $\omega_{k,j}$. Starting with the kernel placed at the top-left corner of the input layer, the input neurons overlapping with the kernel are multiplied with the corresponding kernel weights and the products are summed to generate the value of the neuron $z'_{i,j}$ at the convoluted layer, aligned with the center of the kernel:

$$z'_{i,j} = \phi \left( \sum_{k=1}^{3} \sum_{l=1}^{3} z_{i+k,j+l} \omega_{k,l} \right),$$  

where $\phi$ is an activation function responsible for the so-called ‘detector stage’. The kernel is then displaced to overlap with other neurons in the input layers and the same weights of the convolution kernel $\omega_{k,j}$ are used to generate the value of another neuron in the convoluted layer, according to Eq. (17).

The process is then repeated until the whole input layer is convoluted, as seen by clicking in the animation of the lower panel of Fig. 12. Convolution is often followed by a pooling layer that ‘summarizes’ sections of the convoluted layer into a single output value for example by computing the average of the values of the convoluted layer, or the maximum value (maxpooling). Pooling is often applied since it makes the output invariant with respect to global changes in the same image such as translation or rotation of the input in applications to image processing.

When compared to feedforward NN layers, CNN offer significant advantages for enhanced performance since they typically include fewer links between neurons (i.e., sparsity), and fewer parameters (i.e., shared weights). Even for the simple example shown in Fig. 12 we note that a feedforward transformation from $7 \times 7$ to $5 \times 5$ layer would require $7 \times 7 \times 5 \times 5$ weights, while the CNN requires only $3 \times 3$ since the same parameters are shared by all of the transformations that generate the convoluted layer. In addition, the CNN introduces sparsity of connectivity, with much fewer links between input and output neurons. Note, for example, that the neuron at the top-left corner of the input layer is linked only to the neuron at the

Figure 12: Top: Convolution of a $7 \times 7$ neural layer, using a $3 \times 3$ kernel with stride=1, padding=0, and dilation=1 to generate a $5 \times 5$ convoluted layer. Middle: Animation of convolution. Bottom: Comparison of convolution layers obtained with $3 \times 3$ kernels, with dilation = 1 (left) and dilation = 2 (right), respectively.
Convolution layers are defined by the hyperparameters of the kernel (parameters defined by the user), shown in Fig. fig:cnn (bottom panel), including the kernel dimensions as defined by the height \( (k_h) \) and width \( (k_w) \) for a 2-dimensional kernel, the stride \( (s) \) (i.e., the step size for striding across the input layer), the padding \( (p) \) (or, border parameter defining whether the kernel stops when it reaches the border or beyond), and the dilation rate \( (d) \) defining the spacing between the elements of the kernel. The resulting dimensions (width \( W_{\text{out}} \) and height \( H_{\text{out}} \)) of the resulting convoluted layer can be computed in terms of the kernel parameters and the width \( W_{\text{in}} \) and height \( H_{\text{in}} \) of the input layer, as follows:

\[
W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2 \times p - d \times (k_w - 1) - 1}{s} + 1 \right\rfloor
\]

\[
H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2 \times p - d \times (k_h - 1) - 1}{s} + 1 \right\rfloor
\]

(18)

When the input has \( C_{\text{in}} \) channels (e.g., pixel intensities for blue, red and green), the kernel can have multiple channels and transform the input into an output layer with a single channel. Further, \( C_{\text{out}} \) kernels can be applied to generate an output layer with \( C_{\text{out}} \) channels, as shown in Fig. 13.

Convolutional layers are particularly suitable for extracting features, as in applications to image processing. For example, the edges of an image can be generated simply by convolution of the image with a \( 3 \times 3 \) kernel with positive and negative values:

\[
k = \begin{bmatrix}
1 & 0 & -1 \\
0 & 0 & 0 \\
-1 & 0 & 1
\end{bmatrix}.
\]

(19)

In addition, convolution has valuable properties for image processing since it commutes with translation (i.e., the shift of an input simply leads to a shifted output) and thus generates similar outputs for similar inputs. In addition, convolutions are local due to the finite scope of the kernel and, therefore, preserve the local structure of the input. Various different convolution kernels are typically applied so that various different features can be extracted.

The development of the CNN [AlexNet], shown in Fig. 14 (top panel), and its celebrated victory in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) has revolutionized the field of deep learning. The network achieved a top-5 performance with 15.3% error for recognition and classification of images of 1000 categories, including about 1,200 image per category. AlexNet, with 5 convolutional layers and 3 fully connected layers, has shown that the depth of the model is essential for high performance. Furthermore, AlexNet laid the foundation for the traditional CNN scheme based on a convolutional layer followed by an activation function followed by a max pooling operation, although the pooling operation is sometimes omitted to preserve the spatial resolution of the image. Training of AlexNet, with 61 million parameters, and 600 million connections, as summarized here was made possible due to the utilization of graphics processing units...
(GPUs), and the feasibility of parallelizing the convolutions based on two pathways representing
the split between two GPUs (Fig. 14 (top panel)).

Since AlexNet, the state-of-the-art in CNN architectures has gone deeper and deeper (i.e., AlexNet has only 5 convolutional layers, while the VGG network has 19 layers, and GoogleNet (also codenamed Inception_v1) has 22 layers) (Fig. 14, middle panel). Developing deeper neural networks has been possible by using residual layers that skip over convolutional layers to avoid the vanishing gradient problem, resulting in one of the most groundbreaking developments in the last few years. ResNet has made it possible to train up to hundreds or even thousands of layers and still achieves compelling performance.

Convolutional neural networks have made a significant contribution toward solving the folding protein challenge, with the development of AlphaFold1, a neural network (Fig. 14, bottom panel) for prediction of 3-dimensional protein structures. AlphaFold1 created high-accuracy structures for 24 out of 43 free modeling domains, in the blind assessment of the state of the field competition Critical Assessment of Protein Structure Prediction (CASP13), whereas the next best achieved accuracy for only 14 out of 43 domains.

The AlphaFold1 neural network takes a 2-dimensional input based on the protein primary sequence of amino acid residues and features of the amino acids, and implements a CNN that generates the 2-dimensional contact map of distances $d_{cnn,jk}$ between amino acid residues $j$ and $k$. The predicted $d_{cnn}^{j,k}$ values are then used to calculate the protein backbone torsional angles $\phi$ and $\psi$ of all residues by gradient descent optimization of a function $x = G(\phi, \psi)$ that computes the coordinates $x_j$ of all alpha carbons $j$ and thus the inter-residue distances $d_{jk} = \|x_j - x_k\|$. SGD optimization minimizes the loss between $d_{jk}$ and $d_{cnn,jk}$ w.r.t. $\phi$ and $\psi$. More recently, the CNN of AlphaFold1 has been replaced by an attention transformer, exploiting advances in natural language processing developments, discussed in Sec. 7.1, leading to the development of AlphaFold2 that has achieved performance higher than 87 % at the CASP14 assessment.
5 Graph Convolutional Networks (GCN)

Graph convolutional networks (GCN) are ideally suited for describing molecules, since molecules can be represented by graphs (i.e., nodes connected by edges). The nodes are defined by \( N \times F^{(j)} \) matrices \( H^{(j)} \), corresponding to \( N \) atoms with \( F^{(j)} \) features (Fig. 15). The edges (bonds) are defined by the \( N \times N \) adjacency matrix \( A \) (with \( A_{jk} = 1 \), if atoms \( j \) and \( k \) are linked, and \( A_{jk} = 0 \), otherwise). The adjacency matrix defines the degree matrix as follows:

\[
D_{ij} = \delta_{ij} \sum_{k} A_{ik},
\]

where \( D \) is a diagonal matrix defining the number of edges of each node.

The features \( H^{(j)} \) of nodes in hidden layer \( j \) are computed with so-called propagation rules, described later in this section, pooling transformations (represented by red arrows in Fig. 15 top) that compute the features of each node by convolution with those of its neighbors as defined in the previous layer \( (j-1) \). The convolutional kernels are defined by \( F^{(j-1)} \times F^{(j)} \) matrices of weights \( \omega^{(j-1)} \), corresponding to the numbers \( F^{(j-1)} \) and \( F^{(j)} \) of features in layers \( j-1 \) and \( j \), respectively (vide infra). The weights are trained by gradient descent to ensure that the final output of the GCN matches the labels of a training set of molecules (all of them defined by their corresponding different graphs although with common atomic features).

**Inputs:** The features are initialized as defined according to the atom types and their corresponding neighborhoods in the molecular structure as encoded, for example, by the extended-connectivity fingerprints [ECFPs] also known as circular fingerprints available at DeepChem. The main properties of ECFPs are that (i) they are defined by considering circular atom neighborhoods (Fig. 15 middle panel); (ii) they are rapidly calculated; (iii) they represent substructures; (iv) they can account for a huge number of different molecular features (including stereochemical information); and (v) they represent both the presence and absence of functionality. Also, a differentiable generalization of circular fingerprints has been implemented.

![Figure 15: Top: GCN with two hidden layers for predicting a molecular property (e.g., solubility). Middle: Representation of the molecule with a graph where nodes correspond to atoms, colored according to their circular fingerprints, and edges link the nodes according to the molecular connectivity. Bottom: Adjacency matrix defining the edges of the graph as determined by the molecular connectivity.](image)
5.1 Propagation Rules

Each hidden layer of features \( H^{(j)} \), with \( H^{(0)} = X \), is obtained from the previous layer, according to a propagation rule \( H^{(j)} = f(A, H^{(j-1)} \omega^{(j-1)}) \). In general, \( f(A, H^{(j-1)} \omega^{(j-1)}) = \phi(C(A, H^{(j-1)} \omega^{(j-1)})) \), where \( C \) is a convolution that aggregates the features of each node with those of its linked neighbors, using weights that are optimized by gradient descent. The activation \( \phi \) is typically a ReLU function. Often, the adjacency matrix is incremented with the identity \((A = A + I)\) to aggregate the features of the central node with those of its neighbors during the computation of hidden layers.

Examples of popular propagation rules are:

(i) the \textit{sum rule} \((AH\omega)\) computes the features of the \(i\)-th node in the \(j\)-th hidden layer as the entries of the \(i\)-th row of the convolution of \(A\) and \(H^{(j-1)}\), as follows:

\[
C(A, H^{(j-1)} \omega^{(j-1)}))_i = \sum_{k=1}^{N} A_{ik} H_k^{(j-1)} \omega^{(j-1)}. \tag{20}
\]

(ii) the \textit{mean rule} \((D^{-1}AH\omega)\) averages the values of the neighbors, as follows:

\[
C(A, H^{(j-1)} \omega^{(j-1)}))_i = \frac{1}{\sum_{k=1}^{N} D_{ii}^{-1}} \sum_{k=1}^{N} A_{ik} H_k^{(j-1)} \omega^{(j-1)}, \tag{21}
\]

\[
= \sum_{k=1}^{N} D_{ii}^{-1} A_{ik} H_k^{(j-1)} \omega^{(j-1)},
\]

where \(D_{ii} = \delta_{i,i} \sum_k A_{ik}\) are the elements of the degree matrix. The resulting normalization allows for balanced training of all weights regardless of the node degree (i.e., atom covalency), keeping the aggregated feature on the same order or magnitude as the input features to avoid the problem of exploding gradients. (iii) the \textit{spectral rule} \((D^{-1/2}AD^{-1/2}H\omega)\) normalizes the features of a node, not only taking into consideration the degree of the node, but also the degree of its neighbors, as follows:

\[
C(A, H^{(j-1)} \omega^{(j-1)}))_i = \sum_{l=1}^{N} D_{ii}^{-1/2} \sum_{k=1}^{N} A_{ik} D_{kl}^{-1/2} H_k^{(j-1)} \omega^{(j-1)}, \tag{22}
\]

\[
= \sum_{k=1}^{N} D_{ii}^{-1/2} A_{ik} D_{kk}^{-1/2} H_k^{(j-1)} \omega^{(j-1)}.
\]

Therefore, the spectral rule also keeps the features roughly on the same scale as the input features. The main difference when compared to the mean rule is that it weights more strongly neighbors with low-degree and lower if they have a high-degree. So, it is particularly useful when low-degree neighbors provide more useful information than neighbors with high-degree.

5.2 Prediction of NMR Chemical Shifts by Graph Convolutional Networks

[This tutorial has been designed and developed by Mr. Cantarella (Haote) Li, based on the recent publication by Eric Jonas and Stefan Kuhn, J Cheminform (2019) 11:50 Rapid prediction of NMR spectral properties with quantified uncertainty] A PyTorch tutorial showing how to implement a Graph Convolutional neural network for predictions of NMR chemical shifts of molecules using molecular graphs with atomic features as inputs, can be downloaded as a notebook: GCN_NMR_Cantarella.zip.
5.3 Prediction of Solubilities by Graph Convolutional Networks with DeepChem

A turn-key tutorial on how to make predictions of molecular solubilities using Graph Convolutional neural networks with DeepChem, using molecular graphs as inputs, can be downloaded as a notebook:

- [01_The_Basic_Tools_of_the_Deep_Life_Sciences.ipynb](#)
- [01_The_Basic_Tools_of_the_Deep_Life_Sciences.pdf](#)

5.4 Introduction to classification by Graph Convolutional Networks with DeepChem

A turn-key tutorial on how to classify molecular structures using Graph Convolutional neural networks with DeepChem, using molecular graphs as inputs, can be downloaded as a notebook:

- [06_Introduction_to_Graph_Convolutions.ipynb](#)
- [06_Introduction_to_Graph_Convolutions.pdf](#)
6 Recurrent Neural Networks (RNN)

Up to this point, we have discussed feedforward networks, convolutional neural networks and graph convolutional networks for supervised learning. Those networks take all of the input values of the input layer at once and predict a single output corresponding to the input values. In contrast, recurrent neural networks (RNN) take one input value at a time, and recurrently produce outputs in context of all the previously processed input values. So, RNN’s are ideally suited for predicting the next element of a sequence (i.e., fill in the blank). For example, if the input is ‘TNT is’ the output would be ‘explosive’, or if the input is ‘H₂S is called’ the output would be ‘hydrogen sulfide’). Isn’t that cool? Obviously, a model like that trained with enough sentences from chemistry books and publications would be quite useful. So, can we build it? What is the structure of an RNN?

One way of looking at the structure of a RNN is as a linear chain of feedforward neural networks (from left to right in Fig. 16). From bottom to top, each feedforward network has an input layer (green), hidden layer (blue) and output (red) from bottom to top. The key distinct structural aspect of the RNN is that the hidden layers of the feedforward networks are connected with directional links from left to right. In each feedforward layer, the input generates a hidden state \( h_t = \phi(Ux_t + Vh_{t-1} + b) \) by using an input \( x_t \) and the hidden state \( h_{t-1} \) from the previous feedforward layer. The activation function is typically \( \phi = \tanh \) for which \( \phi' = (1 - \phi^2) \). So, for each input at a time, a feedforward generates an output \( o_t = c + Wh_t \), and passes the hidden state to the next layer. So, in contrast to feedforward and convolutional neural networks, RNNs generate a Markov chain of hidden states \( h_t \), determined by the current input \( x_t \) and the hidden state \( h_{t-1} \) which in turn is determined by all previous inputs \( x_1, x_2, \cdots, x_{t-1} \). So, each output is determined from all previous inputs. Regularization is achieved by sharing the same parameters for all times.

Training a RNN can be tricky since the gradient of the loss \( L \) associated with a large number \( N \) of iterative composition steps with \( h_t = \phi(Ux_t + Vh_{t-1} + b) \), can quickly vanish, or diverge. To see the origin of this difficulty, let us compute the gradient of \( L \) with respect to the hidden states, for the simple example where \( o_t = Wh_t \), with \( L = L(o_{t_N}, o_{t_{N-1}}, \cdots, o_{t_0}) \),

\[
\frac{\partial L}{\partial h_{t_{N-1}}} = \frac{\partial L}{\partial o_{t_N}} W + \frac{\partial L}{\partial o_{t_{N-1}}} \frac{\partial o_{t_N}}{\partial h_{t_N}} \frac{\partial h_{t_N}}{\partial h_{t_{N-1}}} = \frac{\partial L}{\partial o_{t_{N-1}}} W + \frac{\partial L}{\partial o_{t_N}} WV \phi',
\]

\[
= \frac{\partial L}{\partial o_{t_{N-1}}} W + \frac{\partial L}{\partial o_{t_N}} WV (1 - (\phi(Ux_t + Vh_{t-1} + b))^2).
\]

(23)

So, the gradient with respect to \( h_{t_{N-1}} \) depends on \( V \). Analogously, we can show that the gradient of \( L \) with respect to \( h_{t_{N-2}} \) depends on \( V^2 \), and with respect to \( h_{t_0} \) depends on \( V^N \). So, for sufficiently large \( N \), the gradients either vanish when \( V < 1 \), or diverge when \( V > 1 \), and very quickly! In fact, exponentially quickly with the number of steps \( N \). Multiple strategies have been explored to address this numerical challenge.
Echo State Network: One simple approach that also works as a regularizer is the so-called echo state network where constraints are imposed on the parameters \( W, V \) and \( U \) so that the gradients will not vanish or diverge. For example, training only \( W \), with \( V = U = 1 \) and \( \phi = \text{ReLU} \).

Clipping Gradients: Another approach is the so-called clipping gradient method where a maximum value for the gradient is imposed and the weights are evolved with a value of the gradient that is the minimum between the actual gradient and the maximum allowed value. That trick often helps to prevent the divergent gradient problem. Those are problematic since they tend to evolved the weights with big jumps into undesired regions of the parameter space where the loss is very high and the gradients are even higher thus preventing any reasonable convergence. However, the clipping gradient method is not a panacea, so we need other approaches.

Long Delays: Another approach is called long delays, where the hidden states are not connected to the nearest neighbors but rather to the second, or to the \( n \)-th neighbor. That design delays the process of vanishing or diverging gradients since the gradients would no longer depend on \( V^N \) but rather on \( V^{N/n} \).

Leaky Units: Another approach is based on the so-called leaky units where the hidden units summarizing the past are defined, as follows: \( h_{t,j} = \left(1 - \frac{1}{\tau_j}\right) h_{t-1,j} + \frac{1}{\tau_j} \phi(Ux_t + Vh_{t-1} + b) \), where the index \( j \) labels various different components of the vector \( h_t \). Note that the model has two limits, including the limit of \( \tau_j = 1 \) where the model is equivalent to the standard RNN with \( h_{t,j} = \phi(Ux_t + Vh_{t-1} + b) \), and the limit when \( \tau_j \rightarrow \infty \) where \( h_{t,j} = h_{t-1,j} \) so the model is analogous to the echo state, with \( V = 1 \). Implementing different values of \( \tau_j \) for different directions \( j \), allows to implement the influence of the past differently along different components \( j \) corresponding to different features of the model.

Gated RNN: Starting with the Long Short-Term-Memory (LSTM) algorithm, the Gated Recurrent Units (GRU) method proposed in 2014, and the many different variations that have been developed during the past few years address the central issue of either forgetting the past or allowing information from previous steps to pass through a gate and influence the current hidden state.

GRUs are more recent and more flexible than LSTM. Rather than updating the hidden units \( h_{t,j} = \phi(Ux_t + Vh_{t-1} + b) \) as in the standard RNN, GRUs are based on so-called ‘gate units’. The original version by [Cho et al] introduced the so-called update gate \( z_t = \sigma(U^{(z)}x_t + V^{(z)}h_{t-1}) \) usually defined with a sigmoid activation function, and a reset gate that resets the memory to forget the past. It is similarly defined although with different weights \( r_t = \sigma(U^{(r)}x_t + V^{(r)}h_{t-1}) \) that can be open or closed, for letting information flow, or not. Here, \( U^{(r)}, V^{(r)}, U^{(z)}, \) and \( V^{(z)} \) are weight matrices that are learned. The gates update the hidden layer, as follows:

\[
\begin{align*}
    h_t &= z_t \circ h_{t-1} + (1 - z_t) \circ \tilde{h}_t, \\
    \tilde{h}_t &= \tanh(Wx_t + U(r_t \circ h_{t-1})),
\end{align*}
\]

(24)

where the symbol \( \circ \) in the first term corresponds to the element-wise Hadamard product of elements of vector \( z_t \) times corresponding elements of vector \( h_{t-1} \) in the same position. The temporary memory \( \tilde{h}_t = \tanh(Wx_t + U(r_t \circ h_{t-1})) \) is defined just by \( x_t \) when the reset gate \( r_t \) is closed (equal to 0) and therefore zeroes the vector \( Uh_{t-1} \) that brings information from the past. Many other variations of GRUs have been proposed.

The equations for LSTM are more complicated with various types of gates which provide further flexibility at combining current information with information from the past. In fact, GRU was developed to simplify LSTM for problems where that extra flexibility was not necessary. Nevertheless, in practice, LSTM and GRU are implemented as modules that take inputs \( h_t \) and \( x_t \) and generate output \( h_{t+1} \), as implemented for example in PyTorch with \( \text{torch.nn.RNN} \) and \( \text{torch.nn.GRU} \).
Autoencoders are made by connecting two neural networks with a bottleneck layer called latent space, as shown in Fig. [17]. Autoencoders are typically trained for reconstruction of the input by minimizing the loss defined by the difference $|x - \hat{x}|$ between the input $x$ and the output $\hat{x}$, for example, in applications to data compression. The underlying dimensionality reduction when transforming $x \in \mathbb{R}^d$ to $z \in \mathbb{R}^n$ with the encoder neural network is analogous to the transformation $z = U^T x$ of principal component analysis (PCA) where $U$ is a $d \times n$ matrix. The reconstruction with the decoder neural network is analogous to the transformation $\hat{x} = U z$, with $U U^T = I$. Autoencoders can operate like PCA when they are built with linear activation. However, they can also be more general (i.e., generalizations of PCA) by including multiple layers with non-linear activation functions. Autoencoders can also be overcomplete, when the latent space is larger than the input space (i.e., $n > d$). A possible application of overcomplete autoencoders is noise reduction since the trivial 'solution' obtained by copying the input into the latent space, and from latent vector to the output would not reconstruct the clean image simply because the input is noisy.
7.1 RNN, CNN and Multi-Head Attention Autoencoders

Autoencoders have been applied to the problem of sequence to sequence reconstruction/translation, as analyzed by Sutskever et al. also described with LSTM RNNs for both the encoder and decoder networks (Fig. 18, top) by the excellent online tutorial. The corresponding transformer based on multi-head attention (Fig. 18, bottom) has been introduced by the article Attention is All You Need also analyzed by the online tutorial.

In those implementations, the input is a sentence that is tokenized (i.e., broken into the constituent words and symbols which are embedded (i.e., converted to numbers) with the nn.Embedding PyTorch module as 256 entry vectors) while the output is the corresponding sentence in a different language. The resulting transformer is trained by teacher forcing, with an additional input to the decoder corresponding to the translated sentence (Fig. 18).

We note that the context vector generated in latent space by the LSTM encoder summarizes not only the content of the sentence, as defined by the words and symbols, but also the relative order of the sequence of words which is essential for the meaning of the sentence.

Analogous transformers can also be achieved by using non-RNN encoders and decoders, including convolutional models and multi-head attention schemes that are much faster because all of the input values of the sequence are provided at the same time. Furthermore, contrary to the RNNs that require a slow sequential training process, the convolutional models and multi-head attention schemes have the advantage that can be parallelized, so they can be trained much faster than RNNs.

Figure 18: Schematic representations of the sequence to sequence autoencoders based on RNNs with teacher forcing (top) and multi-head attention (bottom).
7.1.1 Attention Mechanism

Figure 18 (lower panel) shows the implementation of the autoencoder for sequence to sequence prediction based on the so-called multi-head attention mechanism (Figure 20). Instead of using an RNN for getting dependency between input values, the encoder implements self-attention to captures information of each input token in context of the sequence, as described below.

Input to Attention: Figure 18 (lower panel) shows that each element of the input sequence (i.e., each input token) is embedded into a vector and combined with a vector that provides the position of that token $i$ in the sequence (i.e., a positional encoding vector), generating the input vector $x_i$. Key, query and value vectors ($k_i = W_k x_i$, $q_i = W_q x_i$ and $v_i = W_v x_i$, respectively) are then generated from $x_i$ using learnable weights $W_k$, $W_q$ and $W_v$.

Dot-Product Attention: The key and query vectors are used to compute softmax attention values

$$w_{ij} = \exp(score_{ij}) / \sum_k \exp(score_{kl}),$$

from the score of similarity $score_{ij} = q_i^T k_j$ is higher when $i$ and $j$ are correlated elements of the input (related elements of the input). These attention values $w_{ij}$ define the level of alignment between input tokens, and thus the level of attention that the $i$-th input should pay to the value $v_j$ of the $j$-th token, as follows $w v_i = \sum_j w_{ij} v_j$. These weighted value vectors are then passed by a linear layer with learnable weights $W_v$ to produce the attention output context vector, as follows: $o = W_v \cdot w v$.

Parallel Multi-Head: The so-called 'multi-head' implementation applies the attention mechanism in parallel to various components of the input and then concatenates and linearly transforms the output into the expected dimension for the context vector, as shown in Figure 20. Such a parallel implementation allows for attending to different parts of the sequence differently.
7.2 Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN)

Variational autoencoders (VAE) are generalized versions of autoencoders where the decoder receives an input sample from a uniform or Gaussian distribution and generates an output that resembles the type of labels used in the training. One form of VAE is based on a context vector defining the parameters of a desired probability distribution that generates the input data for the decoder. In that case, the model could be trained with a loss defined by the reconstruction error plus the KL divergence between the desired sampling distribution and the distribution defined by the context vector. The Gaussian distribution allows for analytic computation of the gradients of the KL divergence.

Variational autoencoders (VAE) are generalized versions of autoencoders where the decoder receives an input sample from a uniform or Gaussian distribution and generates an output that resembles the type of labels used in the training. One form of VAE is based on a context vector defining the parameters of a desired probability distribution that generates the input data for the decoder. In that case, the model could be trained with a loss defined by the reconstruction error plus the KL divergence between the desired sampling distribution and the distribution defined by the context vector. The Gaussian distribution allows for analytic computation of the gradients of the KL divergence.

A decoder of a generative model can also be parametrized to take samples from a uniform, or Gaussian, distribution and generate outputs with a distribution similar to the distribution that characterizes the labels of the training set. The similarity between the two distributions can be measured by the so-called *maximum mean discrepancy* (MMD) distance described below in Sec. 7.2.1.

Another form of generative models are the generative adversarial networks (GAN) that also feed a decoder with data from a uniform or Gaussian distribution and generate outputs that resemble the distribution of labels. GAN are trained by using an adversarial classifier that is trained to become better and better at distinguishing between the generated data and the labels while the generator is trained to become better at generating data that the discriminator cannot discriminate from the original data. A tutorial on how to train a GAN model with MNIST is available here, and a version with conditional generation with additional inputs to the generator and discriminator to condition the output. Also, Python-GAN provides a wonderful discussion and Pytorch implementation.

Both VAE and GAN models provide generative methods based on sampling functions capable of feeding the decoder with inputs that generate new samples resembling the type of data used in the training set. For example, when the model is trained with pictures of human faces, the decoder generates outputs that resemble human faces. When the training set is based on small drug like molecules, such as those from the Zinc library, the sampling process generates drug like molecules, as shown in the tutorial mentioned below. Isn’t that amazing? The VAE tutorial provided in the following link [designed and developed by Mr. Haote Li] shows how to implement a VAE model based on multi-head attention and MMD to generate small drug like molecules, such as those from the Zinc library.

7.2.1 Maximum Mean Discrepancy Method

The goal of this section is to introduce the so-called *maximum mean discrepancy* (MMD) method for comparisons of distribution functions, using the radial basis function (RBF) kernel.

We consider $n$ samples $x_j$ from distribution function $p(x_j)$ and $m$ samples $y_k$ from distribution function $q(y_k)$. Each of the samples is transformed according to a transformation $\phi$ and the means of the transformed data are computed, as follows: $\mu_p = \sum_{j=1}^{n} \phi(x_j)$ and $\mu_q = \sum_{k=1}^{m} \phi(y_k)$ to obtain the MMD distance between the two distributions $d_{pq} = \|p - q\|^2$ as estimated by the distance
between the two means, $d_{pq} = |\mu_p - \mu_q|^2$:

$$d_{pq} = \left( \frac{1}{n} \sum_{j=1}^{n} \phi(x_j)^T - \frac{1}{m} \sum_{k'=1}^{m} \phi(y_{k'})^T \right) \left( \frac{1}{n} \sum_{j=1}^{n} \phi(x_j) - \frac{1}{m} \sum_{k=1}^{m} \phi(y_k) \right),$$

$$= \sum_{j,j'=1}^{n} \phi(x_{j'})^T \phi(x_j) + \sum_{k,k'=1}^{m} \phi(y_{k'})^T \phi(y_k) - \sum_{j=1}^{n} \sum_{k'=1}^{m} \phi(y_{k'})^T \phi(x_j) - \sum_{k=1}^{m} \sum_{j'=1}^{n} \phi(x_{j'})^T \phi(y_k),$$

$$= \frac{1}{n^2} \sum_{j,j'=1}^{n} K(x_{j'}, x_j) + \frac{1}{m^2} \sum_{k,k'=1}^{m} K(y_{k'}, y_k) - \frac{1}{nm} \sum_{j=1}^{n} \sum_{k'=1}^{m} K(y_{k'}, x_j) - \frac{1}{nm} \sum_{j'=1}^{n} \sum_{k=1}^{m} K(x_{j'}, y_k),$$

(25)

where the radial basis function (RBF) kernel $K(x_j, y_{k'}) = exp(-|x_j - y_{k'}|^2) / (2\sigma^2)$, with the hyperparameter $\sigma$ defining the variance, ensures matching of all of the moments of the two distributions.
7.3 Time Series Prediction: Dynamical Mode Decomposition

The goal of this section is to introduce the so-called dynamical mode decomposition (DMD) method to predict the evolution of a dynamical system, after learning the principal modes of propagation from observations of a time series of data \( v_1, v_2, \cdots, v_n \), collected at \( n \) times \( t_j = (j - 1)\tau \), equally spaced by time intervals \( \tau \). Typically, the instantaneous data \( v_j \) is an array of \( N \) numbers with \( N \gg n \). When \( v_j \) are the concentrations of components of a mixture at time \( t_j \) during a chemical reaction, the task is to predict the evolution of concentrations at subsequent times after analyzing the evolution during the first \( n \) time steps.

The DMD method finds the eigenvalues \( \lambda_k \) and eigenvectors \( \Gamma_k \) of the eigenvalue problem

\[
A\Gamma_k = \lambda_k \Gamma_k
\]  
(26)

where \( A \) is the \( N \times N \) transformation matrix defined by the following equation,

\[
Y = AX,
\]  
(27)

where \( X = v_1, v_2, \cdots, v_{n-1} \), and \( Y = v_2, v_3, \cdots, v_n \) are the matrices of data collected at \( n \) equally spaced times \( t_j = j\tau \), with \( j = 1 - n \). We note that \( A \) can be diagonalized according to the similarity transformation \( \Gamma^\dagger A \Gamma = \lambda \), where \( \lambda \) is the diagonal matrix of eigenvalues, and \( \Gamma \) the matrix of eigenvectors of \( A \).

