

Esquemas de Codificação Espaço-Temporal Baseados em Caos com Decodificação por Aprendizagem Profunda

Carlos E. C. Souza, Rafael Campello, Cecilio Pimentel e Daniel P. B. Chaves

Resumo— Neste trabalho é proposto um esquema de codificação espaço-temporal baseado em atratores caóticos tridimensionais. Codificadores de estados finitos são utilizados para mapear sequências de informação nos sinais caóticos que são transmitidos por múltiplas antenas. É proposta uma arquitetura de rede neural para decodificar o sinal transmitido por aprendizagem profunda. O desempenho do esquema proposto é avaliado em relação à taxa de erro de blocos por aprendizagem profunda e pelo algoritmo de Viterbi. É demonstrado que a rede proposta é capaz de aprender a estrutura do código espaço-temporal baseado em caos.

Palavras-Chave— Comunicação caótica, codificação espaço-temporal, dinâmica simbólica, aprendizagem profunda.

Abstract— In this work we propose a space-time trellis code scheme based on three-dimensional chaotic attractors. Finite-state encoders are employed to map binary sequences into chaotic signals which are transmitted by multiple antennas. We propose a neural network architecture to decode the transmitted signal by machine learning. The frame error rate of the proposed scheme is analyzed by deep learning and the Viterbi algorithm. We show that the proposed neural network is capable of learning the chaos-based space-time code.

Keywords— Chaos communication, space-time trellis codes, symbolic dynamics, deep learning.

I. INTRODUÇÃO

Sistemas de comunicação baseados em caos vêm sendo explorados na literatura para o projeto de sistemas de comunicação sem fio [1], [2]. Em [3], [4] mostra-se que sinais caóticos gerados por atratores tridimensionais transmitidos em canais multipercurso conservam a topologia do atrator. O expoente de Lyapunov também é conservado em canais multipercurso, portanto não há perda de informação topológica decorrente do efeito do canal [3]. Estas propriedades evidenciam que sistemas dinâmicos caóticos são promissores candidatos para o projeto de sistemas de comunicação sem fio.

Esquemas de codificação espaço-temporal em treliça (STTC, *space-time trellis code*) foram propostos inicialmente em [5] com o objetivo de aumentar a taxa de transmissão e o desempenho de sistemas de comunicação em canais sem fio. Nestes esquemas, a diversidade espacial obtida pela utilização de múltiplas antenas resulta em melhor desempenho quando comparado com sistemas com uma única antena [5]–[7].

Em [8] é proposto um sistema de modulação codificada em treliça baseado em atratores caóticos tridimensionais e é

feito um estudo de caso para o atrator Sprott D. A evolução dinâmica das trajetórias no atrator é representada pela dinâmica simbólica gerada por uma partição rotulada da seção de Poincaré. A dinâmica simbólica das trajetórias é representada por grafos construídos a partir de um conjunto de palavras proibidas. Um codificador de estados finitos (FSE, *finite-state encoder*) é construído para mapear sequências de informação nas sequências restritas geradas pela dinâmica simbólica das trajetórias caóticas.

Este trabalho tem duas contribuições principais. A primeira é a incorporação da dinâmica caótica, representada pela dinâmica simbólica das trajetórias caóticas, em sistemas STTC baseados em caos (CB-STTC, *chaos-based space-time trellis code*). Para este fim, as restrições provenientes das propriedades dinâmicas e topológicas do fluxo caótico são exploradas no projeto de um sistema de comunicação com múltiplas antenas. Em particular, a metodologia desenvolvida em [8] é estendida para projetar um CB-STTC baseado no atrator caótico Sprott D. A segunda contribuição é a investigação de técnicas de aprendizagem profunda para a decodificação do CB-STTC. Propomos uma arquitetura de rede neural que combina redes convolucionais e redes recorrentes para investigar a capacidade de aprender a estrutura do CB-STTC proposto. A estratégia de treinamento da rede é discutida e são realizadas comparações com esquemas clássicos de decodificação. A decodificação de códigos de bloco lineares via técnicas de aprendizagem profunda tem sido considerada na literatura [9], [10].

Este trabalho está dividido em seis seções. Na Seção II apresenta-se o atrator Sprott D, sua dinâmica simbólica e o FSE associado. O sistema de comunicação e o CB-STTC proposto são detalhados na Seção III. A arquitetura da rede neural é detalhada na Seção IV. Na Seção V o desempenho do sistema proposto é analisado. Finalmente, na Seção VI são feitas as considerações finais.

