Continuous Variable Quantum Perceptron

F. Benatti[1,2], S. Mancini[3,4], S. Mangini[1]

[1]Dipartimento di Fisica, Università di Trieste, Strada Costiera 11, I-34151, Trieste, Italy

[2]Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Strada Costiera 11, I-34151, Trieste, Italy

[3]School of Science and Technology, University of Camerino I-62032 Camerino, Italy

[4]Istituto Nazionale di Fisica Nucleare, Sezione di Perugia, Via A. Pascoli, I-06123 Perugia, Italy

Abstract

We present a model of Continuous Variable Quantum Perceptron (CVQP) whose architecture implements a classical perceptron. The necessary non-linearity is obtained via measuring the output qubit and using the measurement outcome as input to an activation function. The latter is chosen to be the so-called ReLu activation function by virtue of its practical feasibility and the advantages it provides in learning tasks. The encoding of classical data into realistic finitely squeezed states and the use of superposed (entangled) input states for specific binary problems are discussed.

1 Introduction

Quantum Machine Learning brings together Machine Learning and Artificial Intelligence on one side and Quantum Information and Computation on the other one. Both sides have recently witnessed a series of breakthroughs which herald them as fundamental ingredients of future technologies. As a consequence a steadily growing amount of research has been focussing upon whether these two fields could benefit from each other. Many generalizations of quantum architectures for machine learning tasks and, vice versa, classical machine learning aided quantum computational architectures are currently being explored and tested [1, 2, 3, 4, 5].

In the following we focus on the study of a possible quantum implementation of a classical perceptron, the backbone of any learning algorithm, in a Continuous Variable (CV) quantum architecture [6], based on harmonic oscillator-like degrees of freedom, instead of *discrete*, spin-like variables. Some generalizations of perceptron models in the quantum regime have already been proposed [7, 8], but most of them are built on top of a discrete quantum system, made of *qubits*. Since a vast class of learning algorithms requires calculation of derivatives, which in turn needs continuous quantities to be evaluated, it is important to explore a possible quantum implementation of a perceptron that be continuous in nature.

Models of quantum perceptrons based on the pseudo eigen-projectors of position operators are easily mathematically constructible; however, for all practical purposes one has to investigate how to concretely implement these formal architectures by means of approximations based on square-integrable states with fairly well, but not perfect continuous localization properties. In doing this, one is then forced to consider the cost in energy that need be spent to counterbalance non-perfect state localization and the inaccuracies that it introduces. We shall investigate the relation between the energy cost and the probability of classification error in the case of the AND and XOR rules. In addition, it is worth underlining that the perceptron activation function

considered in this work is the ReLu (Rectified Linear Unit) which recently proved to be an optimal choice for learning taks [9, 10], and we shall introduce a measurement protocol devoted to its implementation. We shall also show that, in tackling the XOR and AND problems, no apparent advantage results from linear superpositions and entanglement possibly present in the input states of the quantum perceptron.

The structure of the work is as follows. In Section 2 all the necessary concepts about classical perceptrons are reviewd, in Section 3 the proposed model for a continuous quantum perceptron with the ReLu activation function is presented. In Section 4 realistic and feasible input states for the quantum perceptron are considered and in Section 5 the performance for the AND problem is studied while in Section 6 the XOR problem is addressed using linear superpositions. Finally, in Section 7, an outlook of possible further research directions is drawn.
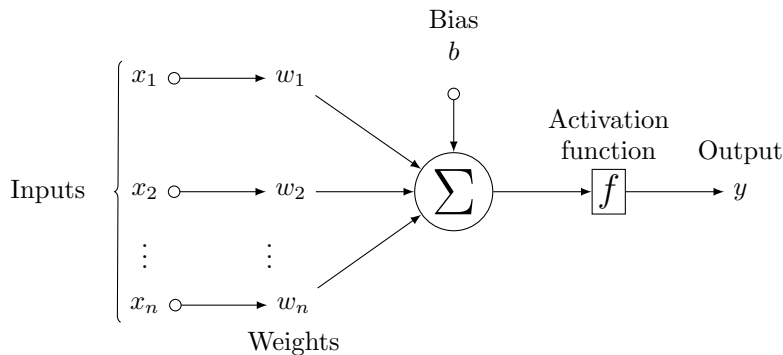
2   Classical Perceptron

Figure 1: Schematic representation of the mathematical model of a perceptron.

By perceptron it is meant a mathematical model that mimics the functioning of a natural perceptron. Given an input $\vec{x} = (x_1, \ldots, x_n)$ with $x_i \in \mathbb{R}$, a perceptron computes an affine transformation with real parameters $\vec{w} = (w_1, \ldots, w_n)$ and b, called *weights* and *bias*, respectively:

$$\vec{x} \longmapsto z := \vec{x} \cdot \vec{w} + b \ . \tag{1}$$

Subsequently, the perceptron evaluates on the output z an *activation function* $f : \mathbb{R} \to \mathbb{R}$, eventually yielding the final result $y := f(z)$.

There exist different possible activation functions, some being more computationally efficient and others more biologically inspired as the hyperbolic tangent $f(z) = (e^z - e^{-z})/(e^z + z^{-x})$, or the sigmoid function $f(z) = 1/(1 + e^{-z})$. More recently however, the nonlinear function known as Rectified Linear Unit, ReLu for short, has proved to be a very good candidate in learning procedures [9, 10]:

$$\text{ReLu}(z) := z^+ = \max(0, z) \qquad \forall z \in \mathbb{R} \ . \tag{2}$$

In order to illustrate how a classical perceptron can be used, consider the AND binary function on pairs $(x_1, x_2) \in \{-1, 1\}^2$ whose truth table is reported in the following Table 1. The AND function maps the four pairs (x_1, x_2) into 1 if and only if $x_1 = x_2 = 1$ otherwise the output is 0. One says that such a function is computed by the perceptron if the pairs $(x_1, x_2) \neq (1, 1)$

x_1	x_2	x_1 AND x_2
-1	-1	0
1	-1	0
-1	1	0
1	1	1

Table 1: Truth table of the AND problem.

are univocally associated with the output 0 and the pair $(1, 1)$ with the output 1. Choosing the weights $w_1 = w_2 = 1$ and the bias $b = -1$ one gets

$$z = x_1 + x_2 - 1 = \begin{cases} +1 & \text{if} & (x_1, x_2) & = & (1, 1) \\ -1 & \text{if} & (x_1, x_2) & = & (1, -1) \\ -1 & \text{if} & (x_1, x_2) & = & (-1, 1) \\ -3 & \text{if} & (x_1, x_2) & = & (-1, -1) \end{cases},$$

whence $\text{ReLu}(z) = 1$ only in the first case, whence the pair $(1, 1)$ is then univocally separated from the other three ones for which $\text{ReLu}(z) = 0$.

Unfortunately, the hope for a full classification of the whole of 16 binary logical functions by a classical perceptron is hindered by the exclusive-OR (XOR) logical function. Its truth table is reported in Table 2. Indeed, the XOR function outputs 1 only if only one of the input values is 1. Such a classification problem cannot be solved by a one-layer classical perceptron as it is not linearly separable; namely, the points $(1, -1)$, $(-1, 1)$ in the (x_1, x_2) plane cannot be univocally separated (by means of a line) from the points $(1, 1)$, $(-1, -1)$.

x_1	x_2	x_1 XOR x_2
-1	-1	0
1	-1	1
-1	1	1
1	1	0

Table 2: Truth table for the XOR function.