**Linear Model:** The dynamics describing the evolution of the data is simulated by considering that \( X \) is a continuous time-dependent variable,

\[
X(\tau) = AX(0),
\]  
(28)

with \( A = e^{A\tau} \) defined as the propagator so

\[
X(t) = e^{A\tau}X(0),
\]  
(29)

and \( \dot{X}(t) = AX(t) \). Note that \( \lambda = \Gamma^\dagger e^{A\tau} \Gamma = e^{\Gamma^\dagger A\tau} \), showing that \( \Gamma \) also diagonalizes \( A \) since the exponential of a matrix is diagonal only if the matrix is diagonal. Thus, \( A \) and \( \dot{A} \) have the same eigenvalues \( \lambda_k \) and eigenvectors \( \Gamma_k \).

Any arbitrary initial state of the data of interest \( X(0) = \Psi_0 \) can then be expanded in the basis of eigenvectors of \( A \), as follows:

\[
\Psi_0 = \sum_{k=1}^n c_k \Gamma_k,
\]  
(30)

where the coefficients \( c_k = \sum_j S_{kj}^{-1} \langle \Gamma_j | \Psi_0 \rangle \) are computed in terms of the inverse of the overlap matrix \( S_{kj} = \langle \Gamma_k | \Gamma_j \rangle \), since the eigenvectors \( \Gamma_j \) are not orthogonal when the matrix \( A \) is not symmetric, or Hermitian. The dynamics of the system can then be propagated, according to Eq. (29), as follows:

\[
\Psi_t = e^{A\tau} \Psi_0 = \sum_{k=1}^n c_k e^{\lambda_k \tau} \Gamma_k,
\]  
(31)

**Eigenvalues and Eigenfunctions:** The eigenvalues and eigenvectors of \( A \) are efficiently computed, according to the DMD method, by using the singular value decomposition \texttt{svd} of \( X \), as follows.
We introduce the substitution $X = U \Sigma V^T$ into Eq. (27), where $U$ is the matrix of eigenvectors of $XX^T$ with diagonal matrix of eigenvalues $\Sigma^2$ since $XX^T = (U \Sigma V^T)(V \Sigma U^T)^T = U \Sigma^2 U^T$, thus $XX^T U = U \Sigma^2$. Analogously, we can show that $V$ is the matrix of eigenvectors of $X^TX$ with the same diagonal matrix of eigenvalues $\Sigma^2$. The elements of $\Sigma^2$ thus define the entries of $\Sigma$ (the so-called singular values) which can be all positive (since $\Sigma^2$ defines $\Sigma$ out of an arbitrary phase). Singular values smaller than a given threshold value $\epsilon$ can be neglected to keep only $n \leq N$ non-zero singular values larger than a desired threshold value, $\epsilon$. Dropping out singular values smaller than $\epsilon$ is a form of regularization that filters out noise in the data and define a reduced dimensionality model that can be solved very efficiently.

Substituting the svd $X = U \Sigma V^T$ into Eq. (27), we obtain: $Y = AU \Sigma V^T$, where $U$ is an $N \times n$ matrix. Having, $U$, $\Sigma$ and $V$, we can compute the small $n \times n$ matrix $\hat{A} = U^T AU$, as follows: $\hat{A} = U^T Y V \Sigma^{-1}$. Then, we solve the small eigenvalue problem,

$$\hat{A} \Gamma_k = \lambda_k \Gamma_k,$$

and we obtain the $n$ eigenvalues $\lambda_k$, corresponding to the first $n$ eigenvalues of the large $N \times N$ matrix $A$ since $\hat{A}$ and $A$ are related by the similarity transformation $\hat{A} = U^T AU$.

Furthermore, we can obtain the first $n$ eigenvectors $\Gamma_k$ of $A$, from the eigenvectors of $\hat{A}$, as follows: $\Gamma_k = U \hat{\Gamma}_k$. This can be shown by solving for $\hat{\Gamma}_k$, as follows: $\hat{\Gamma}_k = U^T \Gamma_k$, and substituting into Eq. (32) together with $\hat{A} = U^T AU$ to obtain: $U^T A \Gamma_k = \lambda_k U^T \hat{\Gamma}_k$. Therefore, it is clear that $\Gamma_k = U \hat{\Gamma}_k$ is an eigenstate of $A$ since $A \Gamma_k = \lambda_k \Gamma_k$.

Alternatively, when $\lambda_k \neq 0$, we can compute $\hat{\Gamma}_k$, as follows:

$$\hat{\Gamma}_k = \lambda_k^{-1} Y V \Sigma^{-1} \Gamma_k.$$

Equation (33) is obtained by introducing the substitution $\Gamma_k = U \hat{\Gamma}_k$ into the l.h.s. of Eq. (33), while $Y = AX$, and $X = U \Sigma V^T$ into the r.h.s, and multiplying both sides of the equation by $U^T$, as follows: $U^T U \hat{\Gamma}_k = \lambda_k^{-1} U^T AU \Sigma V^T V \Sigma^{-1} \Gamma_k$ to obtain $\lambda_k \hat{\Gamma}_k = \hat{A} \Gamma_k$. Alternatively, the eigenstates of $A$ can be written, as follows: $\hat{\Gamma}_k = \lambda_k \Gamma_k = Y V \Sigma^{-1} \Gamma_k$ since any multiple of the eigenstate $\Gamma_k$ is also an eigenstate.

**Exercise:** Implement the DMD method for analyzing and predicting the evolution of a 2-dimensional Gaussian $\psi(x, y) = e^{-(x-x_0(t))^2-(y-y_0(t))^2}$ with an oscillatory dynamics defined by $x_0(t) = 0.5 \cos(\omega_x t)$ and $y_0(t) = 0.1 \cos(\omega_y t)$, with $\omega_x = \pi/4$ and $\omega_y = 0.5 + \pi/4$, where $t = 30$ is the total propagation time.

**Solution:** The following [python script](https://github.com/PGelss/scikit_tt) shows the implementation of the DMD method for analyzing the evolution of a 2-dimensional Gaussian. The script also includes the tt implementation which requires installation of scikit tt as described at [https://github.com/PGelss/scikit_tt](https://github.com/PGelss/scikit_tt).

```python
import numpy as np
import os
import scipy.linalg as lin
from scikit_tt.tensor_train import TT
import scikit_tt.data_driven.tdmd as tdmd
import matplotlib.pyplot as plt
import scikit_tt.utils as utl
import time as _time

def gau(r,r0):
    x=r-r0
    return(np.exp(-x**2**2))
```
def dmd_exact(x_data, y_data):
    # decompose x
    u, s, v = lin.svd(x_data, full_matrices=False, overwrite_a=True,
                      check_finite=False, lapack_driver='gesvd')
    # construct reduced matrix
    reduced_matrix = u.T @ y_data @ v.T @ np.diag(np.reciprocal(s))
    # find eigenvalues
    eigenvalues, eigenvectors = lin.eig(reduced_matrix, overwrite_a=True,
                                         check_finite=False)
    # sort eigenvalues
    ind = np.argsort(eigenvalues)[::-1]
    dmd_eigenvalues = eigenvalues[ind]
    # compute modes
    dmd_modes = y_data @ v.T @ np.diag(np.reciprocal(s)) @
                eigenvectors[:, ind] @ np.diag(np.reciprocal(dmd_eigenvalues))
    # overlap matrix
    nm = np.size(dmd_eigenvalues)
    S = np.zeros((nm,nm),dtype=complex)
    for j in range(nm):
        for k in range(nm):
            S[j][k]=np.vdot(dmd_modes[:,j],dmd_modes[:,k])
    # invert S
    Sinv = lin.inv(S)
    return S, Sinv, dmd_eigenvalues, dmd_modes

def run():
    nt=30
    npt=50
    nmo=4
    wx=np.pi/npt
    wy=np.pi/npt + .5
    dz=4/npt
    z=(np.arange(npt)-npt/2)*dz
    # generate time dependent data
    data = np.zeros((npt,npt,nt))
    for k in range(nt):
        rx=0.7*np.cos(k*wx)
        ry=0.3*np.cos(k*wy)
        for i in range(npt):
            for j in range(npt):
                data[i][j][k] = gau(z[i],rx) * gau(z[j],ry)
    # visualize time-dependent data
    nt2=nt
    stri=np.int(nt/nt2)
    f = plt.figure(figsize=plt.figaspect(1.))
    for j in range(nt2):
        i=j*stri
        ax = f.add_subplot(1, 1, 1, aspect=0.5)
        ax.imshow(np.real(data[:,:,i]), cmap='jet')
        plt.axis('off')
        plt.pause(.25)
        plt.clf()
    # construct tensors y and x corresponding to y = A x
    number_of_snapshots = data.shape[-1] - 1
\(x = \text{data[:, :, 0:} \text{number_of_snapshots].reshape(}
\text{data.shape[0] * data.shape[1], number_of_snapshots)}\)
\(y = \text{data[:, :, 1:} \text{number_of_snapshots + 1].reshape(}
\text{data.shape[0] * data.shape[1], number_of_snapshots)}\)

# apply exact DMD
\(S, Sinv, \text{eigenvalues_dmd, modes_dmd = dmd_exact(x, y)}\)
# Check \(S \ast Sinv\)
\#print("ov=", S.dot(Sinv))

# reshape result for comparison
\(\text{modes_dmd = modes_dmd.reshape([data.shape[0], data.shape[1],}
\text{number_of_snapshots])}\)

# plot 4 modes
\(f = \text{plt.figure(figsize=plt.figaspect(1.75))}\)
\text{for j in range(nmo):}
\quad ax = f.add_subplot(2, 2, j + 1, aspect=0.5)
\quad ax.imshow(np.real(modes_dmd[:, :, j]), cmap='jet')
\quad plt.axis('off')
\quad ev = eigenvalues_dmd[j]
\quad plt.title(r'\lambda ~ = ~ ' + str(\%.2f \% np.real(ev)) + ' + \'
\quad \str(\%.2f \% np.imag(ev)) + r'\i'\))

# tt DMD implementation
# construct x and y tensors and convert to TT format
\(x = \text{TT(data[:, :, 0:} \text{number_of_snapshots, None, None, None]})\)
\(y = \text{TT(data[:, :, 1:} \text{number_of_snapshots + 1, None, None, None]})\)
# define lists
\(\text{eps=0}\)
\(\text{eigenvalues_tdmd = [None]}\)
\(\text{modes_tdmd = [None]}\)
# apply exact TDMD TT
\(\text{eigenvalues_tdmd, modes_tdmd = tdmd.tdmd_exact(x, y, threshold=eps)}\)
# convert to full format for comparison and plotting
\text{modes_tdmd = modes_tdmd.full()[:, :, :, 0, 0, 0]}

# plot 4 modes
\(ff = \text{plt.figure(figsize=plt.figaspect(1.75))}\)
\text{for j in range(nmo):}
\quad ax = ff.add_subplot(2, 2, j + 1, aspect=0.5)
\quad ax.imshow(np.real(modes_tdmd[:,:,j]), cmap='jet')
\quad plt.axis('off')
\quad ev = eigenvalues_tdmd[j]
\quad plt.title(r'\lambda ~ = ~ ' + str(\%.2f \% np.real(ev)) + ' + \'
\quad \str(\%.2f \% np.imag(ev)) + r'\i'\))

\# Expansion coefficients of initial state in terms DMD modes
\(\text{rx=0.7}\)
\(\text{ry=0.3}\)
\(\text{nm=n.p.size(eigenvalues_dmd)}\)
\(\text{nm2 = 20}\)
\#print("n.p.size(eigenvalues_dmd)="\, nm)
\text{ck=n.p.zeros(nm,dtype=complex)}\)
for k in range(nm2):
    for i in range(npt):
        for j in range(npt):
            for jj in range(nm):
                ck[k] = ck[k] + Sinv[k][jj]*np.conjugate(modes_dmd[i,j,jj]) \
                        * gau(z[i],rx) * gau(z[j],ry)

# time-dependent reconstructed data
data = np.zeros((npt,npt,nt),dtype=complex)
for k in range(nt):
    norma=0.0
    for i in range(npt):
        for j in range(npt):
            for kk in range(nm2):
                aa = np.angle(eigenvalues_dmd[kk])
                ra = np.absolute(eigenvalues_dmd[kk])
                data[i][j][k] = data[i][j][k] + \
                                modes_dmd[i,j,kk] * ck[kk] * ra**k * np.exp(1j*k*aa)
                norma = norma + data[i][j][k] * dz**2
    print("norma2=",norma)
# visualize time-dependent reconstructed data
f = plt.figure(figsize=plt.figaspect(1.))
for j in range(nt):
    ax = f.add_subplot(1, 1, 1, aspect=0.5)
    ax.imshow(np.real(data[:, :,j]), cmap='jet')
    plt.axis('off')
    plt.pause(.25)
plt.clf()
plt.show()
7.4 Hybrid Quantum-Classical Neural Network

This section shows how to include a quantum layer into a neural network and create a so-called hybrid quantum-classical neural network (QCNN), as described in the [qiskit tutorial](https://qiskit.org/tutorials/quantum-classical/09_qcnn).

Figure 21 shows the implementation of a quantum hidden layer by using a quantum circuit that evolves a given input quantum state using unitary transformations parametrized by the output of a previous classical layer in the neural network. The result of measurements of the evolved quantum state provides expectation values that are used as inputs for a subsequent classical layer in the neural network. In Figure 21, $\sigma$ is a nonlinear function and $h_i$ is the value of neuron $i$ at each hidden layer. $R(h_i)$ represents any rotation of the quantum state about an angle equal to $h_i$, while $y$ is the final prediction generated from the network.

A tutorial notebook with the implementation of the hybrid QCNN with PyTorch, as applied to the classification of images of two types of digits (0 or 1) from the MNIST dataset is available at [vic_hqcnn.ipynb](https://example.com/vic_hqcnn.ipynb) and [vic_hqcnn.pdf](https://example.com/vic_hqcnn.pdf). For simplicity, the quantum circuit evolves a single qubit and involves a single trainable parameter $\theta$ corresponding to the unitary $R_y$-rotation by the angle $\theta$:

$$R_y(\theta) = \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}$$

(34)

Since the quantum circuit involves 1 parameter, it is necessary to ensure the network condenses neurons down to size 1, which is accomplished by creating a typical CNN with two fully-connected layers. The value of the last neuron of the fully-connected layer is fed as the parameter $\theta$ into the quantum circuit. The circuit measurement then serves as the final prediction for 0 or 1, as provided by a $\sigma_z$ measurement.

The tutorial also includes examples of quantum circuits for implementation of quantum algorithms that we will discuss in future lectures, that could be run on a classical quantum simulator, or on the [IBM quantum computer](https://ibmq.com) after setting up an account.

What about backpropagation? how do we calculate gradients when the quantum circuit is involved. We can view the quantum circuit as a black box and the gradient of this black box with respect to its parameters can be calculated by finite differences, incrementing the inputs $\theta$ of the quantum circuit by $\pm s$. The gradient is simply the difference between our quantum circuit evaluated at $\theta + s$ and $\theta - s$. Thus, we can systematically differentiate our quantum circuit as part of a larger backpropagation routine. This closed form rule for calculating the gradient of quantum circuit parameters is known as the [parameter shift rule](https://example.com/parameter-shift-rule).
7.5 Variational Optimization with Hybrid Neural Networks

The goal of this subsection is to explain how to use the hybrid quantum-classical neural network, shown in Fig. 21, to find the ground state of a molecule described by a Hamiltonian \( \hat{H} \).

The ground state is typically obtained according to the variational theorem by optimization of the parameters \( \theta \) that define a trial wavefunction \( |\Psi(\theta)\rangle = \hat{U}(\theta)|0\rangle \), so that the expectation value of the energy \( E_\theta = \langle \Psi(\theta)|\hat{H}|\Psi(\theta)\rangle \) is as small as possible. Therefore, we can initialize the input layer as a single neuron with input \( x_1 = 1 \), a linear hidden layer with as many neurons as necessary parameters \( \theta \) for the unitary transformations that would be applied in the quantum circuit. Each of the qubits are rotated with parameters defined by the activation of the hidden layer. The resulting quantum state is then used to compute the loss, defined as the expectation value of the energy.

7.5.1 Quantum Computation

The expectation value of the energy can be computed by quantum phase estimation, as described later in Sec. 17.2, using the following quantum circuit:

\[
|0\rangle \xrightarrow{\text{QFT}_n} |\Psi(\theta)\rangle \xrightarrow{\hat{U}} |\Psi(\theta)\rangle \quad (35)
\]

where \( U = e^{-i\hat{H}\tau/\hbar} \) and \( E_\theta = \langle \Psi(\theta)|\hat{H}|\Psi(\theta)\rangle = \sum_j p_j E_j \).

7.5.2 Hybrid Quantum-Classical Computation

The expectation value of the energy can also be computed with a classical computer by computing a single point with \( \Psi(\theta) \) using a standard software for electronic structure calculations (e.g., Gaussian, PySCF, etc.) leading to a hybrid quantum-classical method.

7.5.3 Variational Quantum Eigensolver

A particular case of the hybrid quantum-classical methodology is the so-called variational quantum eigensolver (VQE) method, where the quantum circuit is a variational form to facilitate the implementation by using as few parameters as possible.

Implementation on the IBM Experience: A turn-key tutorial on how to implement the VQE algorithm with Qiskit on Colab or the IBM Experience is available as a tutorial, and can be downloaded as a notebook: [VQE_vic.ipynb](VQE_vic.ipynb), or [VQE_vic.pdf](VQE_vic.pdf).
8 Qubits and Gates

This section introduces the representation of qubits in the Bloch sphere, as points on the surface of a unit sphere. The states of qubits can be represented in the basis of eigenstates $|0\rangle$ and $|1\rangle$ of the $\sigma_z$ operator,

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$  \hspace{1cm} \text{(36)}

with eigenvectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$  \hspace{1cm} \text{(37)}

with eigenvalues 1 and $-1$, respectively, since $\sigma_z|0\rangle = |0\rangle$ and $\sigma_z|1\rangle = -|1\rangle$. Therefore, an arbitrary qubit $|\psi\rangle$ can be represented in that basis as follows:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle,$$  \hspace{1cm} \text{(38)}

with $\alpha$ and $\beta$ two complex numbers whose squares give the probability of observing the qubit in state $|0\rangle$ and $|1\rangle$, respectively. Therefore, $\alpha^2 + \beta^2 = 1$.

When both $\alpha$ and $\beta$ are real numbers, we can represent them with a single real number $0 \leq \theta \leq \pi$, as follows: $\alpha = \cos \frac{\theta}{2}$ and $\beta = \sin \frac{\theta}{2}$. An arbitrary state can be represented with two real numbers: $0 \leq \theta \leq \pi$, and $0 \leq \phi < 2\pi$, as follows (Fig. 22):

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle = \begin{pmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{pmatrix}. \hspace{1cm} \text{(39)}$$

Therefore, all qubit states can be represented by a point on the sphere of unit radius, while the operators are represented by the axes. For example, the state $|0\rangle$ is represented by the point on the $z$ axis with $z = 1$, while the state $|1\rangle$ is also on the $z$ axis with $z = -1$.

8.1 Single qubit gates

Single qubit gates are unitary transformations represented by $2 \times 2$ matrices that rotate the qubit on the Bloch sphere, without changing its norm. An example is the NOT gate $X = \sigma_x$ that converts $|0\rangle$ into $|1\rangle$ and viceversa. For example,

$$X|0\rangle = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = |1\rangle$$  \hspace{1cm} \text{(40)}

Another example is the Hadamard gate $H = (\sigma_x + \sigma_z)/\sqrt{2}$,

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \hspace{1cm} \text{(41)}$$

that transforms $|0\rangle$ into $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and $|1\rangle$ into $|\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. \hspace{1cm}
8.2 Rotations

The Pauli matrices $\sigma_x$, $\sigma_y$, and $\sigma_z$ define the rotation operators,

$$R_r(\theta) = e^{-i \frac{\theta}{2} \sigma_r} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} \sigma_r$$

that rotate any qubit state by an angle $\theta$ about the axis $r = \{x, y, z\}$. Equation (42) can be obtained by Taylor expansion, since the square of a Pauli matrix is equal to the identity (i.e., $\sigma_r^2 = I$). Therefore,

$$e^{-i \frac{\theta}{2} \sigma_r} = \cos\left(\frac{\theta}{2} \sigma_r\right) - i \sin\left(\frac{\theta}{2} \sigma_r\right),$$

with

$$\cos\left(\frac{\theta}{2} \sigma_r\right) = 1 - \frac{1}{2} \left(\frac{\theta}{2} \sigma_r\right)^2 + \frac{1}{4} \left(\frac{\theta}{2} \sigma_r\right)^4 + \cdots$$

and

$$\sin\left(\frac{\theta}{2} \sigma_r\right) = \frac{\theta}{2} \sigma_r - \frac{1}{2} \left(\frac{\theta}{2} \sigma_r\right)^3 + \frac{1}{4} \left(\frac{\theta}{2} \sigma_r\right)^5 + \cdots$$

Analogously, $\sin\left(\frac{\theta}{2} \sigma_r\right) = \frac{\theta}{2} \sigma_r - \frac{1}{2} \left(\frac{\theta}{2} \sigma_r\right)^3 + \frac{1}{4} \left(\frac{\theta}{2} \sigma_r\right)^5 + \cdots$.

As an example, we consider the rotation of $|0\rangle$ by an angle $\theta$ about the $y$ axis, implemented as follows:

$$e^{-i \frac{\theta}{2} \sigma_y}|0\rangle = \cos \frac{\theta}{2} |0\rangle - i \sin \frac{\theta}{2} \sigma_y |0\rangle$$

or

$$e^{-i \frac{\theta}{2} \sigma_y} = \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}$$

In particular, when $\theta = \pi/2$,

$$e^{-i \frac{\pi}{4} \sigma_y}|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) = |+\rangle$$

Similarly,

$$e^{-i \frac{\pi}{4} \sigma_y}|1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} (-|0\rangle + |1\rangle) = |--\rangle$$

We note that rotation operators can be used for implementing a variety of gates. For example, $H = R_y(-\pi/2)X$, since

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

8.3 Multiple qubits

The state of a register with multiple qubits is defined by the tensor product of the states of the qubits. A simple example is the state of a register with only two qubits in states $|0\rangle$ and $|1\rangle$, respectively. In that case, the state of the register is the $|0\rangle \otimes |1\rangle$, which can also be written as $|0\rangle|1\rangle$, or simply as $|01\rangle$. In vector representation, it is written as follows:

$$|01\rangle = |0\rangle \otimes |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$
8.4 The CNOT gate

The controlled NOT gate, or CNOT gate, is defined by the following matrix:

\[
\text{CNOT} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{pmatrix}
\] (49)

It is a 2-qubit gate that flips the second qubit (i.e., the target qubit) only when the first one (i.e., the control gate) is \(|1\rangle\), as follows:

\[
|a\rangle \otimes |x\rangle \rightarrow |a\oplus x\rangle
\] (50)

It is essential in the construction of quantum circuits since it can entangle qubits, as follows:

\[
\sum_x \alpha_x |x\rangle \otimes |a\rangle \rightarrow \sum_x \alpha_x |x\rangle |a\oplus x\rangle
\] (51)

Note that for the resulting superposition state a measurement of the first qubit determines the outcome of a measurement of the second qubit, and vice versa. An interesting case is when the ancilla qubit \(|a\rangle\) is initialized as \(|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}\). Remember that the NOT gate transforms \(|0\rangle\) into \(|1\rangle\) and \(|1\rangle\) into \(|0\rangle\), so the NOT gate transforms \(|-\rangle\) into \(-|-\rangle\). As a result, the CNOT gate flips the sign of the second qubit only when the first one is \(|1\rangle\), so the outcome is, as follows:

\[
\sum_x \alpha_x |x\rangle \otimes |-\rangle \rightarrow \sum_x \alpha_x |x\rangle (-1)^x
\] (52)

Note that the second qubit remains unchanged since the phase is kicked back to the first qubit, the so-called phase ‘kickback’ trick implemented in many quantum algorithms.

8.5 Hadamard Gate

The Hadamard gate for multiple qubits transforms each of the qubits according to the Hadamard gate. A simple example is the Hadamard transformation for a register with only \(n = 2\) qubits for which the Hadamard gate is represented, as follows: \(H^{\otimes 2} = H \otimes H\). When applied to a register initialized in the \(|00\rangle\) state, we obtain:

\[
H^{\otimes 2} |00\rangle = H|0\rangle \otimes H|0\rangle = |+\rangle \otimes |+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x\rangle
\] (53)

For \(n\) qubits, the Hadamard gate transforms the \(|00\cdots0\rangle\) state, or simply as \(|0\rangle\), as follows:

\[
H^{\otimes n} |0\rangle = \frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x\rangle
\] (54)

Also, it is important to note that the Hadamard gate is its own inverse, so \(H^{\otimes n} H^{\otimes n} = 1\) and therefore the Hadamard transform of a uniform superposition generates the state \(|0\rangle\), as follows:

\[
H^{\otimes n} |0\rangle = H^{\otimes n} \frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x\rangle = |0\rangle
\] (55)

\footnote{Implementation on the IBM Experience: A turn-key tutorial on how to define quantum circuits with Qiskit on Colab or the IBM Experience is available as a Qiskit tutorial and can be downloaded as a notebook CNOT_vic.ipynb and CNOT_vic.pdf}
9 Superconducting Circuits: IBM Quantum Computer

The LC-circuit (or resonant circuit), shown in Fig. 23 is an electrical circuit that consists of an inductor, L, connected in series with a capacitor, C (i.e., and idealized model of the RLC-circuit where the resistance is assumed to be zero, thus no energy dissipation). As the capacitor starts discharging, a current is generated and as it passes through the inductor it generates a magnetic flux, described by the Biot-Savart law. As the current increases, the magnetic flux increases, and that change in the magnetic flux induces a voltage across the inductor, described by Faraday’s law: \( V_{\text{ind}} = -\frac{d\Phi}{dt} \), that reverts the current and recharges the capacitor. So, the charge sloshes back and forth like a harmonic oscillator with all of the electrons behaving as a single coordinate charge carrier.

Starting with the open circuit, the capacitor can be charged with an applied voltage \( V \). The initial charge of the capacitor is \( Q = CV \), where \( C = \epsilon a/d \) is its capacitance defined by the dielectric constant \( \epsilon \) of the material in between the plates, the area of the plates \( a \) and the distance \( d \) between the plates.

Upon closing the circuit, the voltage of the capacitor \( V = Q/C \) equals the voltage across the inductor determined by \( V_{\text{ind}} = -\frac{d\Phi}{dt} \), where the time-dependent flux

\[
\Phi(t) = \int_{-\infty}^{t} dt' V(t') = L i(t)
\]

is defined by the time-dependent current \( i(t) \), and the inductance \( L = \mu N^2 A / l \) of the solenoid with core magnetic permeability \( \mu \), length \( l \), number of turns \( N \), and area \( A \).

We obtain the equation of motion of the time-dependent charge \( Q(t) = \int_{-\infty}^{t} dt' i(t') \) by using Kirchoff’s law of voltages (i.e., the sum of voltages around a closed loop is zero, \( \sum V_i = 0 \)):

\[
\begin{align*}
V &= V_{\text{ind}} \\
\frac{Q}{C} + L \frac{d}{dt} i(t) &= 0 \\
\frac{Q}{C} + L \frac{d^2}{dt^2} Q(t) &= 0
\end{align*}
\]
and solving for \( Q(t) \), we obtain:

\[
Q(t) = Q(0) \cos(\omega_0 t),
\]

Equation (58) shows that the \( Q(t) \) behaves like a harmonic oscillator with resonant frequency \( \omega_0 = 1/\sqrt{LC} \). Equation (57) is essentially Newton’s second law: \( F = ma \), with \( F = -Q/C \), \( m = L \) and \( a = \ddot{Q} \). Therefore, the effective momentum is \( p = mv = L \dot{Q} \), so the kinetic energy is \( T = \frac{p^2}{2m} = \frac{1}{2} \dot{Q}^2 \) and since \( F = -dU/dQ \), the potential energy is \( U = Q^2/(2C) \), giving the Hamiltonian,

\[
H = T + V = \frac{L}{2} \dot{Q}^2 + \frac{Q^2}{2C}.
\]

which is essentially the Hamiltonian of a harmonic oscillator with coordinate \( Q \) and mass \( C \). Therefore, the energy levels of the circuit are quantized (Fig. 24, top panel), just like the energy levels of a harmonic oscillator (or energy levels of an atom, or molecule), so these superconducting circuits are often called artificial atoms.

The quantum state of the circuit can undergo transitions upon interaction with an external fields (usually in the microwave frequency range). In fact, the resonance frequency \( \omega_0 = 1/\sqrt{LC} \) defines the energy level spacing \( \Delta E = \hbar \omega \) and can be engineered during fabrication of the device by the macroscopic dimension of the device since, as mentioned above, \( L = \mu N^2 A/l \) and \( C = \varepsilon a/d \). Note that at low temperatures (10-15 mK) achievable with a dilution refrigerator the population of the first excited state is negligible. For example, when \( \omega = 2GHz \), the Boltzmann population of the first excited state at 10 mK is \( e^{-\hbar \omega/(k_BT)} = e^{-10} \).

Alternatively, we can describe the circuit in terms of the dynamics of the flux \( \Phi(t) = L i(t) \) by taking the time derivative of both sides of Eq. (57), as follows:

\[
\begin{align*}
\dot{Q}/C + L \frac{d^2}{dt^2} i(t) &= 0, \\
\dot{Q} + C \frac{d^2}{dt^2} \Phi(t) &= 0, \\
\Phi/L + C \frac{d^2}{dt^2} \Phi(t) &= 0,
\end{align*}
\]

which is essentially Kirchoff’s law of currents (i.e., the sum of input and output currents at any node of the circuit is zero, \( \sum_k i_k = 0 \)), with the first term giving the current from the inductor \( i_L = \Phi/L \) and the second term is the current to the capacitor \( i_C = \dot{Q} = C \dot{V} = CV \).

