problems Bi-directional LSTM is mainly applied for filing missing values in observations and showed goof model performance compared to LSTM but required more than twice the time for training.

C. Gated Recurrent Unit (GRU)

The GRU is one of the variants of the RNN network and an alternative version of LSTM proposed by Cho et al. in 2014. If LSTM networks have separate input and output gates then in GRU they are combined, thus reducing the number of parameters [7].

$$z_t = \delta(x_t U^z + h_{t-1} W^z) \quad (11)$$
$$r_t = \delta(x_t U^r + h_{t-1} W^r) \quad (12)$$
$$\widetilde{h_t} = tanh(x_t U^h + h_{t-1} W^h * r_t) \quad (13)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h_t} \quad (14)$$

When the reset gate $\mathrm{r_t}$ transfers important information from the previous hidden state and new input to a next state, then the update gate $\mathrm{z_t}$ is responsible for transferring information from the previous hidden state and reset gate to the next hidden. While LSTM stores its longer-term dependencies in the cell state and short-term memory in the hidden state, the GRU stores both in a single hidden state. GRU uses less memory and is faster than LSTM, however, LSTM is more accurate when using datasets with longer sequences [7]. The literature review on the comparison between LSTM and GRU advantages regarding time series forecasting showed that the there is no achieved consensus among researchers agreeing on distinct advantages one of them. For instance, if in [16] GRU forecasted better, then in [6] argued that GRU and LSTM accuracy are at same level. However, taking into account the difference in the architecture of these networks, it is still difficult to select for time series forecasting. Most reported experiments mainly refer to problems other than time series forecasting.

D. Quantum-inspired LSTM

The theory of quantum computations is an important and rapidly developing area of deep learning theory in our days. One of the most widely used quantum algorithms that incorporated into classical deep learning methods in classical computers are variation quantum algorithms (VQA) [13]. These algorithms showed good results on optimization and classification problems [14]. In the quantum field there are two concepts,superposition and entanglement, which depend on encoding methods. In order to map classical data into quantum states encoding procedures should be applied. One of the most widely used is amplitude encoding, where the input time series? vector normalized in the range $[0\ ;\frac{\pi}{2}]$ and applied to Bloch sphere aplitudes states ilustrated in Fig.1. The north pole and the south are the basis states $|0\rangle$ and $|1\rangle$ and by rotation in this unit presents possible state combinations.

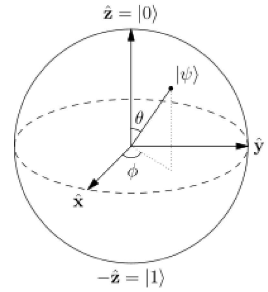

Figure 3. Geometrical representation of Bloch sphere

The entalgment can be achived by appying controled gates for each qubit states. Amplitude of each single qubit after entanglement through gates can be optimized by classical optimization methods used in neural networks training. There are limited studies on time serires forecasting with quantum enhanced LSTM [15]. However, more research is required on the problem of learning time series in the quantum field.

E. Related work: Time series forecasting

Time series forecasting is one of the main research tasks across many disciplines. The reason behind this interest relates to future values x_{t+1} that they can produce based on historical observations (Eq. 13).

$$x_{t+1} = f(x_t, x_{t-1}, x_{t-2}, \ldots., x_{t-n}) + error \quad (15)$$

Variations that are not explained by the observations are included in the error term. In the relevant literature, there have been numerous successful applications in different fields, such as finance and medicine. Among the most popular linear methods in time-series forecasting are auto-regressive integrated moving average, partial least squares, lasso regression. Univariate time series forecasting most often uses methods that are auto-regressive models that detect the previous signals. This models advanced forms are ARMA and ARIMA, which combine past signal with moving average and differencing technics. The specification of the level of differencing allows capturing complex patterns. However, it has also been found that many real time series seem to follow non-linear behavior and the liner approach is insufficient to represent their dynamics. Thus in the most relevant literature have been presented a wide range of models based on neural networks [2, 4, 5]. Neural networks are efficient in forecasting time series, especially with LSTM and GRU variants of recurrent networks. But still there are a gaps in the literature and with the most prominent gap raises questions on the capturing trend, seasonality and cyclical components. The trend component of time series is one of the important researching aspects in forecasting and there have been mixed opinions among researchers. Real world time series encapsulate several components translating seasonal, cycles or trend features of observed processes. A trend component can be long or short term, in classical forecasting methods this component is removed by applying detrending methods. There are several methods such as moving average filters, differencing at level and order. The literature review on effects of detrending application in forecasting shows that there are limited experimental studies with solid theoretical