3 Continuous Variable Quantum Perceptron Model

The continuous quantum degrees of freedom we are going to use in the following are of photonic type described by annihilation and creation operators $\hat{a}_i$, $\hat{a}_i^\dagger$, $[\hat{a}_i, \hat{a}_i^\dagger] = 1$, or more conveniently by the quadrature position and momentum operators

$$\hat{x}_i = \frac{\hat{a}_i + \hat{a}_i^\dagger}{\sqrt{2}}, \quad \hat{p}_i = \frac{\hat{a}_i - \hat{a}_i^\dagger}{i\sqrt{2}}, \qquad [\hat{x}_i, \hat{p}_j] = i\,\delta_{ij} . \tag{3}$$

Within the formalism of continuous quantum computation by means of photons, the pseudo-eigenstates of $\hat{x}_i$, $\hat{x}_i|x_i\rangle = x_i|x_i\rangle$, play the same role as does the computational basis $\{|0\rangle, |1\rangle\}$ in discrete qubit systems; namely, any action on generic optical states can be described in terms of them.

A quantum circuit implementing the behaviour of a classical perceptron when acting on a quantum continuous optical input is summarized by the scheme in Fig. 2.
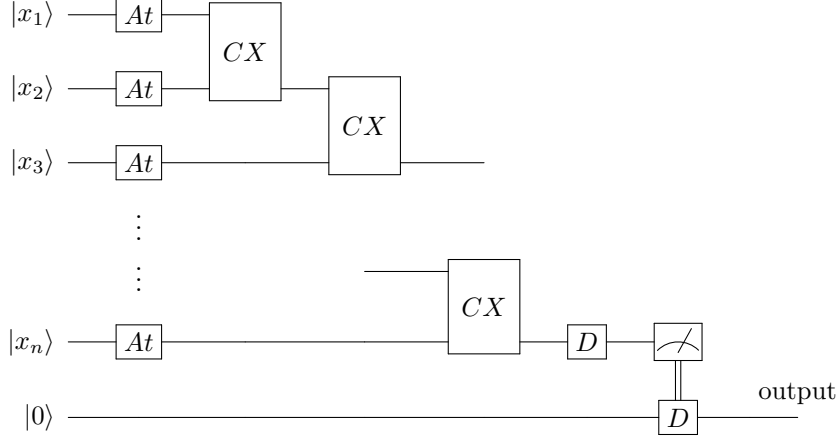
3

Figure 2: Scheme for a continuous valued quantum perceptron

A multimode input state consisting of a common eigenstate $|x_1, x_2, \ldots, x_n\rangle$ of the quadrature operators $\hat{x}_j$, $j = 1, 2, \ldots, n$, $\hat{x}_j |x_1, x_2, \ldots, x_n\rangle = x_j |x_1, x_2, \ldots, x_n\rangle$, undergoes three successive steps: in the first one each component of the input state is affected by a series of attenuators At which multiply each eigen-positions x_j by a real scaling factor η_j with $|\eta_j| \leq 1$. During the second step the attenuated eigen-positions are recursively added by means of suitably Controlled Addition CX gates; finally, in the last step, by means of a Displacement gate D, a bias is added to the last eigen-position. After these three steps, the activation function is implemented by acting on the last position eigen-state by performing a threshold ideal homodyne measurement.

The initial register consists of the n input state encoding the data to be processed and of an additional ancillary system, initialized in the pseudo-position eigenstate $|0\rangle$. Its role is to correctly implement the activation function and to propagate the result of the quantum perceptron. The visual representation of the later as a quantum circuit is shown in Figure 2 and can be schematically summarized as follows:

- **Attenuation**
  The attenuation process, ─$\boxed{At}$─, implements the multiplication of each eigen-position x_j by a corresponding weight $|\eta_j| \leq 1$:

$$|x_1, x_2, \ldots, x_n\rangle \to |\eta_1 x_1, \eta_2 x_2, \ldots, \eta_n x_n\rangle \ . \tag{4}$$

Such a transformation can be obtained by means of squeezing unitary gates. In fact, the single qumode squeezing operator

$$S(r) = \exp(i\,r\,(\hat{x}\hat{p} + \hat{p}\hat{x})) = \exp\bigl(r\,(a^2 - (a^\dagger)^2)\bigr) \ , \quad r \in \mathbb{R}^+ \ , \tag{5}$$

yields

$$S^\dagger(r)\,\hat{x}\,S(r) \ = \ \mathrm{e}^{-2r}\,\hat{x} \ , \qquad S^\dagger(r)\,\hat{p}\,S(r) = \mathrm{e}^{2r}\,\hat{p} \tag{6}$$

$$S^\dagger(r)\,\hat{a}\,S(r) \ = \ \hat{a}\,\cosh(2r) - \hat{a}^\dagger\,\sinh(2r) \ . \tag{7}$$

From the first expression in (6) it follows that

$$S(r)\,|x\rangle = \mathrm{e}^{-r}\,\bigl|\mathrm{e}^{-2r}x\bigr\rangle \ . \tag{8}$$

4

Setting $\eta = \mathrm{e}^{-2r}$, the following transformation is obtained:

$$|x_1, x_2, \ldots, x_N\rangle \to \sqrt{\eta_1 \eta_2 \cdots \eta_n} \, |\eta_1 x_1, \eta_2 x_2, \ldots, \eta_n x_n\rangle \ . \tag{9}$$

Thus, the strengths η_j of the attenuation processes implement the weights of the classical perceptron. Notice that the attenuation process performed by the squeezing operator $S(r)$ in (5) yields $0 \leq \eta = \mathrm{e}^{-2r} \leq 1$. In order to implement negative weights, one can use a simple Phase Shift gate $R(\phi) := \exp\left[i\phi\hat{a}^\dagger\hat{a}\right]$ with phase $\phi = \pi$. In this case the Attenuation gate would comprise both a Squeezing and a Rotation gate, $-\boxed{At}- = -\boxed{S}\boxed{R}-$, leading to the compound transformation:

$$|x\rangle \to \sqrt{\eta} \, |\eta x\rangle \to \sqrt{\eta} \, |-\eta x\rangle \ . \tag{10}$$

- **Controlled Addition**
  The controlled addition gate $-\boxed{}\atop{CX}$ processes two inputs by keeping the first one unaltered and adding it to the second one. The operator responsible for a such a process can be easily seen to be the following one:

$$CX := \exp\left[-\frac{i}{\hbar}\hat{x}_1 \otimes \hat{p}_2\right] \ , \qquad CX(|x_1\rangle \otimes |x_2\rangle) = |x_1\rangle \otimes |x_1 + x_2\rangle \ . \tag{11}$$

There are many possible implementations of such a gate, via two suitable beam-splittings that come before and after a squeezing gate [18] or via quantum-non demolition processes [11, 12, 13, 14]. The combined action on the attenuated state of the $n - 1$ CX gates of the circuit in Fig. 2 is then given by

$$|\eta_1 x_1, \eta_2 x_2, \ldots, \eta_n x_n\rangle \to |\eta_1 x_1, \eta_1 x_1 + \eta_2 x_2, \ldots, \eta_n x_n\rangle \to \ldots$$
$$\ldots \to \left|\eta_1 x_1, \eta_1 x_1 + \eta_2 x_2, \ldots, \sum_{i=1}^{n} \eta_i x_i\right\rangle \ . \tag{12}$$

Namely, they iteratively sum the position of one qumode system to the following one, eventually obtaining the sought after weighted sum as eigen-position of the last qumode.