Therefore, the Hamiltonian can be written in terms of the flux with kinetic energy \( T = \frac{C}{2} \dot{\Phi}^2 \) and potential energy \( U = \frac{\Phi^2}{2L} \), since Eq. (60) is essentially \( F = ma \), with \( m = C \), \( a = \frac{d^2}{dt^2} \Phi(t) \), \( F = -dU/d\Phi = -\Phi/L \), and momentum \( p = mv = C \Phi = CV = Q \):

\[
H = \frac{1}{2C}(C\Phi)^2 + \frac{1}{2L} \Phi^2,
\]

\[
= \frac{1}{2C} \dot{Q}^2 + \frac{1}{2L} \Phi^2.
\]

Figure 24: Schematic energy diagram of the harmonic LC-circuit (top), compared to the anharmonic version where the solenoid is replaced by a Josephson junction (bottom).
Therefore, the coordinate $\Phi$ and the momentum $Q$ are conjugate variables with the usual commutator of coordinates and momenta ($[\hat{\phi}, \hat{Q}] = i\hbar$).

The Hamiltonian introduced by Eq. (61) can also be written, as follows:

$$H = 4E_c\hat{n}^2 + \frac{1}{2}E_L\hat{\phi}^2,$$

where we have introduced the reduced charge $\hat{n} = Q/(2e)$, and the phase $\hat{\phi} = 2\pi\Phi/\Phi_0$, with $\Phi_0 = \hbar/(2e)$ the flux quanta, corresponding to the operators for the Josephson junction introduced below, including the number of Cooper pairs and phase across, as well as the charging energy $E_c = (e^2/2C)$ and the inductive energy $E_L = \Phi_0^2/4\pi L$. In fact, replacing the solenoid by a Josephson junction (a sandwich of superconducting layers with an insulating layer in between, such as two pieces of aluminum with a thin layer of aluminum oxide in between) makes the so-called transmon (i.e., a single Josephson junction shunted by a large capacitance) which as we show below is essentially an anharmonic version of the LC-circuit (Figure 24, lower panel). An advantage of the anharmonic circuit is that it allows for selective transitions between energy levels (e.g., transition between the ground state and the first excited state) by interaction with an external field with the characteristic frequency of the ground to the first excited state transition, while the harmonic-LC circuit is less controllable since the same frequency induces $\Delta\nu = \pm 1$ transitions for all energy levels.

### 9.1 Transmon: Capacitively Shunted Junction

The transmon qubit introduced by Schoelkopf, Girvin and Devoret has been adopted by companies such as IBM (Fig. 25), Google and Rigetti as the so-called circuit quantum electrodynamics (cQED) technology by analogy to the QED experiments that manipulate transitions in real atoms by interaction with photons in optical cavities (Fig. 51). In fact, the transmon is a modified version of the conceptually simpler Cooper pair box, introduced by Devoret (i.e., a superconducting box (‘island’) connected to a superconductor reservoir by Josephson junctions), where the Cooper pair charge states in the box are analogous to the states of photons in an optical cavity of QED experiments, where the phase conjugate to the number of electrons is analogous to the phase of the field in the cavity.

Some advantages of cQED when compared to QED are:
1. the parameters of the artificial atoms can be engineered with great flexibility
2. many qubits can be placed in a transmission line resonator to make them all interact with each other (Fig. 25)
3. qubits and coupling parameters are highly tunable and allow for strong coupling

We note that although the Josephson junction is made up of millions of atoms and electrons, the electrons form Cooper pairs that condensed into a single collective ground state, below the critical temperature of the superconductor. The Cooper pairs can also coherently tunnel through the thin insulating barrier, leading to a phase difference, $\phi$, between the macroscopic wavefunctions on each side of the barrier. The tunneling current yields a new current-flux relationship, described by the Josephson relation,

$$i(t) = I_0 \sin \phi(t),$$  \hspace{1cm} (63)

where $\phi(t) = 2\pi \Phi(t)/\Phi_0$, with $\Phi_0 = h/2e$ the flux quanta and $I_0$ the critical current of the superconductor set by the fabrication parameters of the junction. Note that the relationship between $i(t)$ and $\Phi$ is non-linear, so the transmon is simply a non-linear inductor. That non-linearity leads to a Hamiltonian for the transmon that is essentially Harmonic oscillator with an additional Kerr-type nonlinear term (Eq. (66)). So, the transmon is essentially a Kerr Non-linear Resonator (KNR).

**Transmon Hamiltonian:** The equation of motion $F = ma$ for the transmon flux, analogously to Eq. (60), can be obtained by using Kirchoff’s law of currents (i.e., the sum of input and output currents at any node of a circuit is zero, $\sum_i I_i = 0$):

$$I_0 \sin(2\pi \Phi/\Phi_0) + C \dot{\Phi} = 0,$$  \hspace{1cm} (64)

where the first term is the current from the transmon and the second term is the current to the capacitor. Therefore, the potential energy is $U = -I_0\Phi_0\cos(2\pi \Phi/\Phi_0)$ is anharmonic (Fig. 24 lower panel), and the kinetic energy is $T = C\Phi^2/2 = Q^2/(2C)$, giving the Hamiltonian of the transmon:

$$H = \frac{Q^2}{2C} - I_0\Phi_0\cos(2\pi \Phi/\Phi_0),$$

$$= 4 E_c \hat{n}^2 - E_J \cos \hat{\phi},$$  \hspace{1cm} (65)

with $E_c = e^2/(2C)$, $\hat{n} = Q/(2e)$, and $E_J = I_0\Phi_0$ the Josephson energy replacing the inductive energy of the LC-circuit. Equation (65) shows that a small ratio $E_J/E_c$ provides high anharmonicity, while increasing $E_J/E_c$ makes the cos $\phi$ term dominate.

The comparison of eigenstates of the transmon to those of the LC circuit is provided in the notebook vic_transmon.ipynb and vic_transmon.pdf, where the Hamiltonian is written in terms of the annihilation operator $\hat{c}$. To obtain that expression of the Hamiltonian, we note that according to Eq. (65), which is analogous to Eq. 61, the coordinate and momentum of the transmon correspond to the phase $\hat{\phi}$ and reduced charge $\hat{n}$ operators, respectively. So, we can rewrite them in terms of the annihilation $\hat{c} = \sum_j \sqrt{j+1}|j\rangle\langle j+1|$ and creation $\hat{c}^\dagger$ operators, as follows: $\hat{\phi} = \phi_{zp} (\hat{c} - \hat{c}^\dagger)$ and $\hat{n} = i n_{zp} (\hat{c} + \hat{c}^\dagger)$ where $\phi_{zp} = \left( \frac{2E_J}{E_J^2} \right)^{1/4}$ and $n_{zp} = \left( \frac{E_J}{32E_c} \right)^{1/4}$.

Considering that $E_J/E_c \gg 1$, we have $\phi \ll 1$, so we can expand the potential energy in Taylor series, as follows: $\cos \hat{\phi} = 1 - \frac{1}{2} \hat{\phi}^2 + \frac{1}{4} \hat{\phi}^4 + \cdots$ to obtain:

$$H = -4 E_c n_{zp}^2 (\hat{c} + \hat{c}^\dagger)^2 - E_J \left( 1 - \frac{1}{2!} \phi_{zp}^2 (\hat{c} - \hat{c}^\dagger)^2 + \frac{1}{4!} \phi_{zp}^4 (\hat{c} - \hat{c}^\dagger)^4 + \cdots \right),$$

$$\approx \sqrt{8E_c E_J} \left( \frac{\hat{c} \hat{c}^\dagger + 1}{2} - E_J \frac{E_c}{12} (\hat{c}^\dagger - \hat{c})^4. \right. \hspace{1cm} (66)$$
9.2 SQUID: Tunable Junction

The transmon single Josephson junction is often replaced by a Superconducting QUantum interference Device (SQUID) with two Josephson junctions in a superconducting loop, as shown in Fig. 26, with an externally applied magnetic flux $\Phi_{\text{ext}}$ passing through the loop. As shown below, the Hamiltonian of the SQUID is equivalent to the Hamiltonian of a Josephson junction with energy $E_J(\Phi_{\text{ext}}) = E_J \cos(\Phi_{\text{ext}}/2)$, tunable in situ by simply changing the applied $\Phi_{\text{ext}}$.

**SQUID Hamiltonian:** In this subsection we show that the Hamiltonian of the SQUID is the same as the Hamiltonian of a Josephson junction with energy $E_J(\Phi)$ dependent on the externally applied magnetic flux $\Phi$.

The Lagrangian of the SQUID is given by the difference of kinetic and potential energies, as follows:

$$\mathcal{L} = \frac{1}{2} C_J \left( \Phi_1^2 + \Phi_2^2 \right) + E_J \left( \cos 2\pi \frac{\Phi_1}{\Phi_0} + \cos 2\pi \frac{\Phi_2}{\Phi_0} \right),$$

where the magnetic flux quantum $\Phi_0 = h/2e$ (for convenience, we choose units so that $\Phi_0 = 2\pi$). The flux quantization requires that $\Phi_2 - \Phi_1 = \Phi_{\text{ext}}$. Therefore, defining $\phi = (\Phi_1 + \Phi_2)/2$, we obtain $\Phi_2 = \phi + \Phi_{\text{ext}}/2$ and $\Phi_1 = \phi - \Phi_{\text{ext}}/2$. Therefore,

$$\mathcal{L} = 2C_J \phi^2 + E_J (\cos \phi + \cos \phi_{\text{ext}}),$$

$$= 2C_J \phi^2 + E_J \left( \cos(\phi - \Phi_{\text{ext}}/2) + \cos(\phi + \Phi_{\text{ext}}/2) \right),$$

$$= C_J' \phi^2 + E_J \cos(2\phi_{\text{ext}}/2),$$

with $C_J' = 2C_J$. The Hamiltonian is obtained as usual,

$$H = Q \dot{\phi} - \mathcal{L},$$

with momentum $Q = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = C_J' \phi$, giving

$$H = \frac{Q^2}{2C_J'} - E_J(\Phi_{\text{ext}}) \cos(\phi),$$

$$= 4E_c \dot{n}^2 - E_J(\Phi_{\text{ext}}) \cos(\phi),$$

where $E_c = e^2/2C_J'$, $\dot{n} = Q/2e^2$, and $E_J(\Phi_{\text{ext}}) = 2E_J \cos \frac{\Phi_{\text{ext}}}{2}$. 

![Figure 26: Top: Effective function whose Josephson energy $E_J'$ can be tuned by application of an external magnetic flux $\Phi_{\text{ext}}$.](image)
**10 Grover’s Algorithm**

The goal of this section is to introduce Grover’s algorithm, an algorithm that can find a specific state out of \(N\) possibilities in \(\sqrt{N}\) steps. Considering that a classical search algorithm would have to check one by one each of the possible states, it would take on average \(N/2\) steps and \(N - 1\) steps in the worst case. So, the Grover’s algorithm is quadratically faster than a classical search. It is particularly useful when there is a large number of possible states (so the target state is hard to find ‘the needle in the haystack’) although it is easy to check whether a given state is the target state of interest, or not. For that case, one can define a function \(f(x) = 1\) when \(x\) is the desired state (‘the needle’) and \(f(x) = 0\), otherwise. With \(f(x)\), we can define the so-called ‘oracle’ operator \(\hat{O} = e^{i\pi f(x)}\) that changes the sign of the component along the target state, as follows: \(\hat{O}|x\rangle = -|x\rangle\) when \(|x\rangle\) is the desired state. Otherwise, it leaves the state unchanged: \(\hat{O}|x\rangle = |x\rangle\).

**Algorithm:** Starting with a uniform superposition \(|s\rangle = N^{-1/2} \sum_{x=1}^{N} |x\rangle\) of all possible states \(|x\rangle\), the algorithm repeatedly applies the oracle followed by the so-called diffusion operator,

\[
|\hat{D}| = 2|s\rangle \langle s| - I, \tag{71}
\]

that changes the sign of the component perpendicular to \(|s\rangle\).

**Simple Example:** A simple example is a system with \(N = 4\) possible states, prepared in the uniform superposition

\[
|s\rangle = \frac{1}{\sqrt{N}}(|00\rangle + |10\rangle + |01\rangle + |11\rangle), \tag{72}
\]

with an oracle designed to change the sign of the \(|11\rangle\) component (e.g., \(\hat{O} = I - 2|11\rangle \langle 11|\)), as follows:

\[
\hat{O}|s\rangle = \frac{1}{\sqrt{N}}(|00\rangle + |10\rangle + |01\rangle - |11\rangle), \tag{73}
\]

Remarkably, the state \(\hat{D}\hat{O}|s\rangle\) is more aligned with the target state \(|11\rangle\) than the initial state \(|s\rangle\), as shown geometrically in Fig. (27). The procedure is repeated \(\sqrt{N}\) times to maximize the alignment with the desired state.

**Number of queries:** For the simple example of 4 states, introduced above, Grover’s algorithm transforms the initial superposition into the target state \(|11\rangle\) in one step, thus outperforming a classical search that would need on average 2 steps — i.e., \(\log_2(\#\text{states})\). We note that

\[
\langle 11|s\rangle = \frac{1}{\sqrt{N}}, \tag{74}
\]

so

\[
\hat{O}|s\rangle = |s\rangle - \frac{2}{\sqrt{N}}|11\rangle, \tag{75}
\]

Figure 27: Geometric representation of the first iteration of Grover’s algorithm. The initial superposition state \(|s\rangle\) is rotated by an angle \(\theta\) towards the target vector \(|11\rangle\) by first applying the oracle \(\hat{O}\) that changes the sign of the component along the direction of the target state \(|11\rangle\), and then the diffusion operator \(\hat{D}\) that changes the sign of the component perpendicular to the uniform superposition \(|s\rangle\).
and
\[
\hat{D} \left( |s\rangle - \frac{2}{\sqrt{N}} |11\rangle \right) = (2|s\rangle \langle s| - I) \left( |s\rangle - \frac{2}{\sqrt{N}} |11\rangle \right).
\] 
(76)

Therefore,
\[
\hat{D}\hat{O}|s\rangle = (2|s\rangle \langle s|) \left( |s\rangle - \frac{2}{\sqrt{N}} |11\rangle \right) - \left( |s\rangle - \frac{2}{\sqrt{N}} |11\rangle \right) = (2|s\rangle - \frac{2}{\sqrt{N}} |s\rangle) - |s\rangle + \frac{2}{\sqrt{N}} |11\rangle = \frac{N-4}{N} |s\rangle + \frac{2}{\sqrt{N}} |11\rangle
\] 
(77)

so for a system with 2 qubits, \(N = 4\), we obtain \((\hat{D}\hat{O})^n|s\rangle = |11\rangle\) when \(n = 1\).

More generally we note that, according to Fig. 27, the projection of \((\hat{D}\hat{O})^n|s\rangle\) along the direction of \(|11\rangle\) is \(\sin((2n+1)\frac{\theta}{2})\), with \(\frac{\theta}{2} = \frac{1}{\sqrt{N}}\). To maximize the projection, we need to have \((2n+1)\frac{\theta}{2} = \frac{\pi}{2}\).

So, \((2n+1)\frac{1}{\sqrt{N}} = \frac{\pi}{2}\), and when \(n\) is large, \(2n\frac{1}{\sqrt{N}} \approx \frac{\pi}{2}\), giving \(n \approx \frac{\pi \sqrt{N}}{4}\). So, remarkably, \(n \approx \sqrt{N}\) when \(N\) is large — i.e., quadratically faster than the classical search where \(n\) is of order \(N\).

Another way of showing that \(n \approx \sqrt{N}\) for large \(N\) is by considering how much the amplitude of the target state \(x^*\) is amplified by one step of the algorithm, when its amplitude \(\alpha_{x^*} = 2^{-1/2}\) and therefore the average amplitude of the other states is \(2^{-1/2}(N-1)^{-1/2} \approx (2N)^{-1/2}\) with \(\bar{\alpha} \approx (2^{-1/2} + N2^{-1/2}N^{-1/2})/N \approx (2N)^{-1/2}\). Upon applying the oracle, we obtain \(\alpha_{x^*} = -(2-1/2)^2\), as shown in Fig. 28, bottom panel and inverting about the mean \(\bar{\alpha}\) by applying \(\hat{D}\), we increase it amplitude to \(2^{-1/2} + 2\bar{\alpha}\) (i.e., we increment by \(2\bar{\alpha} = 2(2N)^{-1/2} = (N/2)^{-1/2}\)). Therefore, the number of steps necessary to reach an amplitude of \(2^{-1/2}\) by increments of the order of \((N/2)^{-1/2}\) is \(n = 2^{-1/2}/(N/2)^{-1/2} = \sqrt{N}\). So, the algorithm finds the solution with 50% probability in \(O(\sqrt{N})\) steps.

**Inversion about the mean:** As shown below, the effect of the diffusion operator on an arbitrary state \(|\psi\rangle = \sum_x \alpha_x |x\rangle\), is to change the sign of the component perpendicular to the initial superposition state which is equivalent to inverting the amplitudes relative to their mean value (i.e., amplitude inversion about the mean), as follows:
\[
\hat{D} \sum_{x=1}^{N} \alpha_x |x\rangle = \sum_{x=1}^{N} (2\bar{\alpha} - \alpha_x) |x\rangle,
\] 
(80)

with \(\bar{\alpha} = N^{-1} \sum_{j=1}^{N} \alpha_j\) the amplitude mean value. According to Eq. 80, \(\hat{D}\) inverts the amplitudes about the mean since \((\alpha_x - \bar{\alpha}) = -(2\bar{\alpha} - \alpha_x - \bar{\alpha})\) (Fig. 28). To obtain Eq. 80, we consider the uniform superposition \(|s\rangle = N^{-1/2} \sum_{x=1}^{N} |x\rangle\), so \(\hat{D} = 2|s\rangle \langle s| = 2N^{-1} \sum_{x=1}^{N} \sum_{y=1}^{N} \alpha_x |y\rangle \langle y| |x\rangle - I\), and
\[
\hat{D} |\psi\rangle = 2N^{-1} \sum_{x=1}^{N} \sum_{y=1}^{N} |x\rangle \sum_{x' = 1}^{N} \alpha_{x'} |y\rangle \langle y| |x'\rangle - \sum_{x' = 1}^{N} \alpha_{x'} |x'\rangle,
\] 
\[
= \sum_{x=1}^{N} |x\rangle 2N^{-1} \sum_{y=1}^{N} \alpha_y - \sum_{x=1}^{N} \alpha_x |x\rangle,
\] 
(81)

\[
= \sum_{x=1}^{N} (2\bar{\alpha} - \alpha_x) |x\rangle.
\]

**Quantum Circuit:** The oracle \(\hat{D} = (-1)^{f(|j\rangle)}\) produces a phase inversion on the amplitude of the target state while leaving all the other amplitudes unchanged. It can be implemented by using the so-called phase kickback algorithm, according to the following circuit:
Figure 28: Two examples showing how $\hat{D}$ inverts the amplitudes $\alpha_x$ about the mean $\bar{\alpha}$.

\[
\sum_x \alpha_x |x\rangle - \hat{U}_f \sum_x \alpha_x |x\rangle |f(x) \oplus b\rangle = \sum_x \alpha_x |x\rangle (-1)^{f(x)}|x\rangle
\]

Figure 29: Quantum circuit exploiting the phase kickback algorithm for implementing the unitary $U_f$ corresponding to the conditional phase inversion oracle $\hat{O} = (-1)^{f(x)}$ using $|b\rangle = |-\rangle$ as the second set of qubits. The state $|x\rangle|b\rangle$ is transformed by the unitary, as follows: $U_f|x\rangle|b\rangle = |x\rangle|f(x) \oplus b\rangle$.

When $|b\rangle = |-\rangle$ (right panel), we obtain: $U_f|x\rangle|-\rangle = |x\rangle (-1)^{f(x)}|-\rangle$ since $b$ is unchanged when $f(x) = 0$ (i.e., $f(x) \oplus b = b$) while $f(x) = 1$ transforms the state $|-\rangle = 2^{-1/2}(|0\rangle - |1\rangle)$ into $2^{-1/2}(|f(x) \oplus 0\rangle - |f(x) \oplus 1\rangle) = 2^{-1/2}(|1\rangle - |0\rangle) = (-1)|-\rangle$.

Uniform superposition: The uniform superposition can be prepared by starting with all qubits in the state $|0\rangle$ and applying the Hadamard operator to each qubit, as follows: $\hat{H} \otimes 2 |00\rangle = (\hat{H}_1 \otimes \hat{H}_1)|00\rangle$, since the Hadamard operator $\hat{H}_1 = \frac{1}{\sqrt{2}}(\hat{\sigma}_+ - \hat{\sigma}_-)$ transforms $|0\rangle$ into the symmetric linear combination $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The Hadamard transformation is also its own inverse, so applying it to the uniform superposition generates the state with all qubits in state $|0\rangle$. Furthermore, applying it to an arbitrary state, the Hadamard transform rotates it leaving the component of the uniform superposition pointing along the direction $|0\rangle$. The resulting rotation can be exploited to apply the diffusion operator to an arbitrary state by first applying the Hadamard transform (or inverse quantum Fourier transform) to have the component of the uniform superposition pointing along the direction $|0\rangle$, inverting the phases of all components orthogonal to $|0\rangle$ and then applying the Hadamard transform (or quantum Fourier transform) to leave the state with its original orientation, according to the following circuit:
Figure 30: Top Left: Quantum circuit for inversion of the phases of all components orthogonal to |0⟩ by implementing the unitary $U_g = (-1)^g(x)$, where $g(x) = 0$ if $|x⟩ = |00⋯0⟩$ and $g(x) = 1$, otherwise.

Top Right: Quantum circuit for inverting the amplitudes about their mean, according to the diffusion operator $\hat{D} = 2|s⟩⟨s| - I$, by first applying a Hadamard gate that rotates the state to leave the component of the uniform superposition pointing along the direction |0⟩, applying $U_g$ to invert the phases of all other components, and then applying a Hadamard that rotates the state back to the original orientation. Bottom: Quantum circuit for implementation of one step of the Grover’s algorithm.

Oracle operator: To explain how to construct $\hat{O}$ in terms of unitary operators for the example described above, we note that

$$-|1⟩ = \hat{H}_1\hat{\sigma}_x\hat{H}_1|1⟩,$$

(82)

since the Hadamard operator $\hat{H}_1 = \frac{1}{\sqrt{2}}(\hat{\sigma}_z + \hat{\sigma}_x)$ transforms $|1⟩$ into the antisymmetric linear combination $|-⟩ = \frac{1}{\sqrt{2}}(|0⟩ - |1⟩)$, the NOT operator $\hat{x}$ changes the sign of the state $|-⟩ = \frac{1}{\sqrt{2}}(|0⟩ - |1⟩)$ (by changing $|0⟩$ into $|1⟩$ and $|1⟩$ into $|0⟩$), and applying the Hadamard operator to $|-⟩$ returns the original state $|1⟩$. By using CNOT instead of NOT, as shown in the circuit below:

(83)

we change the sign of $|1⟩$ in the second qubit only when it is preceded by the control qubit $|1⟩$, as necessary when applying $\hat{O}$ to the uniform superposition, according to Eq. 73.

Diffusion operator: The diffusion operator changes the sign of all of the terms orthogonal to the uniform superposition. Therefore, it is possible to implement it by first orienting the state along a convenient direction where it is easier to change the sign of the orthogonal state and then rotate it back to its original orientation. For example, rotating the state $\hat{O}|s⟩$ so that its component along the direction of the uniform superposition points along one of the computational states (e.g., along the $|00⟩$ direction) so one can change the sign of all of the terms that are orthogonal to that direction (e.g., $|00⟩$) and then rotate the resulting state back so that the component along the superposition state points back along its original direction. As mentioned above, to rotate a state so that its component along the uniform superposition points to the direction $|00⟩$ we need to apply the Hadamard gate $H^©2$.

The circuit introduced by Eq. 83 changes the sign of the component along the $|11⟩$ direction. The analogous circuit but with a NOT gate previously applied to the first qubit (and subsequently applied as well to avoid modifying that qubit) would change the sign of the term along $|01⟩$, as follows:

(84)

and the same circuit but with an exchanged roles for the control and target qubits would change the sign of the term associated with the direction $|10⟩$, as follows:

51
so, the complete diffusion operator $\hat{D}$ can be implemented, as follows:

\[
\begin{array}{ccc}
H & \otimes & H \\
X & \rightarrow & X
\end{array}
\]

which can be simplified, as follows:

\[
\begin{array}{ccc}
H & \otimes & X \\
H & \otimes & X \\
H & \otimes & H \\
H & \otimes & X \\
H & \otimes & X \\
H & \otimes & H
\end{array}
\]

since $\hat{H}\hat{H} = I$.

**IBM Experience:** The YouTube videos [24] and [25] show how to implement the Grover algorithm for 2 qubits, as described above, on the IBM Quantum computer.

**Implementation on the IBM Experience:** A turn-key tutorial on how to implement the Grover’s algorithm with Qiskit on Colab or the IBM Experience is available as a Qiskit tutorial, and can be downloaded as a notebook [grover_vic.ipynb] and [grover_vic.pdf].

**Matlab function:** The `gsa.m` Matlab function simulates the Grover algorithm.

**Grover Optimization:** Grover’s quantum computational search procedure can provide the basis for implementing adaptive algorithms. An example of such methods is the Grover adaptive search (GAS) algorithm where the global minimum of a cost function $V$ is iteratively searched for with an adaptive oracle, as follows. Given an initial state $|j_0\rangle$ and its corresponding expectation value $V_0 = \langle j_0 | V | j_0 \rangle$, the oracle $\hat{O} = e^{i\pi f(j)}$ is defined with $f(j) = 1$ for states $|j\rangle$ with expectation value $\langle j | V | j \rangle < V_0$. Applying the Grover algorithm to a uniform superposition, we find a state $|j_1\rangle$ whose expectation value $V_1 < V_0$ after $r = \pi \sqrt{N}/4$ rotations. The oracle is then adapted with $f(j) = 1$ for states $|j\rangle$ with $\langle j | V | j \rangle < V_1$, and the process is iterated $m$ times until convergence to find the global minimum state $|m\rangle$ with $V_m < V_{m-1} < \cdots < V_0$.

As an example of Grover minimization (Fig. 31), we consider the problem of finding the configuration of a conjugated polyene chain with Cartesian atomic coordinates $x_1, \ldots, x_n$, assuming that bond-lengths and bending angles are known but the 1-4 dihedrals are yet to be determined since their $\pi$ or $-\pi$ (cis or trans) configurations must fulfill a constraining set $S$ of interatomic distances $d_{ij} = \|x_i - x_j\|$ determined by NMR. If all interatomic distances are determined, then the problem is trivial and can be solved in $n$ steps. However, the problem is NP-hard when only some of the distances are known.

![Isomers of linoleate](image-url)
Therefore, we need to find the coordinates \( x_1, \cdots, x_n \) that minimize the following cost function:

\[
g(x_1, x_2, \cdots, x_n) = \sum_{(ij) \in S} (d_{ij} - \|x_i - x_j\|)^2,
\]

and thus make \( g = 0 \). We note that only \( n - 3 \) dihedrals have to be specified to define all inter-atomic distances, since all bond-lengths and bending angles are given and the positions of the first 3 atoms are defined by the bond-lengths and bending angles. Since each dihedral can be either cis or trans, we have a total of \( 2^{n-3} \) possible configurations, with only some of them satisfying Eq. (88).

To implement the Grover optimization algorithm, we prepare a register with \( n - 3 \) qubits in a uniform superposition, where the state \( |0\rangle \) of the \( j \)-th qubit corresponds to the cis state of the \( j \)-th dihedral and \( |1\rangle \) corresponds to the trans state of that dihedral. The oracle \( \hat{O} = e^{i\pi f(j)} \) is defined so that \( f(j) = 1 \) if state \( |j\rangle \) satisfies Eq. (88) and \( f(j) = 0 \), otherwise.

### 10.1 Supplement I: Average Deviation Caused by the Oracle

The goal of this subsection is to show that

\[
D_j = \sum_{x=0}^{N-1} \|\psi_j^x - \psi_j\|^2 \leq 4j^2,
\]

where \( |\psi_j\rangle = \sum_y \alpha_{y,j} |y\rangle |\phi_y\rangle \), with \( |\psi_j^x\rangle = UO_x |\psi_{j-1}\rangle \) and \( |\psi_j\rangle = U |\psi_{j-1}\rangle \). According to Eq. (89), the averaged squared deviation \( D_j \) caused by \( j \) calls to the oracle, relative to the state evolving with an empty-oracle, increases no faster than \( O(j^2) \).

**Solution:** Defining \( U_x = UO_x \) and \( \Delta U = U - U_x \), we obtain:

\[
|\psi_j\rangle = (\Delta U + U_x) |\psi_{j-1}\rangle = \Delta U |\psi_{j-1}\rangle + U_x |\psi_{j-1}\rangle,
\]

so, \( |\psi_j\rangle - |\psi_j^x\rangle = \Delta U |\psi_{j-1}\rangle \). Now, substituting \( |\psi_{j-1}\rangle = U |\psi_{j-2}\rangle = (\Delta U + U_x) |\psi_{j-2}\rangle \) into Eq. (90), we obtain:

\[
|\psi_j\rangle = \Delta U |\psi_{j-1}\rangle + U_x |\psi_{j-1}\rangle = \Delta U |\psi_{j-1}\rangle + U_x \Delta U |\psi_{j-2}\rangle + U_x^2 |\psi_{j-2}\rangle,
\]

so, \( |\psi_j\rangle - |\psi_j^x\rangle = \Delta U |\psi_{j-1}\rangle + U_x \Delta U |\psi_{j-2}\rangle \).