II. REPRESENTAÇÃO SIMBÓLICA DO ATRATOR SPROTT D

O sistema Sprott D é definido por [11]

$$\begin{cases} \dot{x}(t) &= -y(t) \\ \dot{y}(t) &= x(t) + z(t) \\ \dot{z}(t) &= x(t)z(t) + ay^2(t) \end{cases} \quad (1)$$

em que $a = 2,3$ é o parâmetro de controle. A dinâmica das trajetórias caóticas no atrator Sprott D pode ser analisada de forma simples utilizando-se uma partição rotulada de uma seção de Poincaré. Neste caso, esta é um plano perpendicular ao plano xy , localizada em $y = 0$. A Fig. 1 mostra o atrator

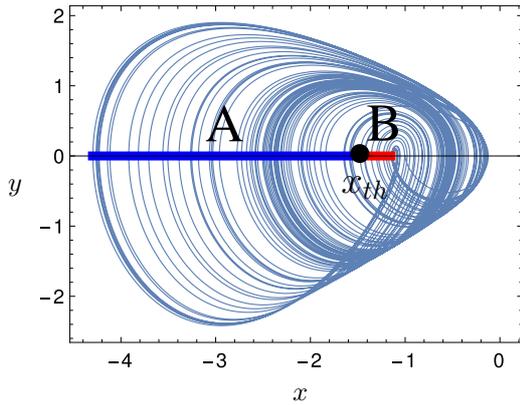


Fig. 1. Atrator Sprott D e sua seção de Poincaré em projeção no plano xy . Também é mostrado o particionamento binário considerado, utilizado para codificar as trajetórias caóticas como sequências binárias, em que cada símbolo define se o cruzamento é à esquerda ou à direita do limiar x_{th} .

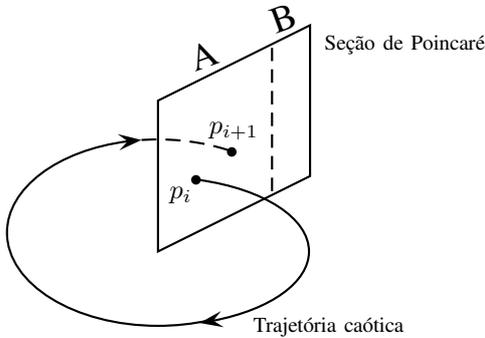


Fig. 2. Cruzamentos entre as trajetórias caóticas e a seção de Poincaré gerando a sequência p .

Sprott D e a seção de Poincaré considerada em projeção no plano xy . Também é mostrado um particionamento binário sobre o alfabeto $\mathcal{A}_2 = \{A, B\}$. O limiar x_{th} é definido como o mínimo do mapa de Poincaré [12].

O mapa de Poincaré $P : X \rightarrow X$ para um atrator caótico tridimensional X é definido por $p_{i+1} = P(p_i)$, em que $(x_i, y_i, z_i) = p_i$ são as coordenadas do i -ésimo cruzamento entre uma trajetória em A e a seção de Poincaré. Seja $p = (p_0, p_1, p_2 \dots, p_{N-1})$ uma sequência de N cruzamentos e seja $s_N = s_0 s_1 s_2 \dots s_{N-1}$, $s_i \in \mathcal{A}_2$ uma sequência simbólica, em que s_i é o rótulo da região da seção de Poincaré em p_i . A Fig. 2 ilustra a dinâmica de cruzamentos na seção de Poincaré do atrator Sprott D. Em seguida, será considerada esta representação para o sistema de comunicação.

A existência de restrições na ocorrência de símbolos adjacentes da sequência s_N para o atrator Sprott D é investigada em [8]. Estas restrições são especificadas por um conjunto de sequências proibidas \mathcal{F} . A partir deste conjunto, projeta-se o FSE de taxa 1/2 ilustrado na Fig. 3 [8]. As sequências geradas percorrendo-se os caminhos no FSE correspondem às sequências geradas pelos sucessivos cruzamentos entre as trajetórias caóticas e a seção de Poincaré, satisfazendo as restrições na dinâmica do atrator. O FSE mapeia sequências

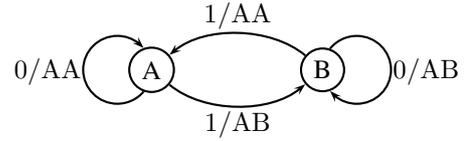


Fig. 3. Codificador de estados finitos para o atrator Sprott D empregado para mapear sequências binárias sem restrição nas sequências simbólicas restritas definidas por \mathcal{F} .

binárias de informação em sequências simbólicas restritas geradas pelo mapeamento de um bit na sequência $s_1 s_2$. Isto equivale a associar cada bit a dois cruzamentos na seção de Poincaré ilustrada na Fig. 1. Para transmitir informação usando o FSE é necessário definir as formas de onda caóticas associadas às transições de estados no FSE.

A. Formas de Onda Caóticas

As formas de onda caóticas são os segmentos de trajetórias conectando as regiões correspondentes da seção de Poincaré para cada transição possível no FSE. Conforme detalhado em [8], o intervalo de tempo entre os cruzamentos não é uniforme devido à aperiodicidade do comportamento caótico. Portanto as formas de ondas caóticas são parametrizadas em termos de uma variável angular θ que varia 2π radianos em cruzamentos consecutivos na seção de Poincaré. Por exemplo, as formas de onda caóticas associadas à transição do estado A para o estado B no FSE são os segmentos de trajetória $x(\theta), y(\theta)$ ou $z(\theta)$ conectando as regiões A a A e em seguida conectando as regiões A a B, o qual denotamos por $w_{AB}^A(\theta)$, $w = x, y, z$, com período angular 4π . O índice superior em $w_{\beta\gamma}^\alpha(\theta)$ indica a região atual da trajetória na seção de Poincaré e os índices inferiores indicam as duas próximas regiões visitadas.

As formas de onda caóticas geradas para uma determinada transição são ligeiramente diferentes a cada transição devido