- **Insertion of a bias**
  Using the n-th qumode as a reading mode, a bias $b \in \mathbb{R}$ can be added to its position by means of a Displacement operator $-\boxed{D}-$:

$$D(b) \ = \ \exp(-i \, b \, \hat{p}) = \exp\left(b\frac{\hat{a}^\dagger - \hat{a}}{\sqrt{2}}\right) \tag{13}$$

$$D^\dagger(b) \, \hat{x} \, D(b) \ = \ \hat{x} + b \ , \quad D^\dagger(b) \, \hat{a} \, D(b) = \hat{a} + \frac{b}{\sqrt{2}} \ . \tag{14}$$

Then, from the first equality in (14), one derives that the initial position eigenstate $|x_n\rangle$ of the last qumode is finally transformed into

$$D(b)|\sum_{i=1}^{n} \eta_i x_i\rangle = |\sum_{i=1}^{n} \eta_i x_i + b\rangle \tag{15}$$

exactly as demanded by the affine transformation (1).

- **Activation function**
Having encoded the affine transformation into the last qumode eigen-position, the last step consists in implementing the ReLu activation function (2) via a threshold measurement on it, $-\boxed{\measuredangle}-$: it mimics the non-linear behaviour of the classical perceptron. Such a task can be performed by an ideal homodyne measurement with POVM elements given by the pseudo-projector operators $P_y = |y\rangle \langle y|$ onto the pseudo-position eigenstates $|y\rangle$. Once performed on the last qumode when its state $\hat{\rho}$, such a measurement yields y with probability $P(y) = \langle y|\hat{\rho}|y\rangle$. Upon receiving y as measurement outcome, using the Displacement gate (13) one displaces the pseudo-position eigenstate $|0\rangle$ of the ancilla in the following way:

$$|0\rangle \to \begin{cases} |0\rangle & \text{if } y \leq 0 \\ |y\rangle & \text{if } y > 0 \end{cases} \tag{16}$$

Such a conditional action, $-\boxed{\measuredangle}$ $\quad$ implements the ReLu activation function (2) and even-

$\qquad\qquad \boxed{D}-$
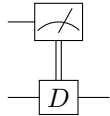
tually encodes the final result into state of the ancilla qumode. In the following, we shall consider more realistic scenarios where position pseudo-eigenstates are substituted by fairly well localized normalizable states. Consequently, the ideal pseudo-eigenstate $|y\rangle$ in (16) will also be substituted by the displaced vacuum state $D(y)|0\rangle$, where now $|0\rangle$ denotes the vacuum state, such that $\hat{a}^\dagger \hat{a}|0\rangle = 0$.

Remark 1 *Other possibilities for the activation function could be implemented substituting the threshold measurement protocol with suitable nonlinear and non gaussian gates, such as the Kerr gate. Unfortunately, due to the non interacting nature of photons and to the lack of sufficiently strong nonlinear materials with low absorption [15], such a gate is very difficult to implement physically. For this reason, measurement induced nonlinearities provide feasible and valuable alternatives. In addition, by means of the described conditional measurement protocol, it is possible to implement the ReLu activation function, which recently provided promising advantages for deep learning tasks. Notice that with this model the quantum implementation of a classical perceptron requires at least n qumodes for the signals to add up and one more ancillary qumode to implement the activation function. In addition, more modes may be necessary for the implementation of the gates used in the computation.*

4   Gaussian Input States

Continuous position eigenstates can not provide actual physical input states to the quantum perceptron because they correspond to Dirac deltas over the continuum of eigen-positions, and are thus not accessible as such in laboratory. However, they can be approximated by means of square integrable Gaussian states fairly well localized around the eigen-positions. Hence it is important to investigate the action of the quantum circuit (2) on Gaussian wave packets of the form

$$|\psi_j\rangle = \frac{1}{(\pi \Delta_j^2)^{1/4}} \int dq_j \, e^{-\frac{(q_j - x_j)^2}{2\Delta_j^2}} |q_j\rangle \ , \tag{17}$$

which are Gaussian weighted normalized superpositions of pseudo-eigenstates $|q_j\rangle$, of the position quadrature $\hat{x}_j$, where the classical datum x_j is encoded as center of the Gaussian weights with width Δ_j. Such states are obtained by acting on the vacuum state $|0\rangle$ first with a Squeezing

gate as in (5) and then with Displacement gate as in (13). Indeed, in position representation, $\langle q|0\rangle = (\exp(-q^2/2)/\sqrt[4]{\pi}$, so that the resulting Displaced-Squeezed vacuum state is [16]:

$$\langle q|\, D(b)S(r)|0\rangle = \frac{1}{(\pi e^{-2r})^{1/4}} \, \exp\left(-\frac{(q-b)^2}{2e^{-2r}}\right) \; , \tag{18}$$

which reduces to (17) when $e^{-r} = \Delta_j$ and $b = x_j$. Notice that, when r becomes large, these states approximate a Dirac delta around x_j. In the following we shall denote by $|x_j, \Delta_j\rangle$ the Displaced-Squeezed vacuum states $D(x_j)S(-\log \Delta_j)|0\rangle$.

It then follows that the input state to the quantum circuit in Fig. 2 is

$$|\Psi\rangle = \bigotimes_{j=1}^{n} |\psi_j\rangle = \prod_{j=1}^{n} \frac{1}{(\pi \Delta_j^2)^{1/4}} \int dq_j \, e^{-\frac{(q_j - x_j)^2}{2\Delta_j^2}} \bigotimes_{j=1}^{n} |q_j\rangle \; . \tag{19}$$

After applying the Attenuation and CX gates of the quantum circuit in FIg. 2, the outcome probability provided by the ideal homodyne detection reads (see Appendix A)

$$P(y, \vec{x}) = \frac{1}{\sqrt{\pi \sum_{j=1}^{N} \eta_j^2 \Delta_j^2}} \, e^{-\frac{(y - b - \sum_{j=1}^{N} \eta_j x_j)^2}{\sum_{j=1}^{N} \eta_j^2 \Delta_j^2}} \; , \qquad \vec{x} = (x_1, x_2, \ldots, x_N) \; . \tag{20}$$

It thus corresponds to a normalized Gaussian, centered around the result $\vec{w} \cdot \vec{x} + b$ of the affine transformation in (1). Depending on the actual outcome of the measurement process, the ancilla qumode will then be displaced by $y = f(z)$ and the final output read out.