Now, substituting \( |\psi_{j-2}\rangle = U |\psi_{j-3}\rangle = (\Delta U + U_x) |\psi_j\rangle \) into Eq. (91), we obtain:

\[
|\psi_j\rangle = \Delta U |\psi_{j-1}\rangle + U_x \Delta U |\psi_{j-2}\rangle + U_x^2 |\psi_{j-2}\rangle = \Delta U |\psi_{j-1}\rangle + U_x \Delta U |\psi_{j-2}\rangle + U_x^2 (\Delta U + U_x) |\psi_{j-3}\rangle,
\]

Therefore, \( |\psi_j\rangle - |\psi_j^x\rangle = \Delta U |\psi_{j-1}\rangle + U_x \Delta U |\psi_{j-2}\rangle + U_x^2 \Delta U |\psi_{j-3}\rangle \).
Analogously, we can repeat the procedure \( j \) times to obtain:

\[
|\psi_j^-\rangle - |\psi_j^+\rangle = U_0^j \Delta U |\psi_{j-1}\rangle + U_x \Delta U |\psi_{j-2}\rangle + U_x^2 \Delta U |\psi_{j-3}\rangle + \cdots + U_x^{(j-1)} \Delta U |\psi_0\rangle = \sum_{k=0}^{j-1} U_x^k \Delta U |\psi_{j-1-k}\rangle,
\]

so

\[
|||\psi_j^-\rangle - |\psi_j^+\rangle||^2 = \| \sum_{k=0}^{j-1} U_x^k \Delta U |\psi_{j-1-k}\rangle \|^2 \leq j \sum_{k=0}^{j-1} \| U_x^{(j-1-k)} \Delta U |\psi_k\rangle \|^2,
\]

as follows. We consider that \( \sum_{i,k=0}^{j-1} (a_k - a_i)^2 = \sum_{i,k=0}^{j-1} (a_k^2 + a_i^2 - 2a_k a_i) \), so

\[
\sum_{i,k=0}^{j-1} (a_k - a_i)^2 = \sum_{i,k=0}^{j-1} a_i^2 + \sum_{i,k=0}^{j-1} a_k^2 - 2 \sum_{i,k=0}^{j-1} a_k a_i,
\]

\[
= j \sum_{k=0}^{j-1} a_k^2 + j \sum_{i=0}^{j-1} a_i^2 - 2 \sum_{i,k=0}^{j-1} a_k a_i,
\]

\[
= j \sum_{k=0}^{j-1} a_k^2 - 2 \| \sum_{k=0}^{j-1} a_k \|^2,
\]

and solving for \( \| \sum_{k=0}^{j-1} a_k \|^2 \), we obtain:

\[
\| \sum_{k=0}^{j-1} a_k \|^2 = j \sum_{k=0}^{j-1} a_k^2 - \frac{1}{2} \sum_{i=0}^{j-1} \sum_{k=0}^{j-1} (a_k - a_i)^2.
\]

Considering that \( (a_k - a_i)^2 \) are positive numbers, we obtain the following bound:

\[
\| \sum_{k=0}^{j-1} a_k \|^2 \leq j \sum_{k=0}^{j-1} a_k^2.
\]

Therefore, according to Eqs. (97) and (93) with \( a_k = \| U_x^{(j-1-k)} \Delta U |\psi_k\rangle \| \), we obtain:

\[
\| \sum_{k=0}^{j-1} U_x^{(j-1-k)} \Delta U |\psi_k\rangle \|^2 \leq j \sum_{k=0}^{j-1} \| U_x^{(j-1-k)} \Delta U |\psi_k\rangle \|^2,
\]

\[
\leq j \sum_{k=0}^{j-1} \| \Delta U |\psi_k\rangle \|^2,
\]

\[
\leq j \sum_{k=0}^{j-1} \| |\psi_{k+1}\rangle - |\psi_{k+1}^+\rangle \|^2,
\]

where the second line of Eq. (98) was obtained by considering that \( U_x \) is unitary so \( U_x^* U = 1 \) and thus \( (U_x^{(j-1-k)} U_x^{(j-1-k)} = 1 \). So

\[
\| U_x^{(j-1-k)} \Delta U |\psi_k\rangle \|^2 = \langle \psi_k | (\Delta U)^* (U_x^{(j-1-k)})^* U_x^{(j-1-k)} \Delta U |\psi_k\rangle = \langle \psi_k | U_x^* U |\psi\rangle = \langle \psi | U^* U |\psi\rangle = \|\psi\|^2.
\]

\[3Unitary trick. For any unitary operator \( U \) (i.e., for which \( U^* U = 1 \)) we have \( \| U |\psi\rangle \|^2 = \| |\psi\rangle \|^2 \) since \( \| U |\psi\rangle \|^2 = \langle \psi | U^* U |\psi\rangle = \langle \psi | U^* U |\psi\rangle = \|\psi\|^2.\]
Now, according to Eq. (??), we obtain:

\[
\| \sum_{k=0}^{j-1} U_x^{j-1-k} \Delta U | \psi_k \rangle \|^2 = \| \Delta U | \psi_k \rangle \|^2 \leq 4j \sum_{k=0}^{j-1} \| \alpha_{x,k} \|^2.
\]  
(99)

Therefore, summing over all possible strings \( x \), we obtain:

\[
\sum_x \| \langle x | \psi_j \rangle - | \psi_j \rangle \|^2 \leq 4j \sum_{k=0}^{j-1} \sum_x \| \alpha_{x,k} \|^2,
\]  
(100)

since \( \sum_x \| \alpha_{x,j} \|^2 = 1 \).

10.2 Supplement II: Optimal Number of Queries

Show that the number of queries \( j \) necessary to identify one out of a sufficiently large number \( N \) of possible states with probability of at least 50 % is bound by \( j \geq c \sqrt{N} \), with \( c \) a small constant.

Solution:

We require \( \| \langle x | \psi_x^j \rangle \|^2 \geq 1/2 \) for any \( x \) to have at least 50 % of successfully identifying \( x \) out of \( N \) possibilities regardless of \( x \). Replacing \( | x \rangle \) by \( e^{i \theta} | x \rangle \) does not change the probability of success, so we can assume \( \langle x | \psi_x^j \rangle = \| \langle x | \psi_x^j \rangle \| \). To obtain the bound we compute the distance,

\[
D_j = \sum_x \| \langle x | \psi_x^j \rangle - | \psi_j \rangle \|^2,
\]

\[
= \sum_x \| \langle x | \psi_x^j \rangle - (| \psi_j \rangle - x) \|^2,
\]

\[
= \sum_x \| \langle x | \psi_x^j \rangle - x \|^2 + \| | \psi_j \rangle - x \|^2 - 2 \| | \psi_j \rangle - x \| \| \psi_x^j \rangle - x \| \cos(\theta),
\]

\[
= \sum_x u_x^2 + \sum_x v_x^2 - 2 \sum_x u_x v_x,
\]  
(101)

where \( u_x = | \psi_x^j \rangle - | x \rangle \) and \( v_x = | \psi_j \rangle - | x \rangle \), so

\[
E_j = \sum_x \| \langle x | \psi_x^j \rangle - x \|^2,
\]

\[
= \sum_x u_x^2 = \mathbf{u} \cdot \mathbf{u},
\]  
(102)

and

\[
F_j = \sum_x \| \langle x | \psi_x^j \rangle - x \|^2,
\]

\[
= \sum_x v_x^2 = \mathbf{v} \cdot \mathbf{v},
\]  
(103)

and \( \mathbf{u} \cdot \mathbf{v} = \sum_x u_x v_x \). Furthermore, \( \mathbf{u} \cdot \mathbf{v} = | \mathbf{u} | | \mathbf{v} | \cos(\theta) \leq | \mathbf{u} | | \mathbf{v} | \), since \( \cos(\theta) \leq 1 \), so

\[
\sum_x u_x v_x \leq | \mathbf{u} | | \mathbf{v} | = \sqrt{\sum_x u_x^2} \sqrt{\sum_x v_x^2},
\]  
(104)
Substituting Eqs. (102), (103) and (104) into Eq. (101), we obtain:

\[ D_j \geq E_j + F_j - 2\sqrt{E_jF_j} = (\sqrt{F_j} - \sqrt{E_j})^2. \]  

(105)

In addition, we have the following bounds for \( E_j \) and \( F_j \):

\[ E_j = \sum_x ||\psi_j^x||^2 - |x||^2, \]

\[ = \sum_x \langle \psi_j^x|\psi_j^x \rangle - \langle \psi_j^x|x\rangle - \langle x|\psi_j^x \rangle + \langle x|x \rangle, \]

\[ \leq \sum_x (1 - \frac{2}{\sqrt{2}} + 1). \]

\[ \leq N(2 - \sqrt{2}), \]  

(106)

Furthermore, for any normalized state \(|\psi_j\rangle\) and complete set of orthonormal states \(|x\rangle\), we have

\[ F_j = \sum_x ||\psi_j|x\rangle||^2, \]

\[ = 2N - 2\sum_x ||\psi_j|x\rangle||\cos(\theta_x), \]

\[ \geq 2N - 2\text{Max}(\sum_x ||\psi_j|x\rangle||), \]  

(107)

To maximize \( \sum_x ||\psi_j|x\rangle|| \) with the constraint \( \sum_x ||\psi_j|x\rangle||^2 = 1 \), we maximize the function \( f(c_0, \ldots, c_{N-1}) = \sum_{x=0}^{N-1} c_x + \gamma(1 - \sum_{x=0}^{N-1} c_x^2) \), with respect to \( c_y = ||\langle x|\psi_j\rangle|| \), as follows:

\[ \frac{\partial f}{\partial c_y} = 1 - \gamma 2c_y = 0, \]  

(108)

giving \( c_y = 1/(2\gamma) \) for all \( y \), and since \( \sum_{y=0}^{N-1} c_y^2 = 1 = N/(4\gamma^2) \), we obtain \( \gamma = \sqrt{N}/2 \) and \( c_y = 1/\sqrt{N} \). Therefore, according to Eq. (107),

\[ F_j \geq 2N - 2N/\sqrt{N} = 2N - 2\sqrt{N}. \]  

(109)

Substituting Eq. (106) and (109) into Eq. (105), we obtain:

\[ 4j^2 \geq D_j \geq (\sqrt{F_j} - \sqrt{E_j})^2, \]

\[ \geq \left( \sqrt{2N - 2\sqrt{N}} - \sqrt{N(2 - \sqrt{2})} \right)^2, \]

\[ \geq 4N - 2\sqrt{N} - N\sqrt{2} - 2\sqrt{N}\sqrt{(2N - 2\sqrt{N})(2 - \sqrt{2})}, \]

\[ \geq N \left( 4 - 2\sqrt{\frac{1}{N}} - \sqrt{2} - 2\sqrt{\frac{2}{\sqrt{N}}}(2 - \sqrt{2}) \right), \]  

(110)

So, for sufficiently large \( N \), we obtain:

\[ j \geq \sqrt{N}\sqrt{\frac{4 - \sqrt{2} - 2\sqrt{2(2 - \sqrt{2})}}{4}}, \]  

(111)
10.3 Supplement III: Average Success Probability of Grover’s Algorithm

Show that the Grover’s algorithm has an average probability \( p \) of finding the quantum state in one of the \( N \) possible states given by the following equation:

\[
2N - 2\sqrt{pN} - 2\sqrt{N(N-1)(1-p)} = D_j = \sum_{x=0}^{N-1} \|\psi^x_j - \psi_j\|^2 \leq 4j^2, \tag{112}
\]

where \( |\psi^x_j\rangle = \sqrt{p}|x\rangle + \sqrt{\frac{1-p}{N-1}} \sum_{y \neq x} |y\rangle \), with \( |\psi_j\rangle = \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} |y\rangle \). Therefore, asymptotically and for \( p = 1 \) Eq. (112) gives the lower bound \( j \geq \sqrt{\frac{N}{2}} \).

Solution: Computing the difference \( |\psi^x_j\rangle - |\psi_j\rangle \), we obtain:

\[
|\psi^x_j\rangle - |\psi_j\rangle = \left(\sqrt{p} - \frac{1}{\sqrt{N}}\right) |x\rangle + \left(\sqrt{\frac{1-p}{N-1}} - \frac{1}{\sqrt{N}}\right) \sum_{y \neq x} |y\rangle, \tag{113}
\]

so

\[
\sum_{x=0}^{N-1} \|\psi^x_j\rangle - |\psi_j\rangle\|^2 = \sum_{x=0}^{N-1} \left(\sqrt{p} - \frac{1}{\sqrt{N}}\right)^2 + \left(\sqrt{\frac{1-p}{N-1}} - \frac{1}{\sqrt{N}}\right)^2 (N-1),
\]

\[
= N \left(\sqrt{p} - \frac{1}{\sqrt{N}}\right)^2 + N \left(\sqrt{\frac{1-p}{N-1}} - \frac{1}{\sqrt{N}}\right)^2 (N-1), \tag{114}
\]

\[
= Np + \frac{N}{N} - 2\sqrt{pN} + N(1-p) + (N-1) - 2\sqrt{\frac{N(1-p)}{N-1}}(N-1),
\]

\[
= -2\sqrt{pN} + 2N - 2\sqrt{N(1-p)(N-1)},
\]

Therefore, for sufficiently large \( N \), the Grover’s algorithm finds one out of \( N \) states in \( \sqrt{N} \) queries, which according to Eq. (111) is optimal. For example, to find one state out of \( N \) with > 50%, we need:

\[
j \geq \sqrt{\frac{N}{2} - \sqrt{2N(1+\sqrt{N-1})}}. \tag{115}
\]
Iterative Power Algorithm: Classical Amplitude Amplification

The goal of this section is to introduce the iterative power algorithm (IPA) that can be thought of as classical computing analogue of a quantum computing algorithm, where an initial state is transformed so that a measurement can reveal the answer to the problem with high probability. We illustrate IPA as applied to global optimization and factorization.

A simple application involves finding the global minimum of the asymmetric double well potential $V(r) = -0.5r^2 + 1.0/(16.0 \times 1.3544) \times r^4 + 0.1 \times r$, shown in the figure.

To find the global minimum, according to the IPA: (1) initialize a pointer state, such as the Gaussian $\psi(r) = \pi^{-1/4}e^{-r^2/2}$ illustrated in the figure (left panel); (2) use the pointer state to compute the expectation value of $V(r)$, as follows: $V_m = \langle \psi | V | \psi \rangle$; (3) update the pointer state by projecting out the amplitude components where $V(r) > V_m$, as follows: $\psi(r) \rightarrow \psi(r) \times [1 - \text{sgn}(V(r) - V_m)]$, and normalize the resulting pointer state (middle panel); (4) Goto (2). Iterate until reaching convergence. The right panel of the figure shows the resulting pointer obtained after 15 iterations. It can be used to reveal the position of the global minimum $r_m = \langle \psi | r | \psi \rangle$ and the value of the potential at the minimum $V_m = \langle \psi | V | \psi \rangle$. Alternatively, the oracle could be defined to transform the state, as follows: $\psi(r) \rightarrow \psi(r) \times \exp(-V(r))/\|\psi(r) \times \exp(-V(r))\|$.

We note that IPA is a particular version of amplitude amplification, as defined by an oracle that iteratively projects and renormalizes the pointer state. It can also be applied for factorization when using the remainder as the cost function in the subspace of prime numbers.

The tarball file that can be downloaded from here includes several versions of the implementation of the IPA introduced in this section, as applied to global optimization in a double well potential (ttdw.py), in a 4-well potential (tt4w.py), in a multiple well (ttmw.py), revealing a transition state proximal to the global minimum (ttts.py), geometry optimization of a cluster of atoms linked by harmonic oscillators (ttho.py and ftho.py), geometry optimization of a $H_2$ molecule using pyscf for the Hartree-Fock calculations (gh2.py), and factorization (gfa.py). Note that some of the programs require Ivan Oseledets’ tensor train toolbox that can be installed from http://github.com/oseledets/ttpy or Alex Gorodetsky’s functional train C3 package. Furthermore, the $H_2$ example requires PySCF.

11.1 Convergence

The goal of this section is to compare the convergence rate of the IPA and Grover’s algorithms.

$$1 = |v_k|^2 = (N - 1) \left( \frac{\lambda_2}{\lambda_1} \right)^{2k} v_{k,\text{max}}^2 + v_{k,\text{max}}^2,$$

$$= v_{k,\text{max}}^2 \left( 1 + (N - 1) \left( \frac{\lambda_2}{\lambda_1} \right)^{2k} \right),$$

(116)
\(u_{k,\text{max}} = \frac{1}{\sqrt{1 + (N - 1) \left( \frac{\lambda_2}{\lambda_1} \right)^{2k}}} \geq \frac{1}{\sqrt{2}},\) \hspace{1cm} (117)

\[\sqrt{N} \geq k \geq \frac{\ln \left( \frac{1}{N-1} \right)}{2\ln \left( \frac{\lambda_2}{\lambda_1} \right)} \] \hspace{1cm} (118)

\[\left( \frac{1}{N-1} \right)^{1/2n} \geq \frac{\lambda_2}{\lambda_1} \] \hspace{1cm} (119)

The logarithmic scaling is comparable to or better than in optimal quantum search algorithms (e.g., the Grover quantum search method, where the number of queries necessary to amplify the amplitude of one out of N possible states scales as \(O(\sqrt{N})\).

12 Bernstein-Vazirani Algorithm: Exponential Speedup from Superpositions

The goal of this section is to explain the Bernstein-Vazirani algorithm. It is one of the earliest quantum algorithms that was able to demonstrate exponential speedup relative to the fastest possible solution achievable with a classical computer. The algorithm introduces fundamental concepts of superposition states that are essential for understanding many other quantum algorithms.

The Bernstein-Vazirani algorithm solves the following problem. Given a function \(f(x) = s \cdot x\) as a black box that transforms the \(n\)-bit string \(|x\rangle\) into a single bit \(f(x) = \{0, 1\}\) using the secret string \(|s\rangle = |s_1 s_2 \cdots s_n\rangle\), find the \(n\) bits of \(|s\rangle\), namely \(s_1, s_2, \cdots, s_n\) by evaluating \(f\) as fewer times as possible.

We note that the classical solution requires \(n\) calls to the function because the string is composed of \(n\) bits and each call to the function returns a single bit (note that to reveal the bit \(s_j\) of the secret string one would call the function with \(|x_1 x_2 \cdots x_n\rangle\) such that \(x_k = \delta_{jk}\) for \(j = 1, 2, \ldots n\)). Remarkably, the quantum computing solution based on the Bernstein-Vazirani algorithm requires a single evaluation of the function, using the quantum circuit shown in Fig. 32. Can you believe it?

Figure 32: Quantum circuit for the Bernstein-Vazirani algorithm.

To understand how the algorithm works, we first note that the Hadamard gate applied on \(n\) qubits, \(H^{\otimes n} = H \otimes H \otimes \cdots \otimes H\), transforms the \(n\)-bit string \(|s\rangle = |s_1 s_2 \cdots s_n\rangle\), as follows:

\[H^{\otimes n} |s\rangle = \sum_{x \in \{0, 1\}^n} \frac{(-1)^{s \cdot x}}{2^{n/2}} |x\rangle.\] \hspace{1cm} (120)
as shown below in Sec. 12.1. Also, the Hadamard transform is its own inverse (i.e., $|s\rangle = H^\otimes n H^\otimes n |s\rangle$). Therefore, the Bernstein-Vazirani algorithm first prepares the state $H^\otimes n |s\rangle$ by making a single call to $f(x)$, and then reveals the string $|s\rangle$ by applying the Hadamard transform and measuring.

To prepare the state $H^\otimes n |s\rangle$, we initialize $n$ working qubits as $|0\rangle$ and the ancilla qubit $|a\rangle$ as $|-\rangle$, and we apply the Hadamard transform to the working qubits to put them in the uniform superposition state, according to Eq. (120):

$$|\rangle H^\otimes n |00\cdots 0\rangle = |\rangle \sum_{x \in \{0,1\}^n} \frac{1}{2^{n/2}} |x\rangle,$$

(121)

Next, we evaluate the function $f(x)$ with the $n$ working qubits still in the uniform superposition, and we transform the ancilla qubit by performing a control-NOT operation with $f(x)$, as follows: $|a \oplus f(x)\rangle$. The strings $|x\rangle$ for which $f(x) = 0$ leave the ancilla qubit unchanged, while the strings for which $f(x) = 1$ transform $|\rangle = \frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle$ into $(-1)|\rangle = \frac{1}{\sqrt{2}} |1\rangle - \frac{1}{\sqrt{2}} |0\rangle$. Therefore, the resulting state is

$$U_f |\rangle H^\otimes n |00\cdots 0\rangle = \sum_{x \in \{0,1\}^n} \frac{1}{2^{n/2}} (-1)^{f(x)} |\rangle |x\rangle,$$

(122)

with the working qubits in the desired state. Applying a Hadamard gate to the working qubits thus reveals the secret string $s$.

### 12.1 Hadamard Transform of Arbitrary Strings

To show that

$$H^\otimes n |s\rangle = \sum_{x \in \{0,1\}^n} \frac{(-1)^{s \cdot x}}{2^{n/2}} |x\rangle,$$

(123)

we consider first the simple case of 2-bit string $|s\rangle = |s_1 s_2\rangle$, where $s_j = \{0,1\}$. When $s_1 = s_2 = 0$, we have $s \cdot x = s_1 \cdot x_1 + s_2 \cdot x_2 = 0$, regardless of the values of $x_1$ and $x_2$. So, Eq. (123) gives the expected result $H^\otimes 2 |00\rangle = |++\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |01\rangle + |10\rangle + |11\rangle)$. Next, when $s_1 = 1$ and $s_2 = 0$, we have $H^\otimes 2 |10\rangle = |+-\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |01\rangle - |10\rangle - |11\rangle)$. Note that if both the input and output strings have a 1 in the same position, then the output is multiplied by $(-1)$. That pattern of negative signs is properly captured by the term $(-1)^{s \cdot x}$. As an exercise, check that the pattern of negative signs is confirmed for the input string $|11\rangle$, and then work out 3-bit strings.

To show that the Hadamard gate is reversible, we compute

$$H^\otimes n H^\otimes n |s\rangle = \sum_{y \in \{0,1\}^n} \sum_{x \in \{0,1\}^n} \frac{(-1)^{s \cdot x}}{2^{n/2}} \frac{(-1)^{y \cdot x}}{2^{n/2}} |y\rangle,$$

(124)
Therefore,

\[ H^\otimes n H^\otimes n |s\rangle = \sum_{x \in \{0,1\}^n} \frac{(-1)^{s \cdot x}}{2^{n/2}} |s\rangle + \sum_{x \in \{0,1\}^n} \frac{(-1)^{s \cdot x}}{2^{n/2}} |y\rangle, \]

\[ = |s\rangle + \sum_{y \neq s} \sum_{x \in \{0,1\}^n} \frac{(-1)^{(y+s) \cdot x}}{2^n} |y\rangle, \]

\[ = |s\rangle + \sum_{y \neq s \perp (y+s)} \sum_{x \in \{0,1\}^n} \frac{(-1)^{(y+s) \cdot x}}{2^n} |y\rangle + \sum_{y \neq s \parallel (y+s)} \sum_{x \in \{0,1\}^n} \frac{(-1)^{(y+s) \cdot x}}{2^n} |y\rangle, \]

(125)

where the second and third term cancel each other since exactly half of the strings \(x\) are orthogonal to \(s + y\).

**Implementation on the IBM Experience:** A turn-key tutorial on how to implement the Vazirani algorithm with Qiskit on Colab or the IBM Experience is available as a [Qiskit tutorial](#) and can be downloaded as a notebook: [vic_vazirani.ipynb](#) or [vic_vazirani.pdf](#)

### 13 Hadamard Gate with Beam Splitters: Mach-Zehnder interferometer

This section introduces the **beam splitter**, as applied to the implementation of a Hadamard gate and phase shifter in an optical interferometer. These optical elements are important since any gate can be implemented just by using beam splitters, phase shifters, photodetectors and single photon sources, allowing for universal quantum computing, as shown the year 2000 by Knill, Laflamme and Milburn (KLM protocol). The optical interferometer provides a very accessible way of analyzing gates and the ‘magic’ of quantum superposition states exploited by quantum computing.

A beam splitter is a crystal that splits an incoming beam of light into two outgoing beams of intensities \(I_c = |E_{c, out}|^2\) and \(I_d = |E_{d, out}|^2\), as shown in (Figure 33). The incoming electric fields

![Figure 33: Beam splitter, a crystal that splits incoming beams with intensities \(I_a = |E_{a, in}|^2\) and \(I_b = |E_{b, in}|^2\) into outgoing beams of intensities \(I_c = |E_{c, out}|^2\) and \(I_d = |E_{d, out}|^2\).]
$E^{in}_{a}$ and $E^{in}_{b}$ could correspond to two different beams, or two modes of a single beam (e.g., two different states of polarization).

The outgoing beam $E^{out}_{c}$ results from the partial reflection of incident beam $E^{in}_{a}$ and partial transmission of beam $E^{in}_{b}$, with reflection and transmission coefficients given by the scattering matrix elements $S_{ac} = Re^{i\phi_{ac}}$ and $S_{bc} = Te^{i\phi_{bc}}$, respectively. Analogously, the outgoing beam $E^{out}_{d}$ results from the partial reflection of incident beam $E^{in}_{b}$ and partial transmission of beam $E^{in}_{a}$, with reflection and transmission coefficients given by the scattering matrix elements $S_{bd} = Re^{i\phi_{bd}}$ and $S_{ad} = Te^{i\phi_{ad}}$, respectively. Energy conservation, $I_{a} + I_{b} = I_{c} + I_{d}$ (i.e., conservation of photons), requires that $R^{2} + T^{2} = 1$ and $\phi_{ad} - \phi_{bd} + \phi_{bc} - \phi_{ac} = \pi$.

A simple example of a scattering matrix that splits incoming beams 50-50 into each of the possible outgoing directions with energy conservation is the matrix where $R = T = 1/\sqrt{2}$, with $\phi_{bd} = -\pi$ and $\phi_{ad} = \phi_{bc} = \phi_{ac} = 0$ (i.e., the Hadamard matrix):

$$BS_{1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$  \hspace{1cm} (126)

Another $S$ matrix that splits incoming beams in each of the possible outgoing directions with energy conservation has $R = T = 1/\sqrt{2}$, and $\phi_{ac} = -\pi$, where $\phi_{ad} = \phi_{bc} = \phi_{bd} = 0$,

$$BS_{2} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$  \hspace{1cm} (127)

To illustrate how the beam splitters transform an incoming state, we consider first an incoming beam with $E^{in}_{a} = 0$ and $E^{in}_{b} = 1$, or $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, as shown in Fig. 34. The outgoing beams after

$$BS_{1}$$ have $E^{in}_{c} = 1/\sqrt{2}$ and $E^{in}_{d} = -1/\sqrt{2}$, or $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$ with 50% intensity on each outgoing beam and a well-defined relative phase (i.e., $E^{out}_{c} = 1/\sqrt{2}$ and $E^{out}_{d} = 1/\sqrt{2}e^{i\pi}$). This example is very similar to the double-slit experiment, with the beam splitter functioning as a simplified version of a double-slit. The relative phases of the outgoing beams are important since they determine the interference phenomena that rules the behavior of the outgoing state, for example, as the beams pass through another $BS_{1}$, as shown in Fig. 35. Note that the second $BS_{1}$ transforms

Figure 34: A beam splitter splits an incoming beam with $E^{in}_{a} = 0$ and $E^{in}_{b} = 1$ into outgoing beams with $E^{out}_{c} = 1/\sqrt{2}$ and $E^{out}_{d} = -1/\sqrt{2}$. BS

62
Figure 35: A beam splitter splits an incoming beam with $E_{a_{in}}^i = 0$ and $E_{b_{in}}^i = 1$ into outgoing beams with $E_{c_{out}}^o = 0$ and $E_{d_{out}}^o = 1$.

the state $\left( \begin{array}{c} 1/\sqrt{2} \\ -1/\sqrt{2} \end{array} \right)$ into the output state $\left( \begin{array}{c} 0 \\ 1 \end{array} \right)$ which is identical to the initial state due to full constructive interference in the lower mode and no intensity in the top one (rather than getting a 50-50 distribution as one would expect in the absence of interference). This is the expected result since the Hadamard matrix is its own inverse.

Changing the second beamsplitter by BS$_2$, as shown in Fig. (36) provides another example that illustrates the essential role of the phases. Here, we obtain full constructive interference in $c$, and no intensity in $d$. These examples show that beam splitters can be used to generate a variety of states. In fact, when combined with phase shifters, we can generate arbitrary states, as shown below. But first, let us show what happens to the intensity of beam $d$ if we block one of the branches, right before passing through the second beamsplitter, as shown in Fig. (37). Note that the lower branch of the state $\left( \begin{array}{c} 1/\sqrt{2} \\ -1/\sqrt{2} \end{array} \right)$ generated by BS$_1$ is absorbed, thus, generating the state $\left( \begin{array}{c} 0 \\ 1/\sqrt{2} \end{array} \right)$ that is transformed by BS$_2$ into $\left( \begin{array}{c} -1/2 \\ 1/2 \end{array} \right)$. Remarkably, we increased the intensity of beam $d$ by actually blocking one of the branches!
Figure 37: A beam splitter splits an incoming beam with $E_{a}^{in} = 0$ and $E_{b}^{in} = 1$ into outgoing beams with $E_{c}^{out} = -1/2$ and $E_{d}^{out} = 1/2$.

**Elitzur-Vaidman test:** The remarkable effect of the block, responsible for increasing the intensity in channel $d$, is the basis of the Elitzur-Vaidman thought experiment. Consider a fragile precious molecule that is destroyed when it absorbs a photon (and absorbs photons with 100% quantum yield). In contrast, a defective state of the molecule does not absorb photons and is not destroyed. So, if the molecule is defective and is positioned like the block in Fig. 37 we get the intensities of Fig. 36 (i.e., no photons are ever detected in channel $d$). When the molecule is good and is placed as the block in Fig. 37, it is destroyed 50% of the times, since a photon from $b$ has only 50% probability of hitting the block. If it is not destroyed, it is because the photon went through the other branch after BS$_1$. In that case, the probability of detecting a photon in channel $d$ is equal to 1/4, as shown in Fig. 37. So, if we detect a photon in channel $d$ we know the molecule is good, as reported by a photon that never interacted with the molecule!

**Hong-Ou-Mandel effect** This effect is another manifestation of interference at a beam splitter demonstrated in 1987 by three physicists from the University of Rochester Chung Ki Hong, Zhe Yu Ou and Leonard Mandel. The experiment shows that the incidence of two indistinguishable photons, one from $a$ and the other from $b$ as described by the incident state $|\psi^{in}\rangle = 1/\sqrt{2}(|01\rangle + |10\rangle)$, produces either both photons in $c$ or both in $d$, with 50-50 probability, according to the outgoing entangled state $|\psi^{out}\rangle = 1/\sqrt{2}(|00\rangle - |11\rangle)$ (Fig. 38). Never one photon coming out from $c$ and the other from $d$.

| $\psi^{in}$ | $|\psi^{out}\rangle = 1/\sqrt{2}(|00\rangle - |11\rangle)$ |
| --- | --- |
| $|0\rangle$ | $|0\rangle$ |
| $|1\rangle$ | $|1\rangle$ |

Figure 38: Experiment demonstrating the Hong-Ou-Mandel effect: when two indistinguishable photons strike a beam splitter, one from $a$ and the other from $b$, they both come out of the same channel, either both from $c$, or both from $d$, with 50-50 probability. Never one from $c$ and the other from $d$. 
Mathematically, we can show this remarkable result, as follows. We build a symmetrized initial state with one photon in \( \langle i.e. \rangle, |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \) and one in \( \langle i.e. \rangle, |0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \), as follows:

\[
|\psi^{in}\rangle = \frac{1}{\sqrt{2}} (|0\rangle|1\rangle + |1\rangle|0\rangle), \quad \text{where} \quad |0\rangle|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad |1\rangle|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.
\]

Applying the Hadamard gate \( H \otimes 2 = H \otimes H \) to \( |\psi^{in}\rangle \), we obtain:

\[
|\psi^{out}\rangle = \frac{1}{2^{3/2}} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}} (|00\rangle - |11\rangle).
\]

Therefore, \(|\langle 01|\psi^{out}\rangle|^2 = |\langle 10|\psi^{out}\rangle|^2 = 0\), while \(|\langle 00|\psi^{out}\rangle|^2 = |\langle 11|\psi^{out}\rangle|^2 = \frac{1}{2}\).