However, differently from the case of the unnormalizable position-eigenstates, after homodyne measurements, states with not sharp eigen-positions yield all possible $y \in \mathbb{R}$ with a given probability distribution that in turn determines a probability of error associated to a wrong pattern classification by the quantum perceptron. Indeed, suppose that $a > 0$ is the correct answer for a given input state $|\Psi_a\rangle$ with associated outcome probability density $P(y)$. Actually, the ReLu activation function is able to discriminate between two kinds of states only, namely either the vacuum state $|0\rangle$, or Gaussian coherent states $D(a)|0\rangle$ with $a > 0$ (see the discussion just before Remark 1). Then,

$$P_{err}(\vec{x}) := \int_{-\infty}^{0} P(y, \vec{x}) \, dy \; , \tag{21}$$

represents the worst probability of miscalculation, that is the probability of obtaining a negative value from the homodyne measurement which in turn leads to an output ancilla qumode in the vacuum state, thus misclassifying the input state.

Furthermore, obtaining Displaced-Squeezed states has an energy cost that can be computed as the mean value of the number operator $\hat{n} = \hat{a}^\dagger \hat{a}$ with respect to such a state which is proportional to the electromagnetic energy content of an optical mode. Hence, as a figure of merit, the *energy cost* of obtaining a Displaced-Squeezed vacuum state from the vacuum state can be taken to be the difference of the mean values of the number operator computed with respect to those states, namely:

$$E(x_j, \Delta_j) := \langle x_j, \Delta_j|\hat{n}|x_j, \Delta_j\rangle - \langle 0|\hat{n}|0\rangle = \frac{|x_j|^2}{2} + \frac{(1 - \Delta_j^2)^2}{4\Delta_j^2} \; . \tag{22}$$

Such an an energy cost diverges when going to large squeezing parameters r, namely when the Squeezed Displaced vacuum approximates a Dirac delta around the displacement parameter. The argument of above then shows that a bound on the expendible energy unavoidably degrades the performances of the quantum perceptron introducing a non-vanishing probability of error. It is thus important to study the trade-off between the energy spent for a better spatial localization and the corresponding lowering of the error probability.

5   Quantum computation of the AND function

In order to study the performances of the quantum circuit in Fig. 2 as a model of quantum perceptron, we now probe it against the AND function introduced in Section 2 and relate the classification errors to the energy bound used in the encoding of the classical data into the input states.

In Fig. 3 it is depicted a quantum circuit that might be used to compute the AND function: in order to correctly classify the input pairs (x_1, x_2), it must ouput the vacuum state $|0\rangle$ if one or both the inputs x_i are negative or the displaced vacuum $|y\rangle$, with $y > 0$, if both are positive.
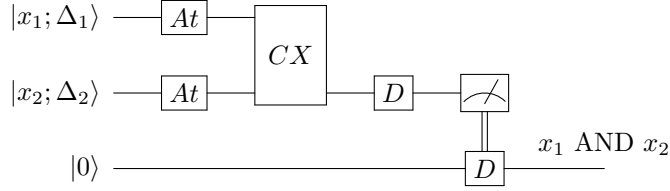

Figure 3: Scheme for implementing and AND function using Displaced-Squeezed vacuum states.

Could one work with position pseudo-eigenstates, then choosing attenuators $\eta_1 = \eta_2 = 1$ and bias $b = -1$ exactly as the classical weights and bias would correctly implements the AND function as with a classical perceptron. Instead, in an actual experimentally feasible context, the use of Displaced-Squeezed input states (18) and the corresponding probability distribution $P(y, x_1, x_2)$ in (20), yields the results reported in Table 3, where the probability of error $P_{err}(x_1, x_2)$ is defined as in (21) namely as the integral of the probability $P(y, x_1, x_2)$ over all possible results leading to misclassification.

Input		$P_{err}(x_1, x_2)$	
x_1	x_2	$r = 0$	$r = 1$
-1	-1	0.13%	$\sim 10^{-14}\%$
-1	+1	15.8%	0.3%
+1	-1	15.8%	0.3%
+1	+1	15.8%	0.3%

Table 3: Probability of misclassification, for each possible input, depending on the squeezing parameter $r = -\log \Delta$, where Δ is the width of the input Displaced-Squeezed vacuum states.

Evidently, a correct implementation of the AND function requires that the input states be squeezed in order to reduce the probability of errors. Already with squeezing factor $r = 1$, a good implementation is obtained. With this choice of squeezing parameter, the worst case scenario has a probability of error of just 0.3%. Thus, this quantum neuron could be actually used for a safe enough implementation of a single AND function. However, the evaluation of more complicated functions, by means of a network composed of multiple copies of such a quantum neuron, could be hindered by the accumulation of single neuron errors, though this effect could be controlled using an higher squeezing. The energy cost due to encoding the classical inputs into Displaced-Squeezed states, to which there contribute the mean values of the number operators $\hat{n} = \hat{a}^\dagger \hat{a}$, amounts to [17]:

$$E_{tot} = E(x_1, \Delta_1) + E(x_2, \Delta_2) = \frac{|x_1|^2 + |x_2|^2}{2} + \frac{(1 - \Delta_1^2)^2}{4\Delta_1^2} + \frac{(1 - \Delta_2^2)^2}{4\Delta_2^2} \ . \tag{23}$$

Remark 2 *Squeezing with $\Delta \ll 1$ is indeed very energy consuming, thus keeping $\Delta \simeq 1$ is preferable for an efficient classification. As a comparison, by encoding the classical input pair into Coherent states (without squeezing them) with displacements $|x_1| = |x_2| = 2$, and using a lower bias, $b = -2$, then, the greatest probability of error amounts to 2.27%, which is about one order of magnitude larger than by encoding through Displaced-Squeezed vacuum states, for as much the same energy cost. In fact, while for the Displaced-Squeezed encoding ($|x_1| = |x_2| = 1$ and $r = 1 \Leftrightarrow \Delta = 1/e$) the energy cost amounts to (23) $E_{tot} \sim 3.76$, using Displaced states with squeezing parameter $r = 0 \Leftrightarrow \Delta = 1$ and displacements $|x_1| = |x_2| = 2$ leads to an energy $E_{tot} = 4$.*
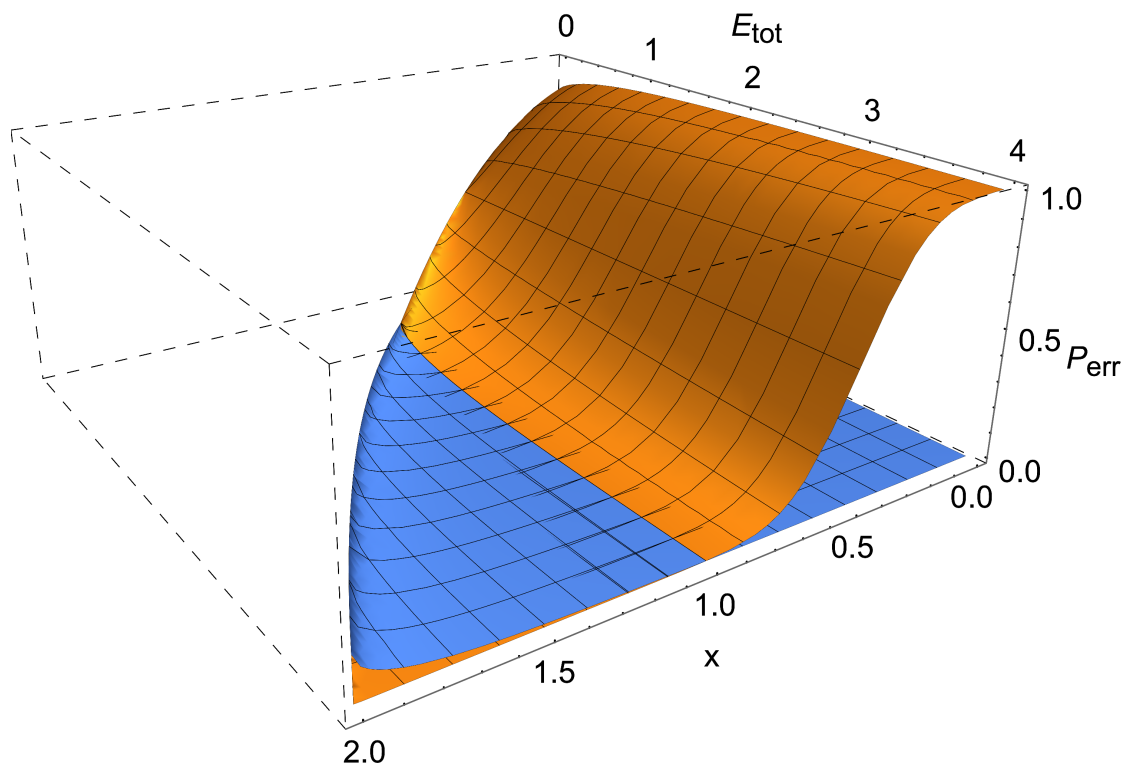

Figure 4: $P_{err}^+(x, E_{tot})$ (yellow surface) and $P_{err}^-(x, E_{tot})$ (blue surface) as functions of the displacement x and energy cost of the state E_{tot}

.

Further insights about the energy cost are shown in the above Fig. 4, where we assumed $|x_1| = |x_2| = |x|$ and allowed for the encoding of the inputs $(\pm 1, \pm 1)$ and $(\pm 1, \mp 1)$ by continuous variables (x, x) and $(x, -x)$ with $\Delta_1 = \Delta_2 = \Delta$. Then, one can invert (23) to have

$$\Delta^2 = \left(1 + E_{tot} - |x|^2\right) - \sqrt{\left(1 + E_{tot} - |x|^2\right)^2 - 1} \tag{24}$$

9

whence, setting $\eta_1 = \eta_2 = 1$ and $b = -1$, we can express (20) as

$$P^+(y, x, E_{tot}) = \frac{1}{\sqrt{2\,\pi\,\Delta^2}} \exp\left[-\frac{(y+1-2x)^2}{2\Delta^2}\right], \qquad x_1 = x_2 = x \qquad (25)$$

$$P^-(y, x, E_{tot}) = \frac{1}{\sqrt{2\,\pi\,\Delta^2}} \exp\left[-\frac{(y+1)^2}{2\Delta^2}\right], \qquad x_1 = -x_2 = x \qquad (26)$$

Leaving outside the case corresponding to the inputs $(-1, -1)$ which according to Table 3 shows low probability of error, according to (21) we will then have two probabilities of error, $P_{err}^+(x, E_{tot})$, corresponding to the case $(+1, +1)$, and $P_{err}^-(x, E_{tot})$, corresponding to the case with inputs with different signs. Figure 4 shows that until the energy is below one, it would be better to use the energy entirely for displacing the vacuum by $|x|$ with no energy consumption due to squeezing, $\Delta = 1$. On the other hand, once the displacement reaches $|x| = 1$, then the eventual extra energy can be used to squeeze the displaced vacuum so to achieve lower probability of error represented by $P_{err}^+(x, E_{tot})$.

Choosing different values for the parameters involved in the computation (including attenuators and bias) allows for different possible encodings, but in any case a compromise between accuracy and energy is needed. In all cases, the instance of the AND function indicates that the quantum circuit in Fig. 3 provides a realistic quantum implementation of a classical perceptron.

6 Quantum computation of the XOR function

As we have seen in Section 2, the exclusive-OR (XOR) logical function represents a limit for a classical perceptron. It is then worthwhile investigating whether the quantum circuit implementation of a classical perceptron can do better in such a case by exploiting quantum state superpositions and/or non-classical correlations. As we shall see, quantum superpositions lead to a probabilistic implementation of the XOR function. We start with analyzing position pseudo-eigenstates in order to get an idea of suitable inputs. Using the quantum circuit in Figure 3, with attenuators set to $\eta_1 = 1, \eta_2 = -1$, bias set to $b = -1$, and classical inputs (x_1, x_2) encoded into the superposition

$$|\psi_{in}\rangle = |x_1, x_2\rangle + |x_2, x_1\rangle \ , \qquad (27)$$

we get

$$|\psi_{in}\rangle \rightarrow |\psi_{out}\rangle = |x_1, x_1 - x_2 - 1\rangle + |x_2, x_2 - x_1 - 1\rangle \ .$$

For input values $x_1 = x_2 = x$, $|\psi_{out}\rangle = |x, -1\rangle + |x, -1\rangle = 2|x, -1\rangle$, whence the only possible result from the homodyne measurement on the second qumode is correctly negative. On the contrary, when $x_1 = -x_2 = x$, then $|\psi_{out}\rangle = |x, 2x - 1\rangle + |-x, -2x - 1\rangle$. In this case, for both $x_1 = 1$, $x_2 = -1$ and $x_1 = -1$, $x_2 = 1$, the second output mode is left in an equally weighted mixture of states $|-3\rangle$ and $|1\rangle$. Then the homodyne measurement statistics will result half of the times in the negative domain and half in the positive one, thus leading to a probabilistic and imperfect classification. Therefore, when an input with $x_1 = x_2$ is passed to the perceptron, it never misclassifies it, while when an input $x_1 = -x_2$ is presented, it has a 50% probability of being misclassified.

Let us make this more rigorous by resorting to normalizable states. Taking for simplicity equal widht for the gaussian states, that is $\Delta_1 = \Delta_2$, a feasible realization of the superposition

state (27) reads:

$$|\psi_{in}\rangle = \frac{1}{C}\Big(|x_1,\Delta\rangle \otimes |x_2,\Delta\rangle + |x_2,\Delta\rangle \otimes |x_1,\Delta\rangle\Big) \tag{28}$$

$$= \frac{1}{(C\pi\Delta^2)^{1/2}} \int dq_1 dq_2 \left(e^{-\frac{(q_1-x_1)^2}{2\Delta^2}} e^{-\frac{(q_2-x_2)^2}{2\Delta^2}} + e^{-\frac{(q_1-x_2)^2}{2\Delta^2}} e^{-\frac{(q_2-x_1)^2}{2\Delta^2}}\right) |q_1,q_2\rangle \ , \tag{29}$$

where C is a normalization constant that amounts to

$$C = 2\left(1 + e^{-\frac{(x_1-x_2)^2}{2\Delta^2}}\right) . \tag{30}$$

A discussion on the realizability of somehow similar states (just squeezed states are replaced by coherent ones) can be found in [19]. Applying the circuit with attenuators $\eta_1 = 1$ and $\eta_2 = -1$, the probability $P(y)$ of the ideal measurement becomes (see Appendix B):

$$P(y) = \frac{1}{C\sqrt{2\pi\Delta^2}}\left(e^{-\frac{(y-b-x_1+x_2)^2}{2\Delta^2}} + e^{-\frac{(y-b-x_2+x_1)^2}{2\Delta^2}} + 2e^{-\frac{(x_1-x_2)^2}{2\Delta^2}} e^{-\frac{(y-b)^2}{2\Delta^2}}\right). \tag{31}$$