**Phase-Shift:** Another interesting experiment is the effect of a piece of glass that introduces a phase-shift in one of the branches, as shown in Fig. 39. The phase shift is described by the gate \( Z(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix} \). The phase-shift can be measured by counting the number of photons detected at \( c \) and \( d \) (i.e., measuring the relative intensity of beams \( c \) and \( d \)).

Figure 39: Mach-Zehnder interferometer demonstrating the effect of a phase shift on the detected field amplitudes \( E_{c}^{out} \) and \( E_{d}^{out} \).

Note that the lower branch of the state generated by \( BS_1 \), \( \begin{pmatrix} 1/\sqrt{2} \\ e^{-i\pi/2} \end{pmatrix} \), gets a phase-shift and becomes \( \begin{pmatrix} 1/\sqrt{2} \\ e^{-i(\pi-\phi)/2} \end{pmatrix} \), so it is transformed by \( BS_2 \) into \( \begin{pmatrix} (-e^{i\phi} - 1)/2 \\ (-e^{i\phi} + 1)/2 \end{pmatrix} \). Therefore, the intensity ratio \( I_c/I_d = (1 + \cos(\phi))/(1 - \cos(\phi)) \) is a simple function of the phase shift.

This method of measuring the phase-shift introduced on a beam of light by a sample, as determined by mixing the signal with a reference beam that comes from the same source as the beam that went through the sample, is called optical homodyne detection. It has the main
advantage of using the reference beam (the so-called the local oscillator) to compensate for any fluctuations in the light source.

14 Phase Kickback

An important step introduced by Eq. (122), is the factorization of $|\bar{-}\rangle$ that multiplies all strings $|x\rangle$, leaving behind its phase $(-1)^f(x)$ as acquired during the conditional-f (XOR) operation when going from the first to the second row of that equation. As a result, the ancilla $|\bar{-}\rangle$ remained unchanged while its acquired phase was ‘kicked back’ to the strings of the superposition. Such a procedure that transfers the phase acquired by a quibit to another quibit(s) in a superposition state is called ‘phase kickback’ and is exploited by many quantum algorithms.

A simple illustrative example of phase kickback is the conditioned implementation of the unitary $U = e^{-\frac{i}{\hbar}Ht}$ that transforms the state $|\phi_k\rangle$, as follows: $U|k\rangle = e^{-i\theta_k}|k\rangle$, with $\theta_k = E_k t/\hbar$. When applied to a qubit prepared in state $|k\rangle$, conditioned to the state of another qubit prepared in a superposition state (e.g., $|+\rangle$), we transfer the phase $\theta_k$ to this other quibit, as follows:

$$|+\rangle \xrightarrow{U} \frac{1}{\sqrt{2}}|0\rangle + e^{-i\theta_k} \frac{1}{\sqrt{2}}|1\rangle$$

since $\left(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle\right) cU|k\rangle = \frac{1}{\sqrt{2}}|0\rangle|k\rangle + \frac{1}{\sqrt{2}}|1\rangle e^{-i\theta_k}|k\rangle = \left(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle e^{-i\theta_k}\right)|k\rangle$.

15 Deutsch Algorithm

The goal of this section is to explain the so-called Deutsch algorithm that exploits the phase kickback trick to determine if a function $f : \{0, 1\} \to \{0, 1\}$ is constant (i.e., $f(0) = f(1)$, and thus $f(0) \oplus f(1) = 0$), or not constant (i.e., ‘balanced’ since $f(0) \neq f(1)$, so $f(0) \oplus f(1) = 1$) by making a single query to the function $f$, as shown in the circuit below. This is a remarkable result since any classical algorithm would have to call $f$ twice to determine whether it is constant, or not. To see how the algorithm works, we note that the initial state $|\psi_0\rangle = |0\rangle|\bar{-}\rangle$ is transformed into state $|\psi_1\rangle$ by the first Hadamard gate, as follows:

$$|\psi_1\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)|\bar{-}\rangle,$$

$$= \frac{1}{\sqrt{2}}|0\rangle|\bar{-}\rangle + \frac{1}{\sqrt{2}}|1\rangle|\bar{-}\rangle.$$  (129)

When applying the function $f$ to $\psi_1$, we obtain:

$$|\psi_2\rangle = \hat{U}_f|\psi_1\rangle,$$

$$= \frac{1}{2} |0\rangle (|0 \oplus f(0)\rangle - |1 \oplus f(0)\rangle) + \frac{1}{2} |1\rangle (|0 \oplus f(1)\rangle - |1 \oplus f(1)\rangle),$$  (130)
We note that \((|0 \oplus f(0)| - |1 \oplus f(0)|) = (-1)^{f(0)} (|0) - |1|)\) regardless of whether \(f(0) = 0\) or \(f(0) = 1\) since \(1 \oplus 1 = 0\) and \(1 \oplus 0 = 1\). Therefore,

\[
|\psi_2\rangle = \frac{(-1)^{f(0)}}{2} |0\rangle (|0\rangle - |1\rangle) + \frac{(-1)^{f(1)}}{2} |1\rangle (|0\rangle - |1\rangle),
\]

(131)

Factorizing \((-1)^{-f(0)}\) which is equal to \((-1)^{f(0)}\), we obtain:

\[
|\psi_2\rangle = (-1)^{f(0)} \frac{1}{\sqrt{2}} \left( |0\rangle + (-1)^{f(0) \oplus f(1)} |1\rangle \right) |\rangle.
\]

(132)

When \(f(0) \oplus f(1) = 0\), the first qubit is in the state \(|\rangle\rangle\), so a Hadamard transformation makes it \(|\rangle\rangle\), and when \(f(0) \oplus f(1) = 1\), the first qubit is \(|\rangle\rangle\), so the Hadamard transformation makes it \(|1\rangle\rangle\). Therefore, a measurement of the first qubit can determine whether the function is constant, or not.

16 Simon’s Algorithm

The goal of this section is to explain Simon’s algorithm, one of the earliest quantum algorithms that was able to demonstrate exponential speedup relative to classical computing.

Simon’s algorithm solves the following problem. We are given a function \(f\) that transforms \(n\) bits into \(n\) bits (e.g., a black box executable program), such that \(f(x) = f(x \oplus s)\), with \(s\) a secret string. So, the resulting \(n\) string is the same for \(x\) and \(x \oplus s\). The problem is to discover the bits of \(|s\rangle\) by making the fewest possible calls to \(f\).

Solving the problem by classical computing requires multiple calls to \(f(x)\), at least until finding two strings for which the output is the same. So, considering that there are \(2^n\) equally probable strings, one can show that classical computing requires of the order of \(\sqrt{2^n}\) evaluations – i.e., an exponential number of evaluations of the function, analogously to the ‘birthday paradox’ problem. Remarkably, solving the problem by quantum computing with Simon’s algorithm requires only a polynomial in \(n\) number of evaluations for \(f(x)\), rather than an exponential number.

Simon’s algorithm requires preparation of the particular superposition \(\frac{1}{\sqrt{2}} |r\rangle + \frac{1}{\sqrt{2}} |r \oplus s\rangle\) where \(|r\rangle\) is a random \(n\)-bit string. That particular superposition state can be generated by initializing \(2n\) qubits as \(|0\rangle\), applying a Hadamard transform \(H^{\otimes n}\) to the first \(n\) qubits to generate a uniform superposition of all possible \(2^n\) random strings, as shown in Eq. (120), and then transforming the second set of \(n\) qubits initialized as \(|0\rangle\) into \(|f(r)\rangle\). Upon measuring the second set of \(n\) qubits (as shown in Fig. 4c), we collapse the first set of \(n\) qubits into the desired superposition since a given value of \(f(r)\) is fulfilled by both \(r\) and \(s \oplus r\).

Finally, we apply a Hadamard transform to the first set of \(n\) qubits (already prepared in the particular superposition \(\frac{1}{\sqrt{2}} |r\rangle + \frac{1}{\sqrt{2}} |r \oplus s\rangle\), as described above) and we measure. It is important to note that the strings \(y\) sampled by measuring the first \(n\) qubits, after the Hadamard transform, satisfy the condition that \(y \cdot s = 0\) (i.e., \(y \in s^\perp\)) since (as shown below) all other strings have
The probability of success is obtained by considering that there are \(2^{n-1}\) equally probable strings \(y\) that are sampled. Sampling a linearly dependent string in the \((n-1)\) step, after having sampled \((n-2)\) linearly independent strings in the previous steps would lead to failure. The probability of that event is equal to the probability of a string \(1/2^{n-1}\) times the number of

\[
\begin{align*}
H^{\otimes n}\frac{1}{\sqrt{2}}|r\rangle + H^{\otimes n}\frac{1}{\sqrt{2}}|r \oplus s\rangle &= \sum_{y \in \{0,1\}^n} \left( \frac{(-1)^{r\cdot y}}{2^{(n+1)/2}} + \frac{(-1)^{(r \oplus s)\cdot y}}{2^{(n+1)/2}} \right) |y\rangle, \\
&= \sum_{y \in \{0,1\}^n} \frac{(-1)^{r\cdot y}}{2^{(n+1)/2}} (1 + (-1)^{s\cdot y}) |y\rangle, \\
&= \sum_{y \in s^\perp} \frac{(-1)^{r\cdot y}}{2^{(n-1)/2}} |y\rangle,
\end{align*}
\]

where \(y \in s^\perp\) fulfill the condition \(s \cdot y = 0\). We note that the probability of each sampled string is given by the square of the expansion coefficient \(\left|\frac{1}{2^{(n-1)/2}}\right|^2 = \frac{1}{2^{n/2}}\). So, the number of strings that are orthogonal to \(s\) \((y \in s^\perp)\) is equal to \(2^n/2 = 2^{n-1}\) (i.e., half the total number of \(n\)-bit strings).

Therefore, by repeating the preparation and measurement \((n-1)\) times, we get a set of \((n-1)\) equations of the form:

\[
\begin{align*}
y_1^{(1)} s_1 + y_2^{(1)} s_2 + \cdots + y_n^{(1)} s_n &= 0 \\
y_1^{(1)} s_1 + y_2^{(1)} s_2 + \cdots + y_n^{(1)} s_n &= 0 \\
\cdots \\
y_1^{(n-1)} s_1 + y_2^{(n-1)} s_2 + \cdots + y_n^{(n-1)} s_n &= 0
\end{align*}
\]

that can be solved for \(s_1, s_2, \ldots, s_n\) to get two possible solutions, including the trivial solution with all \(s_j = 0\) and the non-trivial solution of interest.

The algorithm works only when the \((n-1)\) equations are linearly independent, which can be shown to happen at least with 25\% probability (and exact probability of success close to 29\%), as follows.

The \(\geq 25\%\) probability of success is obtained by considering that there are \(2^{n-1}\) equally probable strings \(y\) that are sampled. Sampling a linearly dependent string in the \((n-1)\) step, after having sampled \((n-2)\) linearly independent strings in the previous steps would lead to failure. The probability of that event is equal to the probability of a string \(1/2^{n-1}\) times the number of
strings that are linearly dependent on of the previously sampled strings. Writing a linear combination of all \((n - 2)\) previously selected strings, with expansion coefficients equal to \(1\) or \(0\) for each string, shows that there are \(2^{n-2}\) possible strings that would be linearly dependent on the previously sampled strings. So, the probability of failure in that last step is \(2^{n-2} \times \frac{1}{2^{n-1}} = \frac{1}{2}\). Analogously, we find that failing in the previous step would have probability \(\frac{1}{4}\), in the previous one \(\frac{1}{8}\), until the very first step for which the probability of failing is \(\frac{1}{2^{n-1}}\) because only the string with all zeros would lead to failure.

Summing the probabilities of failing at the first \((n - 2)\) steps (instead of multiplying them) gives us an upper bound to the probability that the algorithm fails in the first \((n - 2)\) steps. Therefore, multiplying that probability by the probability of independently failing in the last step gives an upper bound to the total probability of failing, as follows. Considering that the geometric series is \(\sum_{j=0}^{n-1} x^j = (1 - x^n) / (1 - x)\), we obtain:

\[
\frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^{n-1}} = \sum_{j=1}^{n-1} 0.5^j - \frac{1}{2} = \sum_{j=0}^{n-1} 0.5^j - 1 - \frac{1}{2} = \frac{(1 - 0.5^n)}{(1 - 0.5)} - 1 - \frac{1}{2} = \frac{(0.5 - 0.5^n)}{0.5} - 1/2 = 1/2 - \frac{1}{2^{n-1}} \leq 1/2.
\]

Therefore, the probability of success in the first \((n - 2)\) steps is at least \(1/2\) and since the probability of success in the last step is \(1/2\), the probability of success is at least \(1/4\).

**Exercise:** Let \(x, y \in \{0,1\}^n\) and let \(s = x \oplus y\). Show that

\[
H^{\otimes n} \frac{1}{\sqrt{2}} |x\rangle + H^{\otimes n} \frac{1}{\sqrt{2}} |y\rangle = \sum_{z \in \{s\}^\perp} (-1)^{x \cdot z} \frac{1}{2^{(n-1)/2}} |z\rangle.
\]

**Solution:**

\[
H^{\otimes n} \frac{1}{\sqrt{2}} |x\rangle + H^{\otimes n} \frac{1}{\sqrt{2}} |y\rangle = \frac{1}{2^{(n+1)/2}} \sum_{z \in \{0,1\}^n} ((-1)^{x \cdot z} + (-1)^{y \cdot z}) |z\rangle,
\]

\[
= \frac{1}{2^{(n+1)/2}} \sum_{z \in \{0,1\}^n} (-1)^{x \cdot z} (1 + (-1)^{y \cdot z}) |z\rangle,
\]

\[
= \frac{1}{2^{(n+1)/2}} \sum_{z \in \{0,1\}^n} (-1)^{x \cdot z} (1 + (-1)^{(s \cdot 2^{(n-1)/2})} |z\rangle,
\]

\[
= \frac{1}{2^{(n+1)/2}} \sum_{z \in \{0,1\}^n} (-1)^{x \cdot z} (1 + (-1)^{s \cdot z}) |z\rangle,
\]

\[
= \frac{1}{2^{(n+1)/2}} \sum_{z \in \{0,1\}^n} (-1)^{x \cdot z} (1 + (-1)^{z}) |z\rangle,
\]

\[
= \frac{1}{2^{(n+1)/2}} \sum_{z \in \{s\}^\perp} (-1)^{x \cdot z} \frac{1}{2^{(n-1)/2}} |z\rangle,
\]

\[
= \frac{1}{2^{(n-1)/2}} \sum_{z \in \{s\}^\perp} (-1)^{x \cdot z} |z\rangle.
\]
Exercise:

We have defined $s^\perp$, but more generally we can let $S$ be a vector subspace of $Z_2^n$, and define $S^\perp = \{ t \in Z_2^n \mid t \cdot s = 0 \text{ for all } s \in S \}$. So our previously defined $s^\perp$ corresponds to $S^\perp$ where $S = \{0, s\}$ is the 2-dimensional vector space spanned by $s$.

- (a) Define $|S\rangle = \sum_{s \in S} \frac{1}{\sqrt{2^n}} |s\rangle$. Prove that $H^\otimes n |S\rangle = \sum_{w \in S^\perp} \frac{1}{\sqrt{2^n}} |w\rangle$.

- (b) For any $y \in \{0,1\}^n$ define $|y + S\rangle = \sum_{s \in S} \frac{1}{\sqrt{2^n}} |y + s\rangle$. What is $H^\otimes n |y + S\rangle$?

Solution:

(a)

$$H^\otimes n |S\rangle = \frac{1}{2^{(n+m)/2}} \sum_{s \in S, w \in \{0,1\}^n} (-1)^{s \cdot w} |w\rangle,$$

$$= \frac{1}{2^{(n+m)/2}} \sum_{s \in S, w \in S^\perp} (-1)^{s \cdot w} |w\rangle + \frac{1}{2^{(n+m)/2}} \sum_{s \in S, w \notin S^\perp} (-1)^{s \cdot w} |w\rangle,$$

$$= \frac{1}{2^{(n+m)/2}} \sum_{s \in S, w \in S^\perp} |w\rangle + \frac{1}{2^{(n+m)/2}} \sum_{s \in S, w \notin S^\perp} (-1)^{s \cdot w} |w\rangle,$$

$$= \frac{2^m}{2^{(n+m)/2}} \sum_{w \in S^\perp} |w\rangle + \frac{1}{2^{(n+m)/2}} \sum_{w \notin S^\perp} (-1)^{s \cdot w} |w\rangle,$$

$$= \frac{1}{2^{(n-m)/2}} \sum_{w \in S^\perp} |w\rangle,$$

since $\sum_{s \in S} (-1)^{s \cdot w} = 0$.

(b)

$$H^\otimes n |y + S\rangle = \sum_{s \in S} \frac{1}{\sqrt{2^{(m+n)}}} \sum_{w \in \{0,1\}^n} (-1)^{y \cdot w + s \cdot w} |w\rangle,$$

$$= \frac{1}{\sqrt{2^{(m+n)}}} \sum_{w \in \{0,1\}^n} (-1)^{y \cdot w} \sum_{s \in S} (-1)^{s \cdot w} |w\rangle,$$

$$= \frac{1}{\sqrt{2^{(m+n)}}} \sum_{w \in S^\perp} (-1)^{y \cdot w} \sum_{s \in S} (-1)^{s \cdot w} |w\rangle + \frac{1}{\sqrt{2^{(m+n)}}} \sum_{w \notin S^\perp} (-1)^{y \cdot w} \sum_{s \in S} (-1)^{s \cdot w} |w\rangle,$$

$$= \frac{1}{\sqrt{2^{(m+n)}}} \sum_{w \in S^\perp} (-1)^{y \cdot w} 2^m |w\rangle,$$

$$= \frac{1}{\sqrt{2^{(n-m)}}} \sum_{w \in S^\perp} (-1)^{y \cdot w} |w\rangle,$$

(139)
17 Quantum Fourier Transform

The goal of this section is to introduce the quantum Fourier transform, the quantum circuit of the classical discrete Fourier transform, and its comparison to the Hadamard transform.

As we discussed earlier, the Hadamard gate $H^{\otimes n} = H \otimes H \otimes \cdots \otimes H$, with

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$  \hspace{1cm} (140)

transforms the $N$-bit string $|s\rangle = |s_1 s_2 \cdots s_N\rangle$, with $N = 2^n$, as follows:

$$H^{\otimes n} |s\rangle = \sum_{x \in \{0,1\}^N} \frac{(-1)^{s \cdot x}}{N^{1/2}} |x\rangle,$$  \hspace{1cm} (141)

with matrix elements $\langle j| H^{\otimes n} |k\rangle = N^{-1/2} (-1)^{j \cdot k}$. Analogously, the quantum Fourier transform of the $N$-bit string $|s\rangle = |s_1 s_2 \cdots s_N\rangle$ is defined, as follows:

$$QFT_N |s\rangle = \sum_{x \in \{0,1\}^N} \frac{w^{s \cdot x}}{N^{1/2}} |x\rangle,$$  \hspace{1cm} (142)

where $w = e^{2\pi i / N}$ is the $N$-th root of unity since $w^N = 1$. Therefore, the matrix elements of the $N$-dimensional quantum Fourier transform are defined, as follows: $\langle j|QFT_N |k\rangle = N^{-1/2} w^{jk}$, defining the QFT in matrix form:

$$QFT_N = \frac{1}{N^{1/2}} \begin{bmatrix} w^0 & w^0 & \cdots & w^0 \\ w^0 & w^1 & \cdots & w^{N-1} \\ w^0 & w^2 & \cdots & w^{2(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ w^0 & w^{N-1} & \cdots & w^{(N-1)^2} \end{bmatrix}.$$  \hspace{1cm} (143)

We note that the Fourier transform of a single qubit coincides with the Hadamard transform, shown in Eq. (140), since $w = -1$ when $N = 2$. Also, $QFT_N |0 \cdots 0\rangle = H^{\otimes n} |0 \cdots 0\rangle$, since $(-1)^{0 \cdot x} = (w)^{0 \cdot x} = 1$ in both Eq. (141) and Eq. (142). We also note that each of the matrix elements $w^{jk} = e^{2\pi i / N}$ are simple phase shifters.

**Inverse Fourier transform:** The inverse Fourier transform is defined, as follows:

$$QFT_N^{-1} |x\rangle = \sum_{r \in \{0,1\}^N} \frac{w^{-r \cdot x}}{N^{1/2}} |r\rangle.$$  \hspace{1cm} (144)

Therefore, the matrix elements of the $N$-dimensional inverse Fourier transform are $\langle j|QFT_N^{-1} |k\rangle = N^{-1/2} w^{-jk}$

Note that according to Eqs. (142) and (144), $QFT_N^{-1} QFT_N |s\rangle = |s\rangle$:

$$QFT_N^{-1} QFT_N |s\rangle = \sum_{x \in \{0,1\}^N} \sum_{r \in \{0,1\}^N} \frac{w^{s \cdot x}}{N^{1/2}} \frac{w^{-r \cdot x}}{N^{1/2}} |r\rangle,$$

$$= \sum_{r \in \{0,1\}^N} \sum_{x \in \{0,1\}^N} \frac{w^{(s-r) \cdot x}}{N} |r\rangle,$$

$$= |s\rangle + \frac{1}{N} \sum_{r \neq s} \sum_{x \in \{0,1\}^N} \frac{(N-1)}{N} w^{(s-r) \cdot x} |r\rangle,$$

with $\sum_{x=0}^{N-1} w^{(s-r) \cdot x} = 0$ when $s - r \neq 0$.  

71
17.1 Properties of the Fourier transform

Property 1: Shift Invariant. Fourier sampling is unaffected by a constant shift since

\[ QFT_n |s + \Delta \rangle = \sum_{x \in \{0,1\}^n} \frac{w^x (s + \Delta)}{n^{1/2}} |x\rangle, \]

\[ = \sum_{x \in \{0,1\}^n} w^{sx} |x\rangle, \quad (146) \]

with \[ |w^{sx} \overline{w^{sx}}|^2 = |w^{sx}|^2, \] since \[ w = e^{i2\pi/n} \] and thus \[ |w^{sx}|^2 = 1. \] Therefore, the probability of sampling string \[ |x\rangle \] is the same before and after the shift.

Property 2: Sum of roots. The sum of the \( n \)-th roots of unity is equal to zero, which can be shown by using the geometric series \( \sum_{j=0}^{n-1} x^j = (1 - x^n) / (1 - x) \), as follows:

\[ \sum_{j=0}^{n-1} w^j = \frac{(1 - w^n)}{(1 - w)} = 0, \quad (147) \]

since \( w^n = 1 \). Note that the same result is obtained by replacing \( w \) by any power of \( w \).

Property 3: Orthogonality. The columns of the Fourier transform are orthonormal vectors that define a coordinate transformation (i.e., a rotation in Hilbert space). Computing the scalar product of columns \( l \) and \( k \) (with \( \overline{w} = e^{-i2\pi/n} \) the conjugate of \( w = e^{i2\pi/n} \), we obtain:

\[ \langle l|k \rangle = \frac{1}{n} \sum_{j=0}^{n-1} \overline{w}^{lj} w^{kj}, \]

\[ = \frac{1}{n} \sum_{j=0}^{n-1} \overline{w}^{(k-l)j}, \]

\[ = \frac{1}{n} \delta_{lk} \sum_{j=0}^{n-1} 1 + \frac{1}{n} (1 - \delta_{lk}) \sum_{j=0}^{n-1} \overline{w}^{(k-l)j}, \]

\[ = \delta_{lk} + \frac{1}{n} (1 - \delta_{lk}) \frac{(1 - w^{n(k-l)})}{(1 - \overline{w}^{(k-l)})}, \]

\[ = \delta_{lk}. \quad (148) \]

where the fourth line was obtained by using Property 2, according to the geometric series \( \sum_{j=0}^{n-1} x^j = (1 - x^n) / (1 - x) \), while the fifth line by using that \( w \) is the \( n \)-th root of unity since \( w^n = 1 \).

Property 4: Efficient Quantum Circuit. The discrete Fourier transform \( (FT_N) \) is defined according to Eq. (143) but without the normalization factor and can be efficiently computed by rearranging the columns of the matrix to have the even numbered columns as the first \( N/2 \) columns followed by the \( N/2 \) odd numbered columns (labeled with index \( k = 0 - (N/2) \)), and rows labeled with
index \( j = 0 - (N/2) \), as follows:

\[
FT(N) = \begin{bmatrix}
  w^{2k} & w^{(2k+1)} \\
  w^{(j+N/2)2k} & w^{(j+N/2)(2k+1)}
\end{bmatrix}.
\] (149)

We note that \( w^N = 1 \), and \( w^{N/2} = -1 \), so \( w^{(j+N/2)(2k)} = w^{(2kj+Nk)} = w^{2kj} \) and \( w^{(j+N/2)(2k+1)} = w^{(2kj+j+Nk+N/2)} = -w^{(2k+1)} \), giving

\[
FT_N = \begin{bmatrix}
w^{2k} & w^{j}w^{2k} \\
w^{2k} & -w^{j}w^{2k}
\end{bmatrix}.
\] (150)

Therefore, the \( FT_N \) can be applied to an \( N \)-bit string \( X_N \) by applying the \( FT_{N/2} \) to \( N/2 \)-bit strings, as follows:

\[
FT_NX_N = \begin{bmatrix}
FT_{N/2} & w^{j}FT_{N/2} \\
FT_{N/2} & -w^{j}FT_{N/2}
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_2 \\
\vdots \\
x_{N-2} \\
x_1 \\
x_3 \\
\vdots \\
x_{N-1}
\end{bmatrix},
\] (151)

where \( FT_{N/2} \) is applied to the top \( N/2 \) even-numbered bits, and \( w^{j}FT_{N/2} \) is applied to the bottom \( N/2 \) odd-numbered bits. The top \( N/2 \) bits are updated by summing the two contributions while the bottom \( N/2 \) bits are updated by subtracting the two contributions, according to the following 'butterfly' circuit:

![Figure 42: Classical 'butterfly' circuit for implementation of the FFT algorithm.](image)

Therefore, the computational cost \( C(N) \) of \( FT_N \) is \( C(N) = 2C(N/2) + O(N) \). If we recurse \( \log(N) \) times, we obtain: \( C(N/2) = 2C(N/4) + O(N/2), \ldots, C(4) = 2C(2) + O(4), \) and, \( C(2) = \)
2C(1) + O(2). So, the total cost of the discrete Fourier transform $F_T N$ implemented recursively is $O(N \log N)$, and is therefore called Fast Fourier Transform (FFT) algorithm [Cooley and Tukey Math. Comp. 19 (1965), 297-301]. By making the Fourier transform $O(N \log N)$ (i.e., much faster than the naive implementation with cost $O(N^2)$, the FFT algorithm enabled a wide range of revolutionary technologies.

Here, we show that the quantum Fourier transform algorithm implements the FT in $O(N \log N)$ elementary operations (i.e., in fewer steps than the number of operations it would take to input the initial $N$-bit state to be transformed (!)). For example, the $QFT_N$ of an input with $N = 2^{1000}$ would take only $10^6$ operations. While this might sound impossible, it is in fact achievable because the $N$-bit input can be encoded in a superposition of states of $n = \log N$ qubits (i.e., $n = 1000$).

Using $n = \log_2(N/2) + 1$ qubits, the quantum circuit that performs the $QFT_N$ looks, as follows:

$$\begin{array}{c}
(n-1) \\
\vdots \\
QFT_{N/2} \\
\vdots \\
H \\
\end{array}$$

Figure 43: Quantum circuit for implementation of the $QFT_N$, with $w = e^{i2\pi/N}$

Note that the quantum circuit, introduced by Fig. 43, has only one $QFT_{N/2}$ unit, in stark contrast with the classical circuit (Fig. 42), since the odd and even numbered bits can be encoded as superposition states of the same set of $n-1$ qubits, using the $n$-th qubit in state $|0\rangle$, or $|1\rangle$, for encoding even or odd numbered bits, respectively. Therefore, the outcome of the $QFT_{N/2}$ resulting from the odd numbered bits (i.e., $S_{j+N/2}$, with $j = 0 - (N/2 - 1)$), is also encoded as a superposition state of the first $n-1$ qubits when the $n$-th qubit is in state $|1\rangle$. To multiply each state $S_{j+N/2}$ of the superposition by $w^j$, we encode $j$ in binary, in terms of $(n-1)$ bits that can be either one or zero ($j_0, \ldots, j_{n-2}$), corresponding to the $n-1$ qubits, as follows: $j = \sum_{k=0}^{n-2} j_k 2^k$, so $w^j = \prod_{k=0}^{n-2} w^{j_k 2^k}$ and can be applied by applying a phase shift defined by $w^{2^k}$ to the outcome of the $k$-th wire when that $k$-th bit is $|1\rangle$ (i.e., when $j_k = 1$) and the $n$-th qubit is also $|1\rangle$.

Remarkably, the butterfly sums and differences of $S_j$ and $S_{j+N/2}$, described in Fig. 42, can be efficiently performed by applying a Hadamard gate to the $n$-th qubit. To understand that magic step, consider the effect of the Hadamard gate on the last qubit when the outcome of the $QFT_{N/2}$ is written as a superposition of all possible $n-1$ strings $|x\rangle$, as follows:

$$|\psi\rangle = \sum_{y \in \{0,1\}} \sum_{x \in \{0,1\}^{n-1}} c_{y,x}|y,x\rangle,$$

$$= |0\rangle \sum_{x \in \{0,1\}^{n-1}} c_{0,x}|x\rangle + |1\rangle \sum_{x \in \{0,1\}^{n-1}} c_{1,x}|x\rangle,$$

where the coefficients $c_{0,x}$ correspond to the superposition of outcomes $S_j$ from even numbered bits, and $c_{1,x}$ correspond to the superposition of $w^j S_{j+N/2}$ resulting from odd numbered bits. Ap-
plying the Hadamard gate to the ancilla \( n \)-th qubit, we obtain:

\[
H | \psi \rangle = \sum_{x \in \{0,1\}^{n-1}} (c_{0,x} | + \rangle + c_{1,x} | - \rangle) | x \rangle,
\]

\[
= \frac{1}{\sqrt{2}} \sum_{x \in \{0,1\}^{n-1}} (c_{0,x} + c_{1,x}) | 0 \rangle + (c_{0,x} - c_{1,x}) | 1 \rangle | x \rangle.
\]

Therefore, the sum and differences correspond to the superposition of the first \( n - 1 \) qubits when the \( n \)-th qubit is \(| 0 \rangle\) and \(| 1 \rangle\), respectively, as provided by the Hadamard transformation with the correct normalization factor.