Furthermore, setting $b = -1$ we can express (31) as

$$P^-(y,x,E_{tot}) = \frac{1}{\sqrt{2\pi\Delta^2}} e^{-\frac{(y+1)^2}{2\Delta^2}}, \qquad\qquad x_1 = x_2 = x \tag{32}$$

$$P^+(y,x,E_{tot}) = \frac{e^{-\frac{(y+1-2x)^2}{2\Delta^2}} + e^{-\frac{(y+1+2x)^2}{2\Delta^2}} + 2e^{-\frac{(y+1)^2+4x^2}{2\Delta^2}}}{(1+e^{-2x^2/\Delta^2})\sqrt{8\pi\Delta^2}} , \qquad\qquad x_1 = -x_2 = x \tag{33}$$

with Δ as in (24). As much as in Section 5, we will then have two probabilities of error $P_{err}^+(x,E_{tot})$ and $P_{err}^-(x,E_{tot})$. They are depicted in Figure 5. We observe that $P^+(y,x,E_{tot})$ is always greater than $1/2$, which means that the quantum perceptron fails the classification most of the time when $x_1 = -x_2$. In the limit of infinite energy, $r \to \infty$, and with $|x| = 1$, one correctly obtains $P_{err}^+ \to 1/2$, that is the same behaviour explained before with pseudo-position eigenstates.

Assuming the 4 possible inputs come with equal probability $1/4$, with increasing energy costs, the quantum circuit tends to answer correctly 75% of the times, which is the same as for a classical perceptron which can always correctly classify at least 3 of the 4 possible inputs. The main difference between the two perceptrons is that while the classical version always misclassifies at least one input, the quantum perceptron acts probabilistically on the inputs, misclassifying some of them at times.
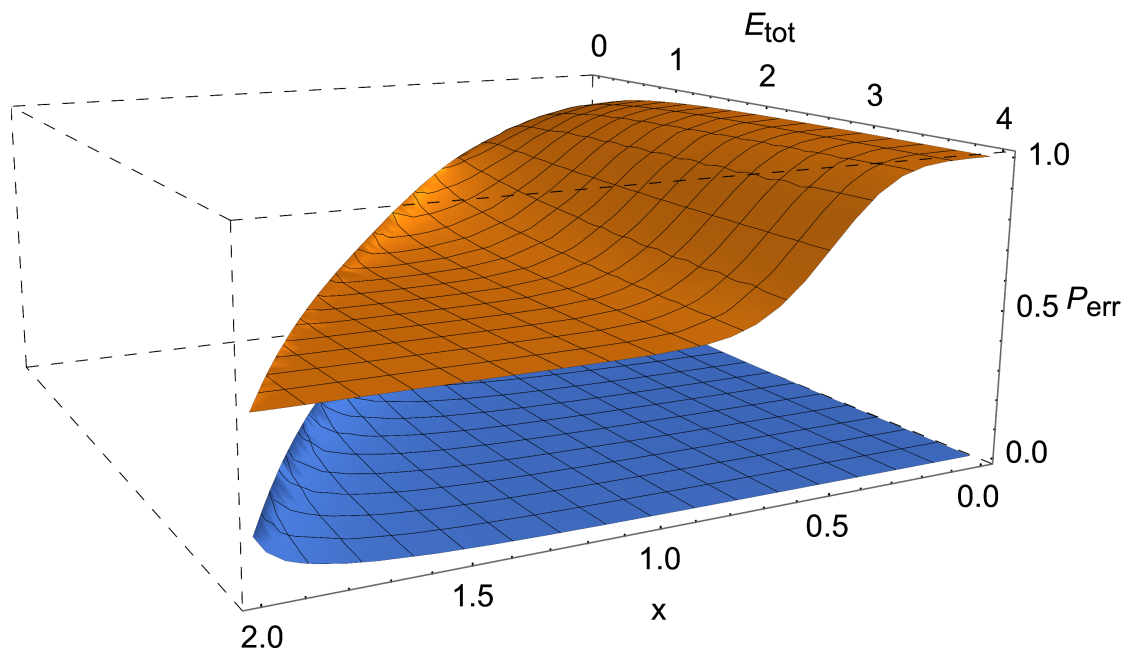
11

Figure 5: $P_{err}^+(x, E_{tot})$ (yellow surface) and $P_{err}^-(x, E_{tot})$ (blue surface) as functions of the displacement $|x|$, and energy cost of the state E_{tot}.

7   Conclusions

A continuous-variable quantum circuit model has been proposed that shows the capability of correctly implementing the desired classification power achievable by a classical perceptron. The use of realistic input states has been considered and shown to provide an efficient trade-off between a proper classification and the energy cost for those inputs.

The quantum nature of the circuit implementing the classical perceptron makes it possible the use of linear superpositions of quantum states; these latter, that have no classical counterpart, could in principle lead to classification advantages. However, our results indicate no quantum improvement when dealing with binary problems like the XOR and AND functions. Furthermore, for the AND function, there are preliminary evidences that also the use of an entangled encoding consisting of a two mode squeezed vacuum does not bring advantages both in terms of probability and of energy cost, with respect to using non-entangled two single mode squeezed states.

Actually, this should not come as a surprise, because the model proposed has been conceived as a direct copy of the classical model of a perceptron, and no further assumptions or ingredients have been used to achieve some kind of quantum advantage. In addition, the proposed quantum circuit falls within the hypothesis of the extended Gottesman Knill theorem [15], which makes the quantum circuit acting on Gaussian inputs efficiently simulatable on a classical computer. Interestingly, no advantages come even when going beyond Gaussian inputs through linear superpositions of Displaced-Squeezed vacuum states.

Therefore, it appears that quantum advantages are not easily achieved and the particular problems addressed in the manuscript are just an evidence of this fact. However, the possibility of processing more inputs at the same time through linear superpositions might require more sophisticated implementations of the perceptron non-linearity to make good use of them.

Nonetheless, it is important to underline that the proposed model works correctly both with position eigenstates and also with the realistic case of Gaussian states. It does then prove to be a good candidate for a quantum perceptron as a backbone of neural networks and machine learning techniques based on continuous quantum computation.

A   Measurement statistics

Given the total input state in (19), using the relation (9), each of its constituent Gaussian wavepacket (17) is attenuated into

$$
At(\eta_j)\,|\psi_j\rangle = \frac{1}{(\pi\Delta_j^2)^{1/4}}\int dq_j\; e^{-\frac{(q_j-x_j)^2}{2\Delta_j^2}}\,At(\eta_j)\,|q_j\rangle = \frac{\sqrt{\eta_j}}{(\pi\Delta_j^2)^{1/4}}\int dq_j\; e^{-\frac{(q_j-x_j)^2}{2\Delta_j^2}}\,|\eta_j q_j\rangle \;. \quad (34)
$$

As a consequence, upon acting with all attenuators and controlled addition gates on $|\Psi\rangle$ one obtains (for simplicity the case with bias $b=0$ is considered):

$$
\left|\tilde{\Psi}\right\rangle = \prod_{j=1}^{n}\left(\frac{\sqrt{\eta_j}}{(\pi\Delta_j^2)^{1/4}}\right)\int dq_1 dq_2\dots dq_n\; e^{-\frac{(q_1-x_1)^2}{2\Delta_1^2}}\,e^{-\frac{(q_2-x_2)^2}{2\Delta_2^2}}\dots\,e^{-\frac{(q_n-x_n)^2}{2\Delta_N^2}}
$$
$$
\times\left|\eta_1 q_1,\eta_1 q_1+\eta_2 q_2,\dots,\sum_{i=1}^{n}\eta_i q_i\right\rangle\;. \quad (35)
$$

Introducing the new variables:

$$
\tilde{q}_j = \sum_{k=1}^{j}\eta_k q_k\;,\qquad j=1,2,\dots,n\;,\quad q_{i+1}=\frac{\tilde{q}_{i+1}-\tilde{q}_i}{\eta_{i+1}}\;, \quad (36)
$$

with Jacobian $J=\prod_{j=1}^{n}\eta_j^{-1}$, the state in (35), can be recast as:

$$
\left|\tilde{\Psi}\right\rangle = \frac{1}{\prod_{j=1}^{n}(\pi\Delta_j^2\eta_j^2)^{1/4}}\int d\tilde{q}_1 d\tilde{q}_2\dots d\tilde{q}_n\; e^{-\frac{(\tilde{q}_1-\eta_1 x_1)^2}{2\eta_1^2\Delta_1^2}}\,e^{-\frac{(\tilde{q}_2-\tilde{q}_1-\eta_2 x_2)^2}{2\eta_2^2\Delta_2^2}}\dots\times
$$
$$
\times\, e^{-\frac{(\tilde{q}_n-\tilde{q}_{n-1}-\eta_n x_n)^2}{2\eta_n^2\Delta_n^2}}\,|\tilde{q}_1,\tilde{q}_2,\dots,\tilde{q}_n\rangle\;. \quad (37)
$$

Such a state is correctly normalized as required by the unitary action of the attenuators and CX gates, the corresponding orthogonal projector $P_{\tilde{\Psi}}=|\tilde{\Psi}\rangle\langle\tilde{\Psi}|$ being

$$
P_{\tilde{\Psi}} = \frac{1}{\prod_{j=1}^{N}(\pi\Delta_j^2\eta_j^2)^{1/2}}\int dq_1\dots dq_n\; dq_1'\dots dq_n'\, e^{-\frac{(q_1-\eta_1 x_1)^2}{2\eta_1^2\Delta_1^2}}\,e^{-\frac{(q_2-q_1-\eta_2 x_2)^2}{2\eta_2^2\Delta_2^2}}\dots\times
$$
$$
\times\, e^{-\frac{(q_n-q_{n-1}-\eta_n x_n)^2}{2\eta_n^2\Delta_n^2}}\,e^{-\frac{(q_1'-\eta_1 x_1)^2}{2\eta_1^2\Delta_1^2}}\,e^{-\frac{(q_2'-q_1'-\eta_2 x_2)^2}{2\eta_2^2\Delta_2^2}}\dots\,e^{-\frac{(q_n'-q_{n-1}'-\eta_n x_n)^2}{2\eta_n^2\Delta_n^2}}\times
$$
$$
\times\,|q_1,q_2,\dots,q_n\rangle\langle q_1',q_2',\dots,q_n'|\;. \quad (38)
$$

13

Since the affine transformation is encoded into the n-th qumode, the remaining ones can be traced out thus yielding the n-th qumode reduced density matrix

$$\hat{\rho}_n = \mathrm{Tr}_{1,\ldots,\,n-1}(P_{\tilde{\Psi}}) = \int dq_1'' \ldots dq_{n-1}'' \left\langle q_1'',\ldots,q_{n-1}'' \middle| \rho \middle| q_1'',\ldots,q_{n-1}'' \right\rangle$$

$$= \frac{1}{\prod_{j=1}^n (\pi \Delta_j^2 \eta_j^2)^{1/2}} \int dq_1 dq_2 \ldots dq_{n-1} \; e^{-\frac{(q_1-\eta_1 x_1)^2}{\eta_1^2 \Delta_1^2}} \cdots e^{-\frac{(q_{n-1}-q_{n-2}-\eta_{n-1}x_{n-1})^2}{\eta_{n-1}^2 \Delta_{n-1}^2}} \times \qquad (39)$$

$$\times \int dq_n dq_n' e^{-\frac{(q_n-q_{n-1}-\eta_n x_n)^2}{2\eta_n^2 \Delta_n^2}} e^{-\frac{(q_n'-q_{n-1}-\eta_n x_n)^2}{2\eta_n^2 \Delta_n^2}} \left| q_n \right\rangle \left\langle q_n' \right| \; .$$

One can check that $\mathrm{Tr}(\hat{\rho}_n) = 1$. Then, the remaining step to perform is the ideal homodyne measurement on the transformed n-th mode. Using the orthogonality relation between position eigenstates $\langle q|y\rangle = \delta(q-y)$, the probability $P(y)$ to obtain a given outcome $y \in \mathbb{R}$ is given by:

$$P(y) = \langle y | \hat{\rho}_n | y \rangle$$

$$= \frac{1}{\prod_{j=1}^n (\pi \Delta_j^2 \eta_j^2)^{1/2}} \int dq_1 dq_2 \ldots dq_{n-1} \; e^{-\frac{(q_1-\eta_1 x_1)^2}{\eta_1^2 \Delta_1^2}} \cdots e^{-\frac{(q_{n-1}-q_{n-2}-\eta_{n-1}x_{n-1})^2}{\eta_{n-1}^2 \Delta_{N-1}^2}} \times \qquad (40)$$

$$\times e^{-\frac{(y-q_{n-1}-\eta_n x_n)^2}{\eta_n^2 \Delta_n^2}} \; .$$

By first integrating with respect to dq_{n-1} and then applying iteratively the relation

$$\int_{-\infty}^{\infty} dx \; e^{-\frac{(y-x-a)^2}{b}} \; e^{-\frac{(x-z-c)^2}{d}} = \sqrt{\frac{\pi}{\frac{1}{b}+\frac{1}{d}}} e^{-\frac{(y-a-c-z)^2}{b+d}} \; , \qquad (41)$$

one gets

$$P(y) = \frac{1}{\sqrt{\pi \sum_{j=1}^N \eta_j^2 \Delta_j^2}} e^{-\frac{(y-\sum_{j=1}^N \eta_j x_j)^2}{\sum_{j=1}^N \eta_j^2 \Delta_j^2}} \; . \qquad (42)$$

Finally, inserting the bias, the measurement outcome probability distribution reads

$$P(y) = \frac{1}{\sqrt{\pi \sum_{j=1}^N \eta_j^2 \Delta_j^2}} e^{-\frac{(y-b-\sum_{j=1}^N \eta_j x_j)^2}{\sum_{j=1}^N \eta_j^2 \Delta_j^2}} \; . \qquad (43)$$