Note that the computational cost of the QFT is \( S(n) = S(n - 1) + \mathcal{O}(n) \). So, the total computational cost over \( n \) recursion layers is \( n + (n - 1) + \cdots + 1 = n(n + 1)/2 \), or \( S(n) = \mathcal{O}(n^2) = \mathcal{O}(\log^2 N) \) for an input of size \( N = 2^n \). The enhanced version of QFT improves it to \( \mathcal{O}(n \log n) \) \cite{Hales2000}. So, the QFT is exponentially faster than the classical FFT algorithm. However, there is caveat! While performing the QFT would be exponentially fast, reading the outcome for a very large \( N \) would still be challenging.

The simplest QFT circuit (beyond the circuit with \( N = 2 \) (i.e., \( n = 1 \) for which the QFT is the Hadamard gate) is the Fourier transform of a state with \( N = 4 \) (i.e., with \( n = 2 \)), which is implemented as follows:

\[
\begin{align*}
| j_1 \rangle & \rightarrow H \rightarrow R_2 \rightarrow \frac{1}{2\pi} (| 0 \rangle + e^{i2\pi(j_1/2 + j_2/4)} | 1 \rangle) \\
| j_2 \rangle & \rightarrow H \rightarrow \frac{1}{2\pi} (| 0 \rangle + e^{i2\pi j_2/2} | 1 \rangle)
\end{align*}
\]

Figure 44: Quantum circuit for \( \text{QFT}_N \), with \( N = 4 \), where \( R_k = e^{i2\pi/2^k} \).

Analogously, the circuit for \( N = 8 \) (i.e., with \( n = 3 \)), is implemented as follows:

\[
\begin{align*}
| j_1 \rangle & \rightarrow H \rightarrow R_2 \rightarrow R_3 \rightarrow \frac{1}{2\pi} (| 0 \rangle + e^{i2\pi(j_1/2 + j_2/4 + j_3/8)} | 1 \rangle) \\
| j_2 \rangle & \rightarrow H \rightarrow R_2 \rightarrow \frac{1}{2\pi} (| 0 \rangle + e^{i2\pi j_2/2} | 1 \rangle) \\
| j_3 \rangle & \rightarrow H \rightarrow \frac{1}{2\pi} (| 0 \rangle + e^{i2\pi j_3/2} | 1 \rangle)
\end{align*}
\]

Figure 45: Quantum circuit for \( \text{QFT}_N \), with \( N = 8 \), where \( R_k = e^{i2\pi/2^k} \).

To obtain the implementation, shown in Fig. 45, we consider \( \text{QFT}_N | j \rangle = 2^{-n/2} \sum_k w^{jk} | k \rangle \).
Writing \( k \) in binary, \( k = \sum_{i=1}^{n} 2^{i-1}k_i \), we obtain:

\[
\text{QFT}_N |j\rangle = 2^{-n/2} \sum_k e^{2\pi i k |j\rangle},
\]

\[
= 2^{-n/2} \sum_{k_{1}=0}^{1} \ldots \sum_{k_{n}=0}^{1} e^{2\pi i \sum_{l=1}^{n} 2^{-l}k_{l}} |k_{1}\rangle \ldots |k_{n}\rangle,
\]

\[
= 2^{-n/2} \otimes_{l=1}^{n} \sum_{k_{l}=0}^{1} e^{2\pi i 2^{l-1}k_{l}} |k_{l}\rangle,
\]

\[
= 2^{-n/2} \otimes_{l=1}^{n} \left( |0\rangle + e^{2\pi i 2^{-1}} |1\rangle \right),
\]

\[
= 2^{-n/2} \otimes_{l=1}^{n} \left( |0\rangle + e^{2\pi i \sum_{m=n-l+1}^{n} 2^{-m-1}} |1\rangle \right),
\]

\[
= 2^{-n/2} \left( |0\rangle + e^{2\pi i 0, j_1} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i 0, j_{n-1}} |1\rangle \right) \ldots \otimes \left( |0\rangle + e^{2\pi i 0, j_{2} \cdots j_{n}} |1\rangle \right),
\]

(154)

where the line before the last one is obtained by writing \( j \) in binary, \( j = \sum_{m=1}^{n} j_m 2^{n-m} \), so \( 2^{-j} = \sum_{m=1}^{n-1} j_m 2^{n-1-m} + \sum_{m=n-l+1}^{n} j_m 2^{n-m-1} \), and for all \( m \) of the first sum \((n-l-m)\) is an integer, so \( e^{2\pi i 2^{-1}} = \left( \prod_{m=n-l+1}^{n} e^{2\pi i j_m 2^{n-m-1}} \right) e^{2\pi i \sum_{m=n-l+1}^{n} j_m 2^{n-m-1}} \), with \( \prod_{m=n-l+1}^{n} e^{2\pi i j_m 2^{n-m-1}} = 1 \). The last line is obtained by introducing the popular notation \( 0, j_1 j_{1+1} \cdots j_n = j_1 2^{-1} + j_{1+1} 2^{-2} + \ldots + j_n 2^{-(1+n-l)} \).

When \( n = 3 \), we obtain as shown in Fig. [45]

\[
\text{QFT}_3 |j\rangle = 2^{-1/2} \left( |0\rangle + e^{2\pi i j_1} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i j_2} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i j_3} |1\rangle \right).
\]

\[
= 2^{-3/2} \left( |0\rangle + e^{2\pi i j_1 j_2} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i (j_1 2^{-1} + j_2 2^{-2})} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i (j_1 2^{-1} + j_2 2^{-2} + j_3 2^{-3})} |1\rangle \right),
\]

\[
= 2^{-3/2} \left( |0\rangle + e^{2\pi i 0, j_3} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i 0, j_2 j_3} |1\rangle \right) \otimes \left( |0\rangle + e^{2\pi i 0, j_1 j_2 j_3} |1\rangle \right).
\]

(155)

### 17.2 Quantum Phase Estimation

Fourier sampling can be used for phase estimation. A simple example is given by the unitary that performs the following transformation \( U |\psi_j\rangle = e^{2\pi i \theta |\psi_j\rangle} \) where \( \theta = j / 2^n \), with \( j \) an integer between 0 and \( 2^n - 1 \) (later we will discuss what happens when \( j \) is not an integer but rather a non-integer real number between 0 and \( 2^n - 1 \)). When \( U \) is applied conditionally, it operates on \( |\psi_j\rangle \), as follows: c-U|\rangle \rangle \langle \rangle \rangle = |\rangle \rangle e^{i k 2^n |\rangle \rangle} \langle \rangle \rangle, \) where |\rangle \rangle is an n-bit string that defines the k number of times that the unitary U is applied on \( |\psi_j\rangle \) (with \( k = 0, \ldots, (2^n - 1) \)).

In the rest of this subsection we show that the phase \( \theta \) can be estimated by first applying c-U to a uniform superposition of states \( |\rangle \rangle \) (prepared, as follows: \( H^{\otimes n} |0\rangle \rangle = \text{QFT}_n |0\rangle \rangle^{\otimes n} \)). Then, we exploit the phase kickback algorithm, we compute the inverse Fourier transform, and we measure, according to the following circuit:

\[
|0\rangle \rangle^{\otimes n} \xrightarrow{H^{\otimes n}} |\psi_j\rangle \xrightarrow{U} \xrightarrow{\text{QFT}_n^{-1}} |j\rangle \xrightarrow{\text{Phase estimate}} |\psi_j\rangle
\]

(156)
We note that the state after the Hadamard gate, right before the conditional unitary, is the uniform superposition \( \frac{1}{2^n} \sum_{k=0}^{2^n-1} |k\rangle |\psi_j\rangle \). Therefore, the state after the unitary is \( \frac{1}{2^n} \sum_{k=0}^{2^n-1} |k\rangle (U^k |\psi_j\rangle) = \frac{1}{2^n} \sum_{k=0}^{2^n-1} |k\rangle e^{i2\pi \theta} |\psi_j\rangle \). Introducing the substitutions \( \theta = j/2^n \) and \( \omega = e^{i2\pi /2^n} \), and invoking the phase kickback trick, the resulting state is

\[
\left( \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle \omega^k \right) |\psi_j\rangle,
\]

which, according to Eq. (142), is equal to the \( (QFT_{2^n} |j\rangle) |\psi_j\rangle \). Therefore, we can obtain \( |j\rangle \) by computing the inverse Fourier transform over the first \( n \) qubits and measuring. Also, note that \( |\psi_j\rangle \) is left unchanged so the process can be repeated multiple times using only one copy of \( |\psi_j\rangle \).

The rest of this section shows how to apply the c-U gate in the computational basis to transform the state \( |k\rangle |\psi_j\rangle \) into the state \( |k\rangle e^{i2\pi \theta} |\psi_j\rangle \), as shown in the circuit of Eq. (156). First, we write \( k \) in binary \( k = \sum_{l=1}^{n} 2^{n-l} k_l \). So, we obtain:

\[
\frac{1}{2^{n/2}} \sum_{k=1}^{2^n} |k\rangle e^{i2\pi \theta} = \frac{1}{2^{n/2}} \sum_{k_1=0}^{1} \cdots \sum_{k_{2^n}=0}^{1} |k_1 \cdots k_{2^n}\rangle e^{i2\pi \sum_{l=1}^{n} k_l 2^{n-l} \theta},
\]

so \( |k\rangle e^{i2\pi \theta} |\psi_j\rangle = \frac{1}{2^{n/2}} \otimes_{l=1}^{n} \sum_{k_l=0}^{1} e^{i2\pi k_l 2^{-l} \theta} |k_l\rangle |\psi_j\rangle = \frac{1}{2^{n/2}} \otimes_{l=1}^{n} (|0\rangle |\psi_j\rangle + |1\rangle e^{i2\pi 2^{-l} \theta} |\psi_j\rangle) \). Therefore,

\[
|k\rangle e^{i2\pi \theta} |\psi_j\rangle = \frac{1}{2^{n/2}} (|0\rangle |\psi_j\rangle + |1\rangle e^{i2\pi 2^{-1} \theta} |\psi_j\rangle) \otimes \ldots \otimes (|0\rangle |\psi_j\rangle + |1\rangle e^{i2\pi 2^{-n} \theta} |\psi_j\rangle),
\]

According to Eq. (159), the circuit to implement the c-U gate in the computational basis is, as follows:

\[
\begin{array}{c}
|0\rangle \\
|0\rangle \\
|0\rangle \\
|\psi_j\rangle \\
\end{array}
\]

\[
\begin{array}{c}
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
H_{\otimes n} \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
U^{2^{n-1}} \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
QFT^{-1} \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

\[
\begin{array}{c}
U^2 \\
U^{2^n} \\
\end{array}
\]

Figure 46: Circuit for quantum phase estimation.

Now, substituting \( j = \sum_{m=1}^{n} j_m 2^{n-m} \), we obtain:

\[
|k\rangle e^{i2\pi \theta} |\psi_j\rangle = \frac{1}{2^{n/2}} \otimes_{l=1}^{n} e^{i2\pi k_l \sum_{m=1}^{n} j_m 2^{n-m} |k_l\rangle |\psi_j\rangle},
\]

Therefore,

\[
|k\rangle e^{i2\pi \theta} |\psi_j\rangle = \frac{1}{2^{n/2}} \otimes_{l=1}^{n} (|0\rangle + |1\rangle e^{i2\pi \sum_{m=1}^{n} j_m 2^{n-m} |k_l\rangle |\psi_j\rangle),
\]

\[
= QFT_{2^n} |j\rangle |\psi_j\rangle,
\]

where the second line of Eq. (161) is obtained by comparing the first line to Eq. (154).

77
Probability of Success in Phase Estimation

For an arbitrary value of \( \theta - i.e., \) regardless of whether \( \theta \) is a multiple of \( 1/2^n \) or not, we obtain that the state after the unitary is
\[
\frac{1}{2^n} \sum_{k=0}^{2^n-1} \sum_{l=0}^{2^n-1} |l\rangle e^{i2\pi k(\theta-l/2^n)} |\psi_j\rangle.
\]
and
\[
\frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{i2\pi k(\theta-j/2^n)} |\psi_j\rangle.
\]
Therefore, the probability of measuring the first \( n \)-qubits in state \( |l\rangle = |j\rangle \) is
\[
P_j = \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{i2\pi k(\theta-j/2^n)} \right|^2.
\]
Similarly, we find that
\[
P_j = \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{i2\pi (\theta-j/2^n)} \right|^2 - \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{i2\pi (\theta-j/2^n)} \right|^2.
\]
Clearly, when \( \theta = j'/2^n \), \( P_j = \delta_{jj'} \). More generally (i.e., when \( \theta \) is not a multiple of \( 1/2^n \)), using the geometric series \( \sum_{k=0}^{2^n-1} x^k = (x^{2^n} - 1)/(x - 1) \), we obtain:
\[
P_j = \left| \frac{1}{2^n} \frac{(e^{i2\pi (\theta-j/2^n)} - 1)}{(e^{i2\pi (\theta-j/2^n)} - 1)} \right|^2.
\]
where \( \theta = j/2^n + \epsilon \), so \( a = e^{i2\pi 2^n \epsilon} - 1 \), and \( b = e^{i2\pi \epsilon} - 1 \). Here, \( \epsilon \) is the difference between the phase \( \theta \) and the best possible \( n \)-bit approximation \( j/2^n \), so \( 0 < \epsilon < 1/2^n \).

Representing \( e^{i2\pi 2^n \epsilon} \) and \( e^{i2\pi \epsilon} \) in the complex plane, as shown in Fig. 47, we find that \( 2\pi|\epsilon|/a < \pi/2 \) since that ratio is maximum for \( 2\pi|\epsilon|/a = \pi \), for which \( a = 2 \) (Fig. 47 left). In addition, we find that \( 2\pi|\epsilon| \geq b \) (Fig. 47 right). Substituting these inequalities into Eq. (164), we obtain:
\[
P_j = \frac{1}{2\pi} \frac{a^2}{b^2},
\]
\[
\geq \frac{1}{2\pi} \frac{(4|\epsilon|2^n)^2}{(2\pi|\epsilon|)^2} = \frac{4}{\pi^2} > 0.4.
\]
Therefore, the probability of measuring all $n$ bits of $j$ correctly (such that $|j/2^n - \theta| \leq 2^{-n}$) is higher than 40%. In addition, since both the first and second closest multiple of $2^{-n}$ are within $2^{-n}$ of the actual value to be determined, the phase is estimated within an error of $1/2^n$ with probability $8/\pi^2 > 0.8$.

Next, we address the probability $pr(|m - j| > k)$ of measuring a value $m$ beyond the $k2^{-n}$ accuracy, as defined by set of $2k$ multiples of $2^{-n}$ closer to $j$, shown in Fig. 48:

$$pr(|m - j| > k) = \sum_{l=-2^{n-1}+1}^{-k} P_{j+l} + \sum_{l=k+1}^{2^{n-1}} P_{j+l},$$

(166)

with $P_{j+l}$ defined according to Eq. (164), with $\theta = j/2^n + \epsilon$:

$$P_{l+j} = \frac{1}{2^{2n}} \left| \frac{e^{i2\pi(2^n\epsilon-l)} - 1}{e^{i2\pi(\epsilon-l/2^n)} - 1} \right|^2.$$  

(167)

Figure 48: $\frac{j}{2^n}$ is the multiple of $\frac{1}{2^n}$ nearest to $\theta$, surrounded by the nearer $2k$ values $\frac{n}{2^n}$ such that $|m - j| \leq k$.

**Bounds:** An expression for $pr$ based on Eq. (166) has been reported [J. M. Chappell et al.]. Here, we use an upper bound of Eq. (167), according to the standard approach (Nielsen and Chuang, p. 224), by considering that the numerator $|e^{i\phi} - 1| = |e^{i\phi/2} - e^{-i\phi/2}| = 2|\sin(\phi/2)| \geq 2|\phi|/\pi$, when $0 < \phi < \pi$ (since $\left| \sin \left( \frac{\phi}{2} \right) \right| \geq \left| \frac{\phi}{2} \right|$). In particular, when $\phi = 2\pi(e - l/2^n)$, we find $|e^{i\phi} - 1| \geq 4|(e - l/2^n)|$. We note that $-\pi \geq \phi = 2\pi(e - l/2^n) \geq \pi$ since $-2^{n-1} + 1 \leq l \leq 2^{n-1}$. So, substituting $l$ by its upper bound $2^{n-1}$ we find $2\pi(e - 2^{-1}) \leq 2\pi(e - l/2^n) - i.e., -\pi \leq 2\pi(e - l/2^n)$. Analogously, substituting $l$ by its lower bound $-2^{n-1} + 1$ we find $2\pi(e - l/2^n) \leq 2\pi(e + 2^{-1} - 1/2^n)$, and since $\epsilon < 2^{-n}$, we obtain $2\pi(e - 1/2^n) \leq \pi$.

Substituting $|e^{i2\pi(\epsilon-l)} - 1| \leq 2$ and $|e^{i2\pi(\epsilon-l/2^n)} - 1| > 4(e - l/2^n)$ into Eq. (168), we obtain:

$$P_{l+j} < \frac{1}{2^{2n}} \left| \frac{1}{2(e - l/2^n)} \right|^2.$$  

(168)
Substituting Eq. (168) into Eq. (166), we obtain:

\[
\begin{align*}
pr(|m - j| > k) &\leq \frac{1}{4} \left( \sum_{l=1}^{k} \frac{1}{(2^k - l)^2} + \sum_{l=k+1}^{2^k-1} \frac{1}{(2^k - l)^2} \right), \\
&< \frac{1}{4} \left( \sum_{l=1}^{k} \frac{1}{l^2} + \sum_{l=k+1}^{2^k-1} \frac{1}{(l-1)^2} \right), \\
&< \frac{1}{4} \left( \sum_{l=1}^{k} \frac{1}{l^2} + \sum_{l=k+1}^{2^k-1} \frac{1}{l^2} \right), \\
&< \frac{1}{2(2^k - 1)}.
\end{align*}
\]

where the second line is obtained by using \(2^n \theta < 1\). In the third line, we introduced the variable transformation \(l = -1\) for the first sum, and \(l = (l - 1)\) for the second sum. Therefore,

\[
\begin{align*}
pr(|m - j| > k) &< \frac{1}{2} \sum_{l=k}^{2^{n-1}-1} \frac{1}{l^2}, \\
&< \frac{1}{2} \int_{k-1}^{2^{n-1}-1} dl \frac{1}{l^2} < \frac{1}{2} \int_{k-1}^{\infty} dl \frac{1}{l^2}, \\
&< \frac{1}{2(k-1)}.
\end{align*}
\]

According to Eq. (170), it is clear that the phase is estimated within an error of \(k/2^n\) with probability at least as high as \(pr(|m - j| < k) = 1 - \frac{k}{2^{n-1}}\).

Therefore, to estimate \(\theta\) up to the first \(r\) bits (with \(r < n\)) — i.e., with an accuracy of \(2^{-r}\), we choose \(k \leq 2^{n-r} - 1\). Using \(n = r + p\) qubits in the first register, the probability of estimating \(\theta\) within the desired error margin is at least \(1 - \frac{1}{2^{n-1}} = 1 - \frac{1}{2(2^r - 2)} = 1 - pr\). Solving for \(p\) in terms of \(pr\), we obtain: \(p = \log_2(2 + \frac{1}{2pr})\). Therefore, the number of qubits necessary to estimate \(\theta\) with accuracy of \(2^{-r}\) and probability of success higher than \(1 - pr\) is \(n = r + \log_2(2 + \frac{1}{2pr})\).

**Probabilistic Algorithm:** As shown above, the probability of measuring all \(n\) bits of \(j\) correctly (such that \(|j/2^n - \theta| \leq 2^{-n}\)) is higher the 40%. Here, we show that a probabilistic algorithm can increase that probability by running the calculations with more bits, taking the most commonly occurring outcome and round it to \(n\) bits of precision. In fact, the probability of getting the correct \(n\) bits with that algorithm approaches 100 % exponentially fast with the number of times the procedure is repeated. The tools necessary to show the capabilities of probabilistic algorithms are the following bounds, or inequalities, of the tails of probability distributions.

**Markov’s inequality:** The Markov’s inequality is based on the first moment \(E(x)\) of a distribution of a random variable \(x\), as follows:

\[
P(x \geq \epsilon) \leq \frac{E(x)}{\epsilon},
\]

indicating that the probability \(P(x \geq \epsilon)\) of sampling \(x \geq \epsilon\) in the upper tail of the distribution is always less or equal than the average value \(E(x)\) divided by \(\epsilon\). To prove Eq. (171), we consider that \(E(x) = P(x \geq \epsilon)E(x \geq \epsilon) + P(x \leq \epsilon)E(x \leq \epsilon)\), with \(E(x \leq \epsilon) \geq 0\) and \(E(x \geq \epsilon) \geq \epsilon\). Therefore, \(E(x) \geq P(x \geq \epsilon)\), as shown in Eq. (171).

Substituting \(\epsilon = aE(x)\), into Eq. (171), we obtain another equivalent form of Markov’s inequality:

\[
P(x \geq aE(x)) \leq \frac{1}{a}.
\]
Chebyshev inequality: Introducing into Eq. (171) the substitution $x = (y - E(y))^2$ with $E(x) = E((y - E(y))^2) = \sigma^2(y)$, and $\epsilon = \alpha^2$, we obtain:

$$P((y - E(y))^2 \geq \alpha^2) \leq \frac{\sigma^2(y)}{\alpha^2}, \quad (173)$$

or

$$P(|y - E(y)| \geq \alpha) \leq \frac{\sigma^2(y)}{\alpha^2}, \quad (174)$$

so

$$P(|y - E(y)| \geq \alpha \sigma(y)) \leq \frac{1}{\alpha^2}, \quad (175)$$

which is known as the Chebyshev inequality based on the second moment $\sigma(y)$ of the distribution of the random variable $y$.

Chernoff Bounds: Here, we obtain sharper bounds (specifically, exponential bounds rather than polynomial bounds as those based on the first and second moments of the distribution) by addressing the particular case of a random variable $x$ defined as the sum of independent random bits $x_j = \{0, 1\}$ with equal probabilities $p_j = p$ that $x_j = 1$. A simple example is a sequence of measurements $j = 1, \cdots, n$, with probability $p_j$ that the $j$-th measurement provides the correct output (e.g., the result within a given range). So, $x = \sum_j x_j$ and $E(x) = \sum_j p_j = np$.

According to Eq. (171), $P(y \geq \epsilon) \leq \frac{E(y)}{\epsilon}$, so introducing the variable transformation $y = e^{tx}$ with $t > 0$ and $\epsilon = e^{ta}$, we obtain:

$$P(e^{tx} \geq e^{ta}) \leq \frac{E(e^{tx})}{e^{ta}}, \quad (176)$$

$$P(x \geq a) \leq e^{-ta} \prod_j E(e^{tx_j}),$$

giving a bound for the upper tail of the probability distribution.

When all independent measurements $j$ have the same probability of success (i.e., $x_j = 1$), the variables are called Bernoulli random variables $x_1, \cdots, x_n$ with $p_j = p \geq 0$, and we obtain:

$$E[e^{tx_j}] = pe^t + (1 - p),$$

$$= 1 + p(e^t - 1),$$

$$\leq e^{p(e^t-1)} = 1 + p(e^t - 1) + \cdots, \quad (177)$$

with $e^t > 1$.

For any $\delta > 0$, defining $t = \ln(1 + \delta) > 0$, we obtain:

$$E[e^{tx_j}] < e^{p\delta}, \quad (178)$$

and

$$e^{-ta} = \frac{1}{(1 + \delta)^a}. \quad (179)$$

Substituting Eqs. (178) and (179) into Eq. (176), with $a = (1 + \delta)np$, where $n$ is an arbitrary integer, we obtain:

$$P(x \geq (1 + \delta)np) < \frac{e^{np\delta}}{(1 + \delta)^{1+\delta}np} = \left[\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right]^{np}. \quad (180)$$
Considering that for this case \( E(x) = np \), we obtain the upper tail Chernoff bound (Theorem A.1.4 in Kaye-Laflamme-Mosca Introduction to Quantum Computing):

\[
P(x \geq (1 + \delta)E(x)) < \left[ \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right]^{E(x)}. \tag{181}
\]

Another form of the upper tail Chernoff bound can be found by taking the logarithm of the r.h.s. of Eq. (181) and using the inequality \( \log(1 + \delta) \geq \delta / (1 + \delta / 2) \), we obtain:

\[
E(x)(\delta - (1 + \delta)\log(1 + \delta)) \leq -\frac{E(x)\delta^2}{(2 + \delta)}, \tag{182}
\]

which allows us to obtain a more convenient (although looser bound) than Eq. (181) by exponentiating, as follows:

\[
P(x \geq (1 + \delta)E(x)) < e^{-\frac{E(x)\delta^2}{(1 + \delta)\log(1 + \delta)}}. \tag{183}
\]

Analogously, we can obtain a Chernoff bound for the lower tail of the probability distribution by introducing the random variable \( y = e^{-tx} \) with \( t > 0 \) and substituting into the Markov's inequality \( P(y \geq \epsilon) \leq \frac{E(y)}{\epsilon} \) given by Eq. (176), with \( \epsilon = e^{-ta} \):

\[
P(e^{-tx} \geq e^{-ta}) = P(x \leq a) \leq \frac{E(e^{-tx})}{e^{-ta}} \leq e^{ta} \prod_j E(e^{-tx_j}), \tag{184}
\]

Defining \( t = -\ln(1 - \delta) > 0 \), we obtain:

\[
e^{-ta} = (1 - \delta)^a. \tag{185}
\]

and

\[
E[e^{-tx_j}] = pe^{-t} + (1 - p),
\]

\[= 1 + p(e^{-t} - 1),
\]

\[\leq e^{pe^{-t} - 1}. \tag{186}
\]

Substituting Eqs. (185) and (186) into Eq. (184), with \( a = (1 - \delta)np \), we obtain:

\[
P(x \leq (1 - \delta)np) \leq \frac{E(e^{-tx})}{e^{-ta}} \leq \frac{e^{np(e^{-t} - 1)}}{(1 - \delta)(1 - \delta - 1)} \leq \left[ \frac{e^{(1-\delta)}(1-\delta)}{1 - \delta(1-\delta)} \right]^{np} \leq \left[ \frac{e^{-\delta}}{1 - \delta(1-\delta)} \right]^{np} \tag{187}
\]

Another form of the Chernoff bound for the lower tail can be obtained by taking the log of the r.h.s. of Eq. (187) and using the inequality \( \ln(1 - \delta) \geq \delta(\delta/2 - 1) / (1 - \delta) \), when \( 0 \leq \delta \leq 1 \), we obtain:

\[
np(-\delta - (1 - \delta)\ln(1 - \delta)) \leq np(-\delta - \delta(1 - \delta)(\delta/2 - 1) / (1 - \delta)),
\]

\[\leq -np\delta^2 / 2. \tag{188}
\]
Exponentiating both sides of Eq. (188), we obtain:

\[
\left[ \frac{e^{-\delta}}{(1 - \delta)^{(1-\delta)}} \right]^{np} \leq \left[ e^{-\frac{\delta^2}{2}} \right]^{np},
\]  

(189)

and substituting into Eq. (187), we obtain (Theorem A.1.3 in Kaye-Laflamme-Mosca Introduction to Quantum Computing):

\[
P(x \leq (1 - \delta)np) \leq \left[ e^{-\frac{\delta^2}{2}} \right]^{np}
\]  

(190)

Practical Application: The Chernoff bound, introduced by Eq. (190), can be used to show that the probability of failure (i.e., the probability that the algorithm gives an answer outside the desired error range from the expected value) is reduced exponentially (i.e., as \( \gamma^n \) with \( 0 < \gamma < 1 \), and \( n \) the number of times the algorithm is repeated with the same input). Therefore, we can amplify the success probability of a bounded-error algorithm simply by repeating the algorithm.

Considering that the probability of success of a single trial is \( p = 1/2 + \beta/2 \), with \( 0 < \beta < 1 \), the number of times the algorithm is expected to provide the correct answer in \( n \) trials is \( E(x) = np = \frac{n}{2} (1 + \beta) \). Defining \( \delta = \beta/(1 + \beta) \) and substituting into Eq. (190), we obtain:

\[
P(x \leq \frac{n}{2}) \leq e^{-\frac{\beta^2}{4(1+\beta)}} = \gamma^n,
\]  

(191)

with \( \gamma = e^{-\frac{\beta^2}{4(1+\beta)}} \). Eq. (191) gives the probability that the correct answer is obtained in fewer than 50 \% of the times after \( n \) trials. Therefore, obtaining the correct answer more that 50 \% of the times with probability higher than \( 1 - \epsilon \) requires \( n \) to be sufficiently large so that \( \gamma^n < \epsilon \), as follows:

\[
e^{-\frac{\beta^2}{4(1+\beta)}} < \epsilon,
\]  

\[n > \frac{4(1+\beta)}{\beta^2} \ln \left( \frac{1}{\epsilon} \right).
\]  

(192)

Coin bias estimation: The Chernoff bounds are useful for all kinds of probabilistic calculations, beyond the problem of phase estimation. Perhaps the simplest example is the problem of estimating the bias of a coin, as determined by a slight curvature (Fig. 49), so it lands with heads facing up more often than 50 \% of the times (i.e., with probability \( p = 1/2 + \beta/2 \), with \( 0 < \beta < 1 \)). So the problem is: how many times \( n \) do we need to toss the coin to determine the bias within \( \delta \) with probability higher than \( 1 - \epsilon \)?

Figure 49 shows that, after tossing the coin \( n \) times, \( j \) is the integer nearest to the expectation value \( \bar{n}_h = n(1+\beta)/2 \). Therefore, \( j \) gives the best possible estimator of the bias \( \beta \), as follows: \( E(\beta) = (2j - n)/n \), with probability that \( j \) of the \( n \) times shows heads facing up:

\[
P^{(n)}_j = \binom{n}{j} p^j (1-p)^{n-j}.
\]  

(193)

According to Eq. (192), when \( n > \frac{4(1+\beta)}{\beta^2} \ln \left( \frac{1}{\epsilon} \right) \), we will obtain heads facing up for more than half of the times (i.e., \( j > n/2 \)), with probability at least \( 1 - \epsilon \).
17.3 Period Finding

The Fourier transform of a periodic superposition state $|f\rangle = \sum_{j=0}^{n-1} a_j |j\rangle$, with period $r$, is a periodic superposition $|\tilde{f}\rangle = \sum_{j=0}^{n-1} \beta_j |j\rangle$ with period $n/r$. A particular example is the function with only one non-zero coefficient per period (i.e., a total of $n/r$ non-zero coefficients), $a_j = \sqrt{\frac{r}{n}}$, for $j = kr$, with $k = 0, 1, 2, \cdots, n/r - 1$, so that $|f\rangle = \sqrt{\frac{r}{n}} \sum_{k=0}^{n/r-1} |kj\rangle$. The discrete Fourier transform of $|f\rangle$ is

$$QFT^{(n)} \sum_{k=0}^{n-1} \sqrt{\frac{r}{n}} |kr\rangle = \sum_{k=0}^{n-1} \sqrt{\frac{r}{n}} \frac{1}{\sqrt{n}} \sum_{l=0}^{n-1} w^{lk} |l\rangle,$$

where $j = lr/n$ while $l'$ is not a multiple of $n/r$. Therefore, the second term is zero since $\sum_{k=0}^{n-1} w^{lrk} = 0$. Considering that $w^n = 1$, we obtain:

$$QFT^{(n)} \sum_{k=0}^{n-1} \sqrt{\frac{r}{n}} |kr\rangle = \sqrt{\frac{1}{r}} \sum_{l=0}^{r-1} |l\rangle.$$

Therefore, measurements of the resulting superposition state after applying the Fourier transform provide equally probable indices $ln/r$ with $l = 0, 1, 2, \cdots, r - 1$. The minimum common factor divisor of all outcomes is $n/r$ which provides the period $r$ since $n$ is known.