B Measurement statistics for the XOR binary function

The state

$$|\psi_{in}\rangle = \frac{1}{(C\pi\Delta_1\Delta_2)^{1/2}} \int dq_1 dq_2 \left(e^{-\frac{(q_1-x_1)^2}{2\Delta_1^2}} e^{-\frac{(q_2-x_2)^2}{2\Delta_2^2}} + e^{-\frac{(q_1-x_2)^2}{2\Delta_1^2}} e^{-\frac{(q_2-x_1)^2}{2\Delta_2^2}} \right) |q_1,q_2\rangle \; , \quad (44)$$

by application of the attenuators ($\eta_1 = 1$, $\eta_2 = -1$) and controlled addition, is transformed into

$$\frac{-1}{(C\pi\Delta_1\Delta_2)^{1/2}} \int dq_1 dq_2 \left(e^{-\frac{(q_1-x_1)^2}{2\Delta_1^2}} e^{-\frac{(q_2-x_2)^2}{2\Delta_2^2}} + e^{-\frac{(q_1-x_2)^2}{2\Delta_1^2}} e^{-\frac{(q_2-x_1)^2}{2\Delta_2^2}} \right) |q_1,q_1-q_2\rangle \; , \quad (45)$$

and changing integration variables $q_1 \to q_1$, $q_1 - q_2 \to q_2$, it is eventually obtained

$$\left| \tilde{\psi}_{in} \right\rangle = \frac{1}{(C\pi\Delta_1\Delta_2)^{1/2}} \int dq_1 dq_2 \left(e^{-\frac{(q_1-x_1)^2}{2\Delta_1^2}} e^{-\frac{(q_2-q_1+x_2)^2}{2\Delta_2^2}} + e^{-\frac{(q_1-x_2)^2}{2\Delta_1^2}} e^{-\frac{(q_2-q_1+x_1)^2}{2\Delta_2^2}} \right) |q_1, q_2\rangle \ .$$

(46)

Then, tracing out the first mode yields

$$\mathrm{Tr}_1 \left(\left| \tilde{\psi}_{in} \right\rangle \left\langle \tilde{\psi}_{in} \right| \right) = \frac{1}{C\pi\Delta_1\Delta_2} \int dq_1 dq_2 dq_2' \left(e^{-\frac{(q_1-x_1)^2}{2\Delta_1^2}} e^{-\frac{(q_2-q_1+x_2)^2}{2\Delta_2^2}} + e^{-\frac{(q_1-x_2)^2}{2\Delta_1^2}} e^{-\frac{(q_2-q_1+x_1)^2}{2\Delta_2^2}} \right)$$

$$\times \left(e^{-\frac{(q_1-x_1)^2}{2\Delta_1^2}} e^{-\frac{(q_2'-q_1+x_2)^2}{2\Delta_2^2}} + e^{-\frac{(q_1-x_2)^2}{2\Delta_1^2}} e^{-\frac{(q_2'-q_1+x_1)^2}{2\Delta_2^2}} \right) |q_2\rangle \langle q_2'| \ ,$$

(47)

so that, after evaluating the integration in dq_1, the probability $P(y) = \langle y | \mathrm{Tr}_1 \left(\left| \tilde{\psi}_{in} \right\rangle \left\langle \tilde{\psi}_{in} \right| \right) |y\rangle$, becomes (31)

$$P(y) = \frac{1}{C\sqrt{\pi(\Delta_1^2+\Delta_2^2)}} \left(e^{-\frac{(y-x_1+x_2)^2}{\Delta_1^2+\Delta_2^2}} + e^{-\frac{(y-x_2+x_1)^2}{\Delta_1^2+\Delta_2^2}} + 2e^{-\frac{(x_1-x_2)^2(\Delta_1^2+\Delta_2^2)}{4\Delta_1^2\Delta_2^2}} e^{-\frac{y^2}{\Delta_1^2+\Delta_2^2}} \right) \ . \quad (48)$$

References

[1] Vedran Dunjko and Hans J Briegel. Machine learning & artificial intelligence in the quantum domain: a review of recent progress *Rep. Prog. Phys. 81 074001*.

[2] Ciliberto Carlo, Herbster Mark, Ialongo Alessandro Davide, Pontil Massimiliano, Rocchetto Andrea, Severini Simone and Wossnig Leonard. Quantum machine learning: a classical perspective. *Proc. R. Soc. A*, **474**.

[3] Peter Wittek. Quantum Machine Learning: What Quantum Computing Means to Data Mining. *Elseiver Science*.

[4] Maria Schuld and Francesco Petruccione. Supervised Learning with Quantum Computers. *Springer*.

[5] Nathan Killoran, Thomas R Bromley, Juan Miguel Arrazzola, Maria Schuld, Nicolás Quesada and Seth Loyd. Continuous-variable quantum neural networks. *arXiv:1806.06871*.

[6] Seth Lloyd and Samuel L. Braunstein. Quantum Computation over Continuous Variables. *Phys. Rev. Lett.* **82**, 1784.

[7] Francesco Tacchino, Chiara Macchiavello, Dario Gerace and Daniele Bajoni. An artificial neuron implemented on an actual quantum processor. *npj Quantum Information*.

[8] Yudong Cao, Gian Giacomo Guerreschi and Alán Aspuru-Guzik. Quantum Neuron: an elementary building block for machine learning on quantum computers. *arXiv: 1711.11240*.

[9] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning.*

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. *The MIT Press.*

[11] K. Bencheikh, J. A. Levenson, Ph. Grangier and O. Grangier. Quantum Nondemolition Demonstration via Repeated Backaction Evading Measurements *Phys. Rev. Lett.* **75**, 3422.

[12] Robert Bruckmeier, Hauke Hansen and Stephan Schiller. Repeated Quantum Nondemolition Measurements of Continuous Optical Waves. *Phys. Rev. Lett.* **79**, 1463.

[13] R. L de Matos Filho and W. Vogel. Quantum Nondemolition Measurement of the Motional Energy of a Trapped Atom. *Phys. Rev. Lett.* **76**, 4520.

[14] S. F. Pereira, Z. Y. Ou and H. J. Kimble. Backaction evading measurements for quantum nondemolition detection and quantum optical tapping. *Phys. Rev. Lett.* **72**, 214.

[15] Stephen D. Barlett, Barry C. Sanders, Samuel L. Braunstein and Kae Nemoto. Efficient Classical Simulation of Continuous Variable Quantum Information Processes *Phys. Rev. Lett.* **88**, 097904.

[16] K. B. Møller, T. G. Jørgensen and J. P. Dahl. Displaced squeezed number states: Position space representation, inner product, and some applications. *Phys. Rev. Lett.* **54**, 5378.

[17] Pieter Kok and Brendon W. Lovett. Introduction to Optical Quantum Information Processing. *Cambridge University Press.*

[18] Nathan Killoran, Josh Izaac, Nicolás Quesada, Ville Bergholm, Matthew Amy and Christian Weedbrook. Strawberry Fields: A Software Platform for Photonic Quantum Computing. *Quantum* 3, 129 (2019).

[19] A Gilchrist, Kae Nemoto, W J Munro, T C Ralph, S Glancy, Samuel L Braunstein and G J Milburn. Schrödinger cats and their power for quantum information processing *J. Opt. B: Quantum Semiclass. Opt. 6, S828 (2004).*