We note that the particular state discussed in this subsection (i.e., a superposition state with only one non-zero coefficient per period) could be prepared by measuring the ancilla bits prepared...
in state $|f(x)\rangle$ when the state of the circuit is in the superposition state $|\psi\rangle = 2^{-n/2} \sum_{x \in \{0,1\}^n} |x\rangle |f(x)\rangle$, as right after applying the unitary of the function in the following circuit:

![Quantum circuit](image)

Figure 50: Quantum circuit for determining the period of $f(x)$ by first preparing a superposition state with only one non-zero coefficient per period in the first $n$ qubits, by measuring the $n$ ancilla bits prepared in state $f(x)$, when the first $n$ working bits are in a uniform superposition. The subsequent QFT then generates a state with non-zero coefficients for those strings with indices that are $n/r$ periodic.

We note that a measurement reporting a value of the function $f(x_j)$ would leave the state in the superposition,

$$|\psi\rangle = \sqrt{\frac{r}{N}} \sum_{k=0}^{\frac{r}{N}-1} |x_j + kr\rangle |f(x_j)\rangle,$$

for which the Fourier transform of the first $n$-bits is shift-invariant (Property 1). Therefore,

$$QFT^{(n)} \left( \sqrt{\frac{r}{N}} \sum_{k=0}^{\frac{r}{N}-1} |x_j + kr\rangle |f(x_j)\rangle \right) = QFT^{(n)} \left( \sqrt{\frac{r}{N}} \sum_{k=0}^{\frac{r}{N}-1} |kr\rangle |f(x_j)\rangle \right),$$

$$= \sqrt{\frac{1}{r}} \sum_{l=0}^{r-1} |\frac{n}{r}\rangle |f(x_j)\rangle.$$ (197)

At this point we must mention that, according to the so-called principle of deferred measurement, it is not even necessary to measure $f(x)$ since measuring commutes with conditioning (i.e., a measurement can be replaced by a CNOT with with an acilla that is measured at the end).

### 17.4 Shor’s Algorithm

The goal of this section is to explain Shor’s algorithm for finding the prime factors of an integer $N$, as applied to finding the prime factors 7 and 3 of $N = 21$.

Shor’s algorithm is essentially the period finding algorithm, discussed in the previous section, applied for the specific case of the function $f(x) = m^x \pmod{N}$, where $m$ is a random integer $m < N$. As an example, we choose $m = 2$ and we evaluate $f(x)$ for $x = 1, 2, 3, \cdots$, as shown in the table below, by considering that $f(x)$ is the remainder of $\frac{m^x}{N}$, or $m^x = Nq + f(x)$, where $q$ is an integer.
We note that \( f(x) \) is a single-valued periodic function, with period \( r = 6 \) that could be efficiently resolved by the period finding algorithm described in the previous section (Fig. 50). Having found the period \( r \), we know that

\[
1 = 2^r \pmod{21},
\]

or

\[
y = 2^{r/2} = 2^3 = 8.
\]

Therefore,

\[
1 = 8^2 \pmod{21},
\]

or

\[
8^2 = 1 \pmod{21}.
\]

So, by period finding we found that \( y = 8 \) is a non-trivial root of unity \((\pmod{N})\), since

\[
8^2 = 1 \pmod{21} \text{ and } (8 + 1) \cdot (8 - 1) = 0 \pmod{21}.
\]

So, we find that 21 divides \((8 + 1) \cdot (8 - 1)\), although it is not divided either \((8 + 1)\) or \((8 - 1)\). How could this be possible? Well, by writing 21 in terms of its prime factors, we find that 21 = 3 · 7, so 3 divides \((8 + 1)\) and 7 divides \((8 - 1)\) \(i.e.,\) the prime factors of \( N \) are the greatest common divisors \( \gcd(N, y + 1) \) and \( \gcd(N, y - 1) \) which can be found very efficiently by Euclid's algorithm. The algorithm has a success rate of at least 50% since \( r \) is even number half of the times and occasionally, we find that \( \gcd(N, m) > 1 \) \(i.e.,\) we find a prime factor for free through the Euclid's algorithm. If \( r \) is odd, we fail, since we cannot find \( y \), so we need to pick another random number \( m \) and try again.
Coherent States

The goal of this section is to introduce coherent states $|\alpha\rangle$, as discussed in the quantum optics community for descriptions of states of coherent light as eigenstates of the annihilation operator:

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle,$$  \hspace{1cm} (198)

where $\alpha$ is a complex number. A coherent state simply has a precise phase, defined by the complex amplitude $\alpha$, although an indefinite number of photons as in a laser beam. In contrast, a state with a fixed number of photons usually has completely arbitrary (random) phase.

The rest of this section shows that we can create coherent states, as follows:

$$\hat{D}(\alpha)|0\rangle = |\alpha\rangle,$$  \hspace{1cm} (199)

where $|0\rangle$ is the vacuum state defined as the ground state of the harmonic oscillator, and $\hat{D}$ is the displacement operator, defined as follows:

$$\hat{D}(\alpha) = e^{\alpha \hat{a}^\dagger - \alpha^* \hat{a}},$$  \hspace{1cm} (200)

The second row is obtained from the first one, making use of the Hausdorff formula $e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}$, with $A = \alpha \hat{a}^\dagger$ and $B = -\alpha^* \hat{a}$, which is valid if $[A,[A,B]] = [B,[A,B]] = 0$ as in this case. Note that $[A,B] = -|\alpha|^2 [\hat{a}^\dagger, \hat{a}] = |\alpha|^2$ since $[\hat{a}^\dagger, \hat{a}] = -1$.

We also note that the inverse must be

$$\hat{D}(\alpha)^{-1} = e^{\frac{1}{2}|\alpha|^2 \hat{a}^\dagger e^{-\alpha \hat{a}}},$$  \hspace{1cm} (201)

since $\hat{D}(\alpha)^{-1}\hat{D}(\alpha) = 1$. The second row of Eq. (201) is obtained from the first one since

$$e^{\alpha \hat{a}^\dagger e^{-\alpha^* \hat{a}}} = e^{-\frac{1}{2}|\alpha|^2 e^{-\alpha \hat{a}^\dagger e^{-\alpha^* \hat{a}}} e^{-|\alpha|^2}}.$$  \hspace{1cm} (202)

Note that multiplying both sides of Eq. (202) by $e^{\alpha \hat{a}^\dagger e^{-\alpha^* \hat{a}}} = \hat{D}(\alpha)e^{\frac{1}{2}|\alpha|^2}$, we obtain:

$$1 = \hat{D}(\alpha)e^{\frac{1}{2}|\alpha|^2 \hat{a}^\dagger} \hat{D}(\alpha) e^{\frac{1}{2}|\alpha|^2} e^{-|\alpha|^2}.$$  \hspace{1cm} (203)

Therefore, according to Eq. (201),

$$\hat{D}(\alpha)^{-1} = \hat{D}(-\alpha) = \hat{D}(\alpha)^\dagger.$$  \hspace{1cm} (204)

The Backer Campbell Hausdorff relation,

$$e^A e^B e^{-A} = B + [A,B] + \frac{1}{2} [A,[A,B]] + \cdots,$$  \hspace{1cm} (205)

can be used with $A = -\alpha \hat{a}^\dagger + \alpha^* \hat{a}$, and $B = \hat{a}$ to show that

$$\hat{D}(\alpha)^\dagger \hat{a} \hat{D}(\alpha) = \hat{a} + \alpha,$$  \hspace{1cm} (206)
since \([A, B] = [−α\hat{a}^† + \alpha^∗\hat{a}, \hat{a}] = α\), and therefore \([A, [A, B]] = 0\). Applying Eq. \ref{eq:206} to the vacuum state \(|0\rangle\), we obtain:

\[
\hat{D}(\alpha)^†\hat{a}\hat{D}(\alpha)|0\rangle = \alpha|0\rangle,
\]

(207)

since \(\hat{a}|0\rangle = 0\), or

\[
\hat{a}\hat{D}(\alpha)|0\rangle = \alpha\hat{D}(\alpha)|0\rangle,
\]

(208)

since \(\hat{D}(\alpha)^† = \hat{D}(\alpha)^−1\). Therefore, according to Eq. \ref{eq:198},

\[
\hat{D}(\alpha)|0\rangle = |\alpha\rangle,
\]

(209)

which is Eq. \ref{eq:199}.

Substituting Eq. \ref{eq:200} into Eq. \ref{eq:209}, we obtain:

\[
|\alpha\rangle = e^{−\frac{1}{2}|\alpha|^2}e^{\alpha\hat{a}^†}e^{−\alpha^∗\hat{a}}|0\rangle,
\]

(210)

and expanding the exponentials in Taylor series (i.e., \(e^A = \sum∞_{n=0} \frac{1}{n!} A^n\), with \(\hat{a}|0\rangle = 0\)), we obtain:

\[
|\alpha\rangle = e^{−\frac{1}{2}|\alpha|^2} \sum_{n=0}^∞ \frac{\alpha^n}{n!} (\hat{a}^†)^n |0\rangle,
\]

(211)

\[
= e^{−\frac{1}{2}|\alpha|^2} \sum_{n=0}^∞ \frac{\alpha^n}{\sqrt{n!}} |n\rangle,
\]

where in the second row, we used \(\hat{a}|n\rangle = \sqrt{n+1} |n+1\rangle\).

We note that the average number of photons in a coherent state \(|\alpha\rangle\) is given by the square of the complex amplitude \(|\alpha|^2\) (the intensity of the wave) since

\[
\langle N \rangle = \langle \alpha|\hat{a}^†\hat{a}|\alpha\rangle,
\]

\[
= \langle \alpha|\alpha^∗\alpha|\alpha\rangle,
\]

\[
= |\alpha|^2\alpha.
\]

(212)

The probability \(P(n)\) of finding \(n\) photons is

\[
P(n) = |\langle n|\alpha\rangle|^2,
\]

(213)

\[
= e^{−|\alpha|^2} \frac{|\alpha|^n}{n!},
\]

defined by the Poisson distribution,

\[
P(n) = \frac{\lambda^n e^{−\lambda}}{n!},
\]

(214)

with \(\lambda = |\alpha|^2\).

**Dynamics:** The dynamical properties of a coherent-state in a harmonic well of frequency \(\omega\) can be described by the time-dependent correlation function,

\[
\langle \alpha|\hat{a}(t)^†\hat{a}|\alpha\rangle = |\alpha|^2 e^{i\omega t},
\]

(215)

where \(\hat{a}(t) = e^{\frac{i\hat{H}t}{\hbar}}\hat{a}e^{−\frac{i\hat{H}t}{\hbar}}\), and \(|\alpha\rangle = \sum_{n=0}^∞ e^{−\frac{|\alpha|^2}{2}} \frac{\alpha^n}{\sqrt{n!}} |n\rangle\). Therefore, the dynamics of a harmonic oscillator and its absorption spectrum can be simulated by the following python script. Analogously, the simulation of a 2-dimensional harmonic oscillator in a 2-level system can be simulated with qudit with the following python script or notebook python script.
To confirm Eq. (215), we note that \( \hat{H}|n\rangle = E_n|n\rangle \), with \( E_n = \hbar \omega (n + 1/2) \), so \( e^{-i\hat{H}t}|\alpha\rangle = \sum_{n=0}^{\infty} e^{-\frac{|\alpha|^2}{2}} \frac{\alpha^n}{\sqrt{n!}} e^{-i\frac{\sigma_n^a}{\hbar}E_n t}|n\rangle \). Further, \( \hat{\alpha}|n\rangle = \sqrt{n}|n-1\rangle \), so we obtain

\[
\hat{a}(t)|\alpha\rangle = e^{i\hat{H}t}\hat{a}e^{-i\hat{H}t}|\alpha\rangle = \sum_{n=0}^{\infty} e^{-\frac{|\alpha|^2}{2}} \frac{\alpha^n}{\sqrt{n!}} e^{-i\left(E_n - E_{n-1}\right)t}\sqrt{n}|n-1\rangle,
\]

\[
= \alpha \sum_{n=0}^{\infty} e^{-\frac{|\alpha|^2}{2}} \frac{\alpha^n}{\sqrt{n!}} e^{-i\omega t}|n\rangle,
\]

\[
= \alpha e^{-i\omega t}|\alpha\rangle.
\]

Therefore, \( \hat{a}|\alpha\rangle = \alpha|\alpha\rangle \), and computing the adjoint we obtain \( \langle \alpha | \hat{a}(t)^\dagger = \alpha^* \langle \alpha | e^{i\omega t}, giving \langle \alpha | \hat{a}(t)^\dagger \hat{a} | \alpha\rangle = |\alpha|^2 e^{i\omega t} \), which is Eq. (215).

18.1 Parity Operator

We define the parity operator, as follows:

\[
\hat{\Pi} = \sum_{j=0}^{\infty} |2j\rangle \langle 2j| - \sum_{j=0}^{\infty} |2j+1\rangle \langle 2j+1|,
\]

\[
= e^{i\pi \hat{a}^\dagger \hat{a}},
\]

where states \( |2j\rangle \) and \( |2j+1\rangle \) in the first row of Eq. (217) are eigenstates of the number operator \( \hat{N} = \hat{a}^\dagger \hat{a} \) with eigenvalues \( 2j \) and \( 2j + 1 \), respectively. Note that both expressions of \( \hat{\Pi} \), introduced by Eq. (217), change the sign of an eigenstate of the number operator when the occupation number is odd, and leave the eigenstate unchanged if the occupation is even. Consequently, when applied to a coherent-states,

\[
\hat{\Pi}|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \hat{\Pi}|n\rangle
\]

\[
= e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} (-1)^n |n\rangle
\]

\[
= e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{(-\alpha)^n}{\sqrt{n!}} |n\rangle
\]

\[
= | - \alpha\rangle.
\]

19 Dicke Model and Jaynes-Cummings Hamiltonian

The goal of this section is to introduce the famous Jaynes Cummings model for the description of a qubit-resonator. In cavity quantum electrodynamics (or cavity QED), the experimental setup involves a single two-level atom (i.e., qubit), with energy level spacing \( \hbar \omega_q \), placed between two highly reflecting mirrors, as shown in Fig. 51, forming a small cavity where photons of a single mode of an electromagnetic field of frequency \( \omega \), are confined and interact with the atom.

The qubit-resonator model system is described by the [Dicke Hamiltonian] (with \( \hbar = 1 \)),

\[
H = \omega_c(a^\dagger a) + \frac{1}{2} \omega_q \sigma_z + g(a^\dagger + a)\sigma_x,
\]

89
Figure 51: Two-level system interacting with an electromagnetic field in a cavity, defined by two mirrors that confine a single mode of an oscillatory electromagnetic field with frequency $\omega_c$.

where $\omega_c$ and $\omega_q$ are the bare frequencies of the resonator and qubit, respectively, and $g$ is the strength of the dipole interaction coupling the atom and field, which is inversely proportional to the volume of the cavity.

Neglecting the high frequency terms according to the rotating wave approximation (RWA), we obtain the [Jaynes-Cummings model] Hamiltonian,

$$
H \approx \omega_c (a^\dagger a) + \frac{1}{2} \omega_q \sigma_z + \frac{g}{2} (a^\dagger \sigma^- + a \sigma^+),
$$

(220)

The coupling $g$ can be made much stronger than the coupling to all the other modes of the electromagnetic field in the surrounding environment simply by making a very small cavity (e.g., with the mirrors very close to each other) so that the volume $V$ of the cavity is small and thus the density of states in the cavity is much larger than the density of states in the environment (so the Purcell factor $F_p = \frac{3}{4\pi} \left( \frac{\lambda}{n} \right)^3 \frac{Q}{P}$ is large). Here, $Q$ is the quality factor, $n$ is the refractive index of the medium and $\lambda$ is the frequency of the free photon. So, the 'transverse' decay rate $\gamma$ of the atom due to coupling with the surrounding field can be made very small. In addition, the photon decay described by the cavity decay $\kappa$ can be made very small by making highly reflecting super-polished mirrors.

Analogously, the experiment can be setup with superconducting circuits (i.e., the so-called 'circuit QED', the superconducting circuit analogue of 'cavity QED') where a transmon qubit Josephson junction (i.e., an 'artificial atom') is capacitively coupled to a mode of an electromagnetic field resonating in a coplanar superconducting waveguide (CPW) resonator (shown in Fig. 52), or with photons in a 3D cavity resonator.

Here, we will describe how to simulate the time-dependent evolution of the 2-level system when the frequency of the photon matches the frequency of the 2-level energy gap. Starting with the two-level system in the excited state and no photons in the cavity, the system exhibits Rabi oscillations due to emission of a single photon and coherent reabsorption of that photon multiple times (the so-called vacuum Rabi oscillations). We discuss how to compute the Rabi oscillations numerically by following a QuTiP lecture tutorial or python script.

Note: Install QuTiP by following the installation instructions:

conda create -n qutip-env python=3
conda install numpy scipy matplotlib pytest pytest-cov jupyter notebook spyder
conda config --append channels conda-forge
conda install qutip

Remember to work with QuTiP in that environment by running

conda activate qutip-env
We also discuss the dispersive limit to resolve the photon number states of a microwave resonator by monitoring a qubit that was coupled to the resonator, as reported by D. I. Schuster et al. Resolving photon number states in a superconducting circuit [Nature 445, 515 (2007)], and how to simulate it with a QuTiP lecture tutorial or python script.

In the dispersive (off-resonance) limit (Fig. 53), the resonator and qubit are far off-resonance, $\Delta \gg g$, where $\Delta = |\omega_r - \omega_q|$ is the detuning between the resonator and the qubit (for example $\omega_r \gg \omega_q$), so only virtual photon exchange is allowed. To see that a single photon has a large effect on the qubit without ever being absorbed, we can transform the Hamiltonian introduced by the RWA of Eq. (220), $H = H_0 + V$, with $H_0 = \omega_c a^\dagger a + \frac{1}{2} \omega_q \sigma_z$ and $V = g (a^\dagger + a) \sigma_x$, according to the similarity transformation $\tilde{H} = e^{\xi H} e^{-\xi}$. Using the Backer Campbell Hausdorff relation, introduced by Eq. (209), we obtain the perturbative expansion $\tilde{H} = H_0 + V + [\xi, H_0] + [\xi, V] + \frac{1}{2!} [\xi, [\xi, H_0]] + \cdots$ Choosing $\xi = g (a \sigma^+ - a^\dagger \sigma^-)$, we find that $[\xi, H_0] = -V$, and keeping terms up to second order in $g$ (i.e., solving to second order after a dispersive approximation), we obtain that the effective Hamiltonian in the dispersive limit is approximately,

$$\tilde{H} = \omega_r (a^\dagger a + 1/2) + \frac{1}{2} \omega_q \sigma_z + \chi (a^\dagger a + 1/2) \sigma_z,$$  \hspace{1cm} (221)  

where $\chi = g^2 / \Delta$. We can view the last term as a correction of the resonator frequency that depends on the qubit state, or a correction to the qubit frequency that depends on the resonator state.
Figure 53: Parameter diagram for the Jaynes-Cummings model defined by the atom-photon coupling strength, $g$, and the detuning frequency $\Delta = \omega_q - \omega_c$, normalized to the decay rates $\Gamma = \max[\gamma, \kappa, 1/T]$.

## 20 Kerr Hamiltonian

The goal of this section is to show that the Kerr Hamiltonian

$$\hat{H}_R = K (\hat{a}^\dagger)^2 \hat{a}^2,$$  \hspace{1cm} (222)

corresponds to the Hamiltonian of the anharmonic oscillator

$$\hat{H}_s = \hbar \omega \left( \hat{N} + \frac{1}{2} \right) + K \hat{N}^2,$$  \hspace{1cm} (223)

but in the reference frame rotating with frequency $\omega$.

The second term in Eq. (223) introduces the anharmonicity, while the first term corresponds to the Hamiltonian of the harmonic oscillator,

$$\hat{H}_h = \hbar \omega \left( \hat{N} + \frac{1}{2} \right),$$

defined in terms of the number operator $\hat{N} = \hat{a}^\dagger \hat{a}$, where the annihilation $\hat{a} = \frac{1}{\sqrt{2}} (\hat{x} + i\hat{p})$ and creation $\hat{a}^\dagger = \frac{1}{\sqrt{2}} (\hat{x} - i\hat{p})$ operators are defined in terms of the reduced coordinate, $\hat{x} = \hat{x} \sqrt{\frac{\hbar}{m \omega}}$, and momentum, $\hat{p} = \hat{p} \sqrt{\frac{\hbar}{m \omega}}$, operators.

We rewrite $\hat{H}_s$ with the 'so-called normal order of creation and annihilation operators (i.e., where for each term the powers of $\hat{a}^\dagger$ are followed by powers of $\hat{a}$) by using the commutator
\[ [\hat{a}, \hat{a}^\dagger] = 1, \text{ as follows:} \]
\[
\hat{H}_s = \hbar \omega \left( \hat{a}^\dagger \hat{a} + \frac{1}{2} \right) + K \hat{a}^\dagger \hat{a} \hat{a}^\dagger \hat{a},
\]
\[
= \hbar \omega \left( \hat{a}^\dagger \hat{a} + \frac{1}{2} \right) + K \hat{a}^\dagger (1 + \hat{a}^\dagger \hat{a}),
\]
\[
= \hbar \omega \left( \hat{a}^\dagger \hat{a} + \frac{1}{2} \right) + K(\hat{a}^\dagger \hat{a} + \hat{a}^\dagger \hat{a} \hat{a}^\dagger \hat{a}),
\]
\[
= (\hbar \omega + K) \hat{a}^\dagger \hat{a} + \frac{\hbar \omega}{2} + K(\hat{a}^\dagger)^2 \hat{a}^2, \tag{225}
\]
\[
= \hat{H}_v + K(\hat{a}^\dagger)^2 \hat{a}^2,
\]
\[
= \hat{H}_v + K(\hat{N}^2 - \hat{N}).
\]

To obtain the anharmonic Hamiltonian \( \hat{H}_R \) in the rotating frame, we consider an initial state \(|\Psi(0)\rangle\) evolving according to such a Hamiltonian \(|\Psi_R(t)\rangle = \hat{U}_R |\Psi(0)\rangle\), where \( \hat{U}_R = e^{-\frac{i}{\hbar} \hat{H}_R t} \). The resulting state \(|\Psi_R(t)\rangle\) can also be generated by first propagating the system in the static frame \(|\Psi_s(t)\rangle = \hat{U}_s |\Psi(0)\rangle\) using the time-evolution operator \( \hat{U}_s = e^{-\frac{i}{\hbar} \hat{H}_s t} \) defined by the Hamiltonian in the static frame \( \hat{H}_s \), and then transform the propagated state to the rotated frame by applying the transformation \( \hat{U} = e^{\frac{i}{\hbar} \hat{H}_R t} \), where \( \hat{H}_v = (\hbar \omega + K) \hat{N} + \frac{\hbar \omega}{2} \) is the Hamiltonian that defines the unitary 'rotation' \( \hat{U} \) with frequency \( \omega \), from the static to the rotated frame. Therefore,
\[
|\Psi_R(t)\rangle = \hat{U}_R |\Psi(0)\rangle,
\]
\[
= \hat{U}_s \hat{U}_R |\Psi(0)\rangle. \tag{226}
\]

The Hamiltonian \( \hat{H}_R \) can now be obtained by computing the time-derivative of \( \frac{\partial \hat{U}_R}{\partial t} = -\frac{i}{\hbar} \hat{H}_R \hat{U}_R \), and solving for \( \hat{H}_R \), as follows:
\[
\hat{H}_R = i\hbar \frac{\partial \hat{U}_R}{\partial t} \hat{U}_R^\dagger,
\]
\[
= i\hbar \frac{\partial \hat{U}_s \hat{U}_R}{\partial t} \hat{U}_R^\dagger,
\]
\[
= i\hbar \frac{\partial \hat{U}_s}{\partial t} \hat{U}_s^\dagger \hat{U}_R + i\hbar \hat{U}_s \frac{\partial \hat{U}_s}{\partial t} \hat{U}_R^\dagger, \tag{227}
\]
\[
= i\hbar \frac{\partial \hat{U}_s}{\partial t} \hat{U}_R + i\hbar \hat{U}_s \frac{\partial \hat{U}_R}{\partial t} \hat{U}_s^\dagger \hat{U}_R^\dagger,
\]
\[
= i\hbar \frac{\partial \hat{U}_s}{\partial t} \hat{U}_R + \hat{U}_s \hat{H}_s \hat{U}_R^\dagger,
\]
\[
= -\hat{H}_v + \hat{U}_s \hat{H}_s \hat{U}_R^\dagger.
\]

Therefore, we obtain:
\[
\hat{H}_R = -\hat{H}_v + e^{\frac{i}{\hbar} \hat{H}_s t} \left( \hat{H}_v + K(\hat{N}^2 - \hat{N}) \right) e^{-\frac{i}{\hbar} \hat{H}_s t},
\]
\[
= K(\hat{N}^2 - \hat{N}), \tag{228}
\]
\[
= K(\hat{a}^\dagger)^2 \hat{a}^2,
\]

where the second line was obtained considering that \( e^{\frac{i}{\hbar} \hat{H}_s t} \) commutes with both \( \hat{H}_v = \frac{\hbar \omega}{2} + (\hbar \omega + K) \hat{N} \), and \( K(\hat{N}^2 - \hat{N}) \), and the third line considering that \( (\hat{a}^\dagger)^2 \hat{a}^2 = (\hat{N}^2 - \hat{N}) \).
The evolution of a coherent state, according to the Kerr Hamiltonian $\hat{H}_R$ can be analytically computed, as follows:

$$e^{-\frac{i}{\hbar}\hat{H}_R t}|\alpha\rangle = \sum_{n=0}^{\infty} e^{-\frac{|\alpha|^2}{2}} \frac{\alpha^n}{\sqrt{n!}} e^{-\frac{i}{\hbar}K(n^2-n)t} |n\rangle,$$

so both the number of photons and the Poisson probability distribution of photons on Fock states $|n\rangle$ remain unchanged.
21 Python

Python is a great general-purpose programming language on its own, but with the help of a few popular libraries (numpy, scipy, matplotlib) it becomes a powerful environment for scientific computing. I expect that many of you will have some experience with Python and numpy; for the rest of you, this section will serve as a quick crash course both on the Python programming language and on the use of Python for scientific computing.

In this tutorial, we will cover:

- Basic Python: Basic data types (Containers, Lists, Dictionaries, Sets, Tuples), Functions, Classes
- Numpy: Arrays, Array indexing, Datatypes, Array math, Broadcasting
- Matplotlib: Plotting, Subplots, Images
- IPython: Creating notebooks, Typical workflows

21.1 A Brief Note on Python Versions

As of January 1, 2020, Python has officially dropped support for python2. We'll be using Python 3.7 for this iteration of the course. You can check your Python version at the command line by running `python --version`. In Colab, we can enforce the Python version by clicking Runtime -> Change Runtime Type and selecting python3. Note that as of April 2020, Colab uses Python 3.6.9 which should run everything without any errors.

```
[6]: !python --version
Python 3.6.9
```

21.1.1 Basics of Python

Python is a high-level, dynamically typed multiparadigm programming language. Python code is often said to be almost like pseudocode, since it allows you to express very powerful ideas in very few lines of code while being very readable. As an example, here is an implementation of the classic quicksort algorithm in Python:

```
[7]: def quicksort(arr):
    if len(arr) <= 1:
        return arr
    pivot = arr[len(arr) // 2]
    left = [x for x in arr if x < pivot]
    middle = [x for x in arr if x == pivot]
    right = [x for x in arr if x > pivot]
    return quicksort(left) + middle + quicksort(right)

print(quicksort([3,6,8,10,1,2,1]))

[1, 1, 2, 3, 6, 8, 10]
```

21.1.1.1 Basic data types
21.1.1.1 Numbers

Integers and floats work as you would expect from other languages:

```
[8]: x = 3
   print(x, type(x))
3 <class 'int'>
```

```
[9]: print(x + 1)     # Addition
    print(x - 1)     # Subtraction
    print(x * 2)     # Multiplication
    print(x ** 2)    # Exponentiation
4
2
6
9
```

```
[10]: x += 1
    print(x)
    x *= 2
    print(x)
4
8
```

```
[11]: y = 2.5
    print(type(y))
    print(y, y + 1, y * 2, y ** 2)
<class 'float'>
2.5 3.5 5.0 6.25
```

Note that unlike many languages, Python does not have unary increment (x++) or decrement (x--) operators.

Python also has built-in types for long integers and complex numbers; you can find all of the
details in the documentation.

21.1.1.2 Booleans

Python implements all of the usual operators for Boolean logic, but uses English words rather than symbols (&&, ||, etc.):

```
[12]: t, f = True, False
    print(type(t))
<class 'bool'>
```

Now we let’s look at the operations:
21.1.1.1.3 Strings

```python
hello = 'hello'  # String literals can use single quotes
world = "world"  # or double quotes; it does not matter
print(hello, len(hello))
```

hello 5

```python
hw = hello + ' ' + world  # String concatenation
print(hw)
```

hello world

```python
hw12 = '{} {} {}'.format(hello, world, 12)  # string formatting
print(hw12)
```

hello world 12

String objects have a bunch of useful methods; for example:

```python
s = "hello"
print(s.capitalize())  # Capitalize a string
print(s.upper())  # Convert a string to uppercase; prints "HELLO"
print(s.rjust(7))  # Right-justify a string, padding with spaces
print(s.center(7))  # Center a string, padding with spaces
print(s.replace('l', '(ell)'))  # Replace all instances of one substring with another
print(' world '.strip())  # Strip leading and trailing whitespace
```

Hello
HELLO
   hello
hello
he(ell)(ell)o
world

You can find a list of all string methods in the documentation.
21.1.1.2 Containers  Python includes several **built-in container types**: lists, dictionaries, sets, and tuples.

1. List item
2. List item
3. List item
4. List item

### 21.1.1.2.1 Lists

A list is the Python equivalent of an array, but is resizeable and can contain elements of different types:

```
[18]: xs = [3, 1, 2]  # Create a list
    print(xs, xs[2])
    print(xs[-1])  # Negative indices count from the end of the list; prints "2"

[3, 1, 2] 2
2
```

Lists can be generated from arrays, as follows:

```
[20]: import numpy as np

int_list = []  # list initialization
int_list = [0,0,1,2,3]  # list with commas
int_list.append(4)  # add 4 to end of the list
int_list.pop(2)  # remove element with index 2

int_list2 = list(range(5))  # make list [0,1,2,3,4]
int_array = np.array(int_list)  # make array [] with no commas: [0 1 2 3 4]
int_array2 = np.arange(5)  # make array [] with no commas: [0 1 2 3 4]
int_list2 = int_array.tolist()  # convert array to list

first = 0
last = 4
float_array = np.linspace(first,last,num=5)

print('int_list=',int_list)
print('int_list2=',int_list2)
print('int_array=',int_array)
print('int_array2=',int_array2)
print('float_array=',float_array)

int_list= [0, 0, 2, 3, 4]
int_list2= [0, 0, 2, 3, 4]
int_array= [0 0 2 3 4]
```

98
int_array2 = [0 1 2 3 4]
float_array = [0.1.2.3.4.]

[21]: xs[2] = 'foo'  # Lists can contain elements of different types
print(xs)
[3, 1, 'foo']

Lists have methods, including append, insert, remove, sort

[22]: xs.append('bar')  # Add a new element to the end of the list
print(xs)
[3, 1, 'foo', 'bar']

[23]: x = xs.pop()  # Remove and return the last element of the list
print(x, xs)
bar [3, 1, 'foo']

As usual, you can find all the gory details about lists in the documentation.

21.1.1.2.2 Slicing  In addition to accessing list elements one at a time, Python provides concise syntax to access sublists; this is known as slicing:

[24]: nums = list(range(5))  # range is a built-in function that creates a list of integers
print(nums)  # Prints "[0, 1, 2, 3, 4]"
print(nums[2:4])  # Get a slice from index 2 to 4 (exclusive); prints "[2, 3]"
print(nums[2:])  # Get a slice from index 2 to the end; prints "[2, 3, 4]"
print(nums[:2])  # Get a slice from the start to index 2 (exclusive); prints "[0, 1]"
print(nums[:])  # Get a slice of the whole list; prints "[0, 1, 2, 3, 4]"
print(nums[:-1])  # Slice indices can be negative; prints "[0, 1, 2, 3]"
nums[2:4] = [8, 9]  # Assign a new sublist to a slice
print(nums)  # Prints "[0, 1, 8, 9, 4]"

[0, 1, 2, 3, 4]
[2, 3]
[2, 3, 4]
[0, 1]
[0, 1, 2, 3, 4]
[0, 1, 2, 3]
[0, 1, 8, 9, 4]

21.1.1.2.3 Loops  You can loop over the elements of a list like this:
```python
[25]: animals = ['cat', 'dog', 'monkey']

for animal in animals:
    print(animal)

cat
dog
monkey

If you want access to the index of each element within the body of a loop, use the built-in `enumerate` function:

```python
[26]: animals = ['cat', 'dog', 'monkey']

for idx, animal in enumerate(animals):
    print('#{}: {}'.format(idx + 1, animal))

#1: cat
#2: dog
#3: monkey
```

### 21.1.1.2.4 List comprehensions:
When programming, frequently we want to transform one type of data into another. As a simple example, consider the following code that computes square numbers:

```python
[27]: nums = [0, 1, 2, 3, 4]
squares = []

for x in nums:
    squares.append(x ** 2)
print(squares)

[0, 1, 4, 9, 16]
```

You can make this code simpler using a list comprehension:

```python
[28]: nums = [0, 1, 2, 3, 4]
squares = [x ** 2 for x in nums]
print(squares)

[0, 1, 4, 9, 16]
```

List comprehensions can also contain conditions:

```python
[29]: nums = [0, 1, 2, 3, 4]

even_squares = [x ** 2 for x in nums if x % 2 == 0]
print(even_squares)

[0, 4, 16]
```

### 21.1.1.2.5 Dictionaries
A dictionary stores (key, value) pairs, similar to a `Map` in Java or an object in Javascript. You can use it like this:
Create a new dictionary with some data:
```
d = {}  # Create a new dictionary
```
```
d = {'cat': 'cute', 'dog': 'furry'}  # Create a new dictionary with some data
```
```
# Get an entry from a dictionary; prints "cute"
print(d['cat'])
```
```
# Check if a dictionary has a given key; prints "True"
print('cat' in d)
```
```
wet
```
```
# Set an entry in a dictionary
print(d['fish'])  # Prints "wet"
```
```
# KeyError: 'monkey' not a key of d
print(d['monkey'])  # KeyError: 'monkey' not a key of d
```
```
N/A
```
```
# Get an element with a default; prints "N/A"
print(d.get('monkey', 'N/A'))  # Get an element with a default; prints "N/A"
```
```
N/A
```
```
# Get an element with a default; prints "wet"
print(d.get('fish', 'N/A'))  # Get an element with a default; prints "wet"
```
```
```
```
# Remove an element from a dictionary
print(d.get('fish', 'N/A'))  # "fish" is no longer a key; prints "N/A"
```
```
N/A
```
```
You can find all you need to know about dictionaries in the documentation.
```
It is easy to iterate over the keys in a dictionary:
```
```
d = {'person': 2, 'cat': 4, 'spider': 8}
```
```
for animal, legs in d.items():
    print('A {} has {} legs'.format(animal, legs))
```
```
A person has 2 legs
A cat has 4 legs
A spider has 8 legs
```
A person has 2 legs
A cat has 4 legs
A spider has 8 legs

Add pairs to the dictionary

```python
[36]: d[‘bird’]=2
```

List keys

```python
[37]: d.keys()
```

```python
[37]: dict_keys([‘person’, ‘cat’, ‘spider’, ‘bird’])
```

List values

```python
[38]: d.values()
```

```python
[38]: dict_values([2, 4, 8, 2])
```

Query values from keys

```python
[39]: d[‘bird’]
```

```python
[39]: 2
```

Dictionary comprehensions: These are similar to list comprehensions, but allow you to easily construct dictionaries. For example:

```python
[40]: nums = [0, 1, 2, 3, 4]
even_num_to_square = {x: x ** 2 for x in nums if x % 2 == 0}
print(even_num_to_square)
```

```python
{0: 0, 2: 4, 4: 16}
```

Convert array to list

21.1.1.2.6 Sets (like dictionaries but with no values, add & remove)

A set is an unordered collection of distinct elements. As a simple example, consider the following:

```python
[41]: animals = {‘cat’, ‘dog’}
print(‘cat’ in animals)  # Check if an element is in a set; prints "True"
print(‘fish’ in animals)  # prints "False"
```

```python
True
False
```

```python
[42]: animals.add(‘fish’)  # Add an element to a set
print(‘fish’ in animals)
print(len(animals))  # Number of elements in a set;
```

```python
True
3
```
Loops: Iterating over a set has the same syntax as iterating over a list; however since sets are unordered, you cannot make assumptions about the order in which you visit the elements of the set:

```python
animals = {'cat', 'dog', 'fish'}
for idx, animal in enumerate(animals):
    print('#{}: {}'.format(idx + 1, animal))
```

#1: fish
#2: dog
#3: cat

Set comprehensions: Like lists and dictionaries, we can easily construct sets using set comprehensions:

```python
from math import sqrt
print({int(sqrt(x)) for x in range(30)})
```

{0, 1, 2, 3, 4, 5}

21.1.1.2.7 Tuples  A tuple is an (immutable) ordered list of values. A tuple is in many ways similar to a list; one of the most important differences is that tuples can be used as keys in dictionaries and as elements of sets, while lists cannot. Here is a simple example:

```python
d = {(x, x + 1): x for x in range(10)}  # Create a dictionary with tuple keys
print(d)
```

```
tt = ()  # initialization of empty tuple
t1 = (66,)  # initialization of tuple with a single value
t = (5, 6)  # Create a tuple
tt = tt+t1+t
print("tt="+tt)
print("tt[2]="+tt[2])
print("tt[1:3]="+tt[1:3])
print("66 in tt", 66 in tt)
```

print(type(t))
print(d[t])
print(d[(1, 2)])
tt = (66, 5, 6)
tt[2] = 6
tt[1:3] = (5, 6)
66 in tt True
<class 'tuple'>
5
1

21.1.1.3 Functions

Python functions are defined using the def keyword. For example:

```python
[48]: def sign(x):
    if x > 0:
        return 'positive'
    elif x < 0:
        return 'negative'
    else:
        return 'zero'

for x in [-1, 0, 1]:
    print(sign(x))
```

negative
zero
positive

We will often define functions to take optional keyword arguments, like this:

```python
[49]: def hello(name, loud=False):
    if loud:
```
print('HELLO, {}'.format(name.upper()))
else:
    print('Hello, {}'.format(name))

hello('Bob')
hello('Fred', loud=True)

Hello, Bob!
HELLO, FRED

21.1.4 Classes A new class creates a new type of object, bounding data and functionality that allows new instances of the type made. Each class instance can have attributes attached to it, so we can make class instances as well as instances to variables and methods for maintaining the state of the class. Instances of the method can have attributes and can modify the state of the class, as clearly described by the [documentation](#).

The syntax for defining classes in Python is straightforward:

```python
[50]: class Greeter:
    
    """ My greeter class ""
    # Constructor (method of construction of class in a specific state)
    v1 = 'papa'  # class variable shared by all instances
    def __init__(self, name_inp):
        # name_inp: argument given to Greeter for class instantiation
        self.name = name_inp  # Create an instance variable maintaining the state
                                # instance variables are unique to each instance
    
    # Instance method
    # note that the first argument of the function method is the instance object
    def greet(self, loud=False):
        if loud:
            print('HELLO, {}'.format(self.name.upper()))
            self.name = 'Haote'
        else:
            print('Hello, {}'.format(self.name))
            self.name = 'Victor'

# Class instantiation (returning a new instance of the class assigned to g):
# Constructs g of type Greeter & initializes its state
# as defined by the class variables (does not execute methods)
g = Greeter('Fred')

# Call an instance method of the class in its current state:
```
Hello, Fred!
HELLO, VICTOR
papa
Hello, Haote!
Hello, Lea!
20

[50]: 'papa'

For loops (iterators). Behind the scenes, the for statement calls iter() on the container object.

[51]: for element in [1,2,3]: # elements of list
    print(element)
for element in (1,2,3): # elements of tuple
    print(element)
for key in {‘first’:1, ‘second’:2, ‘third’:3}: # elements of dictionary
    print(‘key=’,key)
for char in ’1234’:
    print(char)
#for line in open(‘myfile.txt
    # print(line,end=‘‘)
21.1.1.5 Modules A module is a .py file containing Python definitions and statements that can be imported into a Python script, as described in the Python documentation. As an example, after mounting your Google drive as described by the Navigating_tutorial.ipynb Jupyter notebook, use a text editor and write a module with the line:

```python
[greeting = "Good Morning!"
```

Save the document with the name mymod.py. Next, go the the folder where you saved that file and open a notebook with the lines:

```python
import mymod as my
print(my.greeting)
```

You will see that the notebook has imported the variable greeting from the module mymod.py and has invoked the variable as an attribute of the module mymod that was imported as my when printing Good Morning!!.

Modules are very convenient since they allow you to import variables, functions and classes that you might have developed for previous projects, without having to copy them into each program. So, you can build from previous projects, or split your work into several files for easier maintenance.

Within a module, the module's name (as a string) is available as the value of the global variable __name__.

21.1.2 Numpy

Numpy is the core library for scientific computing in Python. It provides a high-performance multidimensional array object, and tools for working with these arrays. If you are already familiar with MATLAB, you might find this tutorial useful to get started with Numpy.

To use Numpy, we first need to import the numpy package:

```python
import numpy as np
```

21.1.2.1 Arrays A numpy array is a grid of values, all of the same type, and is indexed by a tuple of nonnegative integers. The number of dimensions is the rank of the array; the shape of an array is a tuple of integers giving the size of the array along each dimension.
We can initialize numpy arrays from nested Python lists, and access elements using square brackets:

```python
[55]: a = np.array([1, 2, 3])  # Create a rank 1 array
    print(type(a), a.shape, a[0], a[1], a[2])
    a[0] = 5                 # Change an element of the array
    print(a)

<class 'numpy.ndarray'> (3,) 1 2 3
5 2 3
```

```python
[56]: b = np.array([[1,2,3],[4,5,6]])  # Create a rank 2 array
    print(b)

[[1 2 3]
 [4 5 6]]
```

```python
[57]: print(b.shape)
    print(b[0, 0], b[0, 1], b[1, 0])

(2, 3)
1 2 4
```

Numpy also provides many functions to create arrays:

```python
[58]: a = np.zeros((2,2))  # Create an array of all zeros
    print(a)

[[0. 0.]
 [0. 0.]]
```

```python
[59]: b = np.ones((1,2))  # Create an array of all ones
    print(b)

[[1. 1.]]
```

```python
[60]: c = np.full((2,2), 7)  # Create a constant array
    print(c)

[[7 7]
 [7 7]]
```

```python
[61]: d = np.eye(2)  # Create a 2x2 identity matrix
    print(d)

[[1. 0.]
 [0. 1.]]
```

108
21.1.2.2 Array indexing  Numpy offers several ways to index into arrays.

Slicing: Similar to Python lists, numpy arrays can be sliced. Since arrays may be multidimensional, you must specify a slice for each dimension of the array:  

```python
import numpy as np

# Create the following rank 2 array with shape (3, 4)
# [[ 1  2  3  4]
#  [ 5  6  7  8]
#  [ 9 10 11 12]]
a = np.array([[1,2,3,4], [5,6,7,8], [9,10,11,12]])

# Use slicing to pull out the subarray consisting of the first 2 rows
# and columns 1 and 2; b is the following array of shape (2, 2):
# [[2 3]
#  [6 7]]
b = a[:2, 1:3]
print(b)
```

```
[[2 3]
 [6 7]]
```

A slice of an array is a view into the same data, so modifying it will modify the original array.

```python
print(a[0, 1])
b[0, 0] = 77   # b[0, 0] is the same piece of data as a[0, 1]
print(a[0, 1])
```

```
2
77
```

You can also mix integer indexing with slice indexing. However, doing so will yield an array of lower rank than the original array. Note that this is quite different from the way that MATLAB handles array slicing:

```python
# Create the following rank 2 array with shape (3, 4)
a = np.array([[1,2,3,4], [5,6,7,8], [9,10,11,12]])
print(a)
```

```
[[ 1  2  3  4]
 [ 5  6  7  8]
 [ 9 10 11 12]]
```
Two ways of accessing the data in the middle row of the array. Mixing integer indexing with slices yields an array of lower rank, while using only slices yields an array of the same rank as the original array:

```
[66]:
row_r1 = a[1, :]
row_r2 = a[1:2, :]
row_r3 = a[[1], :]
print(row_r1, row_r1.shape)
print(row_r2, row_r2.shape)
print(row_r3, row_r3.shape)

[ 5 6 7 8] (4,)
[[ 5 6 7 8]] (1, 4)
[[ 5 6 7 8]] (1, 4)
```

```
[67]:
# We can make the same distinction when accessing columns of an array:

col_r1 = a[:, 1]
col_r2 = a[:, 1:2]
print(col_r1, col_r1.shape)

[[ 2 6 10] (3,)
[[ 2]
[ 6]
[10]] (3, 1)
```

Integer array indexing: When you index into numpy arrays using slicing, the resulting array view will always be a subarray of the original array. In contrast, integer array indexing allows you to construct arbitrary arrays using the data from another array. Here is an example:

```
[68]:
a = np.array([[1, 2], [3, 4], [5, 6]])

# An example of integer array indexing.
# The returned array will have shape (3,) and
print(a[[0, 1, 2], [0, 1, 0]])

# The above example of integer array indexing is equivalent to this:
print(np.array([a[0, 0], a[1, 1], a[2, 0]]))

[ 1 4 5]
[ 1 4 5]
```

```
[69]:
# When using integer array indexing, you can reuse the same
# element from the source array:
print(a[[0, 0], [1, 1]])
```

```
One useful trick with integer array indexing is selecting or mutating one element from each row of a matrix:

```python
# Create a new array from which we will select elements
a = np.array([[1,2,3], [4,5,6], [7,8,9], [10, 11, 12]])
print(a)

[[ 1 2 3]
 [ 4 5 6]
 [ 7 8 9]
[10 11 12]]

# Create an array of indices
b = np.array([0, 2, 0, 1])
# Select one element from each row of a using the indices in b
print(a[np.arange(4), b])  # Prints "[ 1 6 7 11]"

[ 1 6 7 11]

# Mutate one element from each row of a using the indices in b
a[np.arange(4), b] += 10
print(a)

[[11 2 3]
 [ 4 5 16]
 [17 8 9]
[10 21 12]]
```

Boolean array indexing: Boolean array indexing lets you pick out arbitrary elements of an array. Frequently this type of indexing is used to select the elements of an array that satisfy some condition. Here is an example:

```python
import numpy as np

a = np.array([[1,2], [3, 4], [5, 6]])
bool_idx = (a > 2)    # Find the elements of a that are bigger than 2;
# this returns a numpy array of Booleans of the
~same shape as a, where each slot of bool_idx tells
# whether that element of a is > 2.
```
print(bool_idx)

[[False False]
 [ True True]
 [ True True]]

# We use boolean array indexing to construct a rank 1 array
# consisting of the elements of a corresponding to the True values
# of bool_idx
print(a[bool_idx])

# We can do all of the above in a single concise statement:
print(a[a > 2])

[3 4 5 6]
[3 4 5 6]

For brevity we have left out a lot of details about numpy array indexing; if you want to know more you should read the documentation.

21.1.2.3 Datatypes
Every numpy array is a grid of elements of the same type. Numpy provides a large set of numeric datatypes that you can use to construct arrays. Numpy tries to guess a datatype when you create an array, but functions that construct arrays usually also include an optional argument to explicitly specify the datatype. Here is an example:

```python
[75]: x = np.array([1, 2])  # Let numpy choose the datatype
   y = np.array([1.0, 2.0])  # Let numpy choose the datatype
   z = np.array([1, 2], dtype=np.int64)  # Force a particular datatype

print(x.dtype, y.dtype, z.dtype)
```

int64 float64 int64

You can read all about numpy datatypes in the [documentation](#).

21.1.2.4 Array math
Basic mathematical functions operate elementwise on arrays, and are available both as operator overloads and as functions in the numpy module:

```python
[76]: x = np.array([[1,2],[3,4]], dtype=np.float64)
y = np.array([[5,6],[7,8]], dtype=np.float64)

# Elementwise sum; both produce the array
print(x + y)
print(np.add(x, y))
```
# Elementwise difference; both produce the array
print(x - y)
print(np.subtract(x, y))

\[
\begin{bmatrix}
-4. & -4. \\
-4. & -4. \\
-4. & -4. \\
-4. & -4. \\
\end{bmatrix}
\]

# Elementwise product; both produce the array
print(x * y)
print(np.multiply(x, y))

\[
\begin{bmatrix}
5. & 12. \\
21. & 32. \\
5. & 12. \\
21. & 32. \\
\end{bmatrix}
\]

# Elementwise division; both produce the array
# \[
# \begin{bmatrix}
# 0.2 & 0.33333333 \\
# 0.42857143 & 0.5 \\
# \end{bmatrix}
# \]
print(x / y)
print(np.divide(x, y))

\[
\begin{bmatrix}
0.2 & 0.33333333 \\
0.42857143 & 0.5 \\
0.2 & 0.33333333 \\
0.42857143 & 0.5 \\
\end{bmatrix}
\]

# Elementwise square root; produces the array
# \[
# \begin{bmatrix}
# 1. & 1.41421356 \\
# 1.73205081 & 2. \\
# \end{bmatrix}
# \]
print(np.sqrt(x))

\[
\begin{bmatrix}
1. & 1.41421356 \\
1.73205081 & 2. \\
\end{bmatrix}
\]

Note that unlike MATLAB, \* is elementwise multiplication, not matrix multiplication. We instead use the dot function to compute inner products of vectors, to multiply a vector by a matrix, and to multiply matrices. dot is available both as a function in the numpy module and as an instance method of array objects.
You can also use the @ operator which is equivalent to numpy's dot operator.
```python
print(np.sum(x, axis=0))  # Compute sum of each column; prints "[4 6]"
print(np.sum(x, axis=1))  # Compute sum of each row; prints "[3 7]"

10
[4 6]
[3 7]
```

You can find the full list of mathematical functions provided by numpy in the documentation.

Apart from computing mathematical functions using arrays, we frequently need to reshape or otherwise manipulate data in arrays. The simplest example of this type of operation is transposing a matrix; to transpose a matrix, simply use the `T` attribute of an array object:

```python
[86]:
    print(x)
    print("transpose

    ", x.T)

[[1 2]
 [3 4]]
transpose
[[1 3]
 [2 4]]
```

```python
[87]:
    v = np.array([[1,2,3]])
    print(v)
    print("transpose

    ", v.T)

[[1 2 3]]
transpose
[[1]
 [2]
 [3]]
```

21.1.2.5 Broadcasting

Broadcasting is a powerful mechanism that allows numpy to work with arrays of different shapes when performing arithmetic operations. Frequently we have a smaller array and a larger array, and we want to use the smaller array multiple times to perform some operation on the larger array.

For example, suppose that we want to add a constant vector to each row of a matrix. We could do it like this:

```python
[88]:
    # We will add the vector v to each row of the matrix x,
    # storing the result in the matrix y
    x = np.array([[1,2,3], [4,5,6], [7,8,9], [10, 11, 12]])
    v = np.array([1, 0, 1])
    y = np.empty_like(x)  # Create an empty matrix with the same shape as x
    
    # Add the vector v to each row of the matrix x with an explicit loop
    for i in range(4):
        y[i, :] = x[i, :] + v
```
print(y)

\[
\begin{bmatrix}
  2 & 2 & 4 \\
  5 & 5 & 7 \\
  8 & 8 & 10 \\
 11 & 11 & 13 \\
\end{bmatrix}
\]

This works; however when the matrix \( x \) is very large, computing an explicit loop in Python could be slow. Note that adding the vector \( v \) to each row of the matrix \( x \) is equivalent to forming a matrix \( vv \) by stacking multiple copies of \( v \) vertically, then performing elementwise summation of \( x \) and \( vv \). We could implement this approach like this:

```python
vv = np.tile(v, (4, 1))  # Stack 4 copies of v on top of each other
print(vv)
# Prints 
# [[1 0 1]
#  [1 0 1]
#  [1 0 1]
#  [1 0 1]]
```

\[
\begin{bmatrix}
  1 & 0 & 1 \\
  1 & 0 & 1 \\
  1 & 0 & 1 \\
  1 & 0 & 1 \\
\end{bmatrix}
\]

```python
y = x + vv  # Add x and vv elementwise
print(y)
\[
\begin{bmatrix}
  2 & 2 & 4 \\
  5 & 5 & 7 \\
  8 & 8 & 10 \\
 11 & 11 & 13 \\
\end{bmatrix}
\]
```

Numpy broadcasting allows us to perform this computation without actually creating multiple copies of \( v \). Consider this version, using broadcasting:

```python
import numpy as np

# We will add the vector v to each row of the matrix x,  
# storing the result in the matrix y
x = np.array([[[1,2,3],[4,5,6],[7,8,9],[10,11,12]]])
v = np.array([1,0,1])
y = x + v  # Add v to each row of x using broadcasting
print(y)
```

\[
\begin{bmatrix}
  2 & 2 & 4 \\
  5 & 5 & 7 \\
  8 & 8 & 10 \\
 11 & 11 & 13 \\
\end{bmatrix}
\]
The line $y = x + v$ works even though $x$ has shape $(4, 3)$ and $v$ has shape $(3,)$ due to broadcasting; this line works as if $v$ actually had shape $(4, 3)$, where each row was a copy of $v$, and the sum was performed elementwise.

Broadcasting two arrays together follows these rules:

1. If the arrays do not have the same rank, prepend the shape of the lower rank array with 1s until both shapes have the same length.
2. The two arrays are said to be compatible in a dimension if they have the same size in the dimension, or if one of the arrays has size 1 in that dimension.
3. The arrays can be broadcast together if they are compatible in all dimensions.
4. After broadcasting, each array behaves as if it had shape equal to the elementwise maximum of shapes of the two input arrays.
5. In any dimension where one array had size 1 and the other array had size greater than 1, the first array behaves as if it were copied along that dimension.

If this explanation does not make sense, try reading the explanation from the documentation or this explanation.

Functions that support broadcasting are known as universal functions. You can find the list of all universal functions in the documentation.

Here are some applications of broadcasting:

```python
# Compute outer product of vectors
v = np.array([1,2,3])  # v has shape (3,)
w = np.array([4,5])    # w has shape (2,)
# To compute an outer product, we first reshape v to be a column
# vector of shape (3, 1); we can then broadcast it against w to yield
# an output of shape (3, 2), which is the outer product of v and w:
print(np.reshape(v, (3, 1)) * w)
```

```
[[ 4  5]
 [ 8 10]
[12 15]]
```

```python
# Add a vector to each row of a matrix
x = np.array([[1,2,3], [4,5,6]])
# x has shape (2, 3) and v has shape (3,) so they broadcast to (2, 3),
# giving the following matrix:
print(x + v)
```

```
[[2 4 6]
 [5 7 9]]
```

```python
# Add a vector to each column of a matrix
# x has shape (2, 3) and w has shape (2,).
# If we transpose x then it has shape (3, 2) and can be broadcast
# against w to yield a result of shape (3, 2); transposing this result
```
# yields the final result of shape (2, 3) which is the matrix x with
# the vector w added to each column. Gives the following matrix:

```python
print((x.T + w).T)
```

```
[[ 5  6  7]
 [ 9 10 11]]
```

[95]: # Another solution is to reshape w to be a row vector of shape (2, 1);
# we can then broadcast it directly against x to produce the same
# output.

```python
print(x + np.reshape(w, (2, 1)))
```

```
[[ 5  6  7]
 [ 9 10 11]]
```

[96]: # Multiply a matrix by a constant:
# x has shape (2, 3). Numpy treats scalars as arrays of shape ();
# these can be broadcast together to shape (2, 3), producing the
# following array:

```python
print(x * 2)
```

```
[[ 2  4  6]
 [ 8 10 12]]
```

Broadcasting typically makes your code more concise and faster, so you should strive to use it
where possible.

This brief overview has touched on many of the important things that you need to know about
numpy, but is far from complete. Check out the [numpy reference](#) to find out much more about
numpy.

### 21.1.3 Matplotlib

Matplotlib is a plotting library. In this section give a brief introduction to the matplotlib.pyplot
module, which provides a plotting system similar to that of MATLAB.

[97]: `import matplotlib.pyplot as plt`

By running this special iPython command, we will be displaying plots inline:

[98]: `%matplotlib inline`

#### 21.1.3.1 Plotting

The most important function in matplotlib is plot, which allows you to plot
2D data. Here is a simple example:

[99]: # Compute the x and y coordinates for points on a sine curve
    x = np.arange(0, 3 * np.pi, 0.1)
    y = np.sin(x)
# Plot the points using matplotlib
plt.plot(x, y)

With just a little bit of extra work we can easily plot multiple lines at once, and add a title, legend, and axis labels:

```python
y_sin = np.sin(x)
y_cos = np.cos(x)

# Plot the points using matplotlib
plt.plot(x, y_sin)
plt.plot(x, y_cos)
plt.xlabel('x axis label')
plt.ylabel('y axis label')
plt.title('Sine and Cosine')
plt.legend(['Sine', 'Cosine'])
```

[100]: <matplotlib.legend.Legend at 0x7f78634f2860>
21.1.3.2 Subplots

You can plot different things in the same figure using the subplot function. Here is an example:

```python
# Compute the x and y coordinates for points on sine and cosine curves
x = np.arange(0, 3 * np.pi, 0.1)
y_sin = np.sin(x)
y_cos = np.cos(x)

# Set up a subplot grid that has height 2 and width 1, and set the first such subplot as active.
plt.subplot(2, 1, 1)
# Make the first plot
plt.plot(x, y_sin)
plt.title('Sine')

plt.subplot(2, 1, 2)
# Set the second subplot as active, and make the second plot.
plt.plot(x, y_cos)
plt.title('Cosine')

# Show the figure.
plt.show()
```
You can read much more about the subplot function in the documentation.

21.2 Torch tensor

The pytorch tensor class involves two data attribute (i.e., two data containers), including one analogous to a numpy multidimensional array with the elements of the tensor, and the other container with the gradients of an input function with respect to the tensor elements (Fig. 54). In addition, the tensor class involves the attribute backward method for the so-called backward propagation procedure applied for training neural networks that computes the gradients of the loss function with respect to the elements of the tensor. In addition, PyTorch has modules and functions summarized in Fig. 55.

As an example, we consider the following numpy multiarray

```
[102]: nt=np.ones((2,2))
```

which can be used to build a corresponding torch tensor, with associated gradients, as follows:

```
[103]: import torch
    import torch.nn as nn
    import torch.nn.functional as F
```

Figure 54: Schematic representation of a PyTorch tensor.
```python
import torch.optim as optim

r = torch.tensor(nt, requires_grad=True)

The resulting torch tensor can be combined with other torch tensors to define, for example, a function f as follows:

```python
p = torch.ones((2,2), requires_grad=True)
p2 = p+p
y=(r+2)+p2
z=y+y*3
f = z.mean()
print("r=", r)
print("p=", p)
print("f=", f)
```

```python
print('before backward: r.grad=', r.grad)
```

```python
r= tensor([[1., 1.],
           [1., 1.]], dtype=torch.float64, requires_grad=True)
p= tensor([[1., 1.],
           [1., 1.]], requires_grad=True)
f= tensor(75., dtype=torch.float64, grad_fn=<MeanBackward0>)
before backward: r.grad= None
```

Note that the torch tensor r does not have any gradients (i.e., before backward: r.grad= None) since so far we have not invoked the method backward for any function of r.

Now we can compute the gradient of f with respect to the elements of r by instantiating the backward attribute of f, as follows:

```python
f.backward()
```

```python
We can now check that the tensor r has the correct gradients of f with respect to the 4 elements of r, as follows:
```

```python
print('after backward: r.grad=', r.grad)
```

```python
after backward: r.grad= tensor([[7.5000, 7.5000],
                               [7.5000, 7.5000]], dtype=torch.float64)
```

We can also zero the gradients, as follows:

```python
r.grad.data.zero_()
print('after zero: r.grad=', r.grad)
```

```python
after zero: r.grad= tensor([[0., 0.],
                           [0., 0.]], dtype=torch.float64)
```

```
Figure 55: Summary of PyTorch modules and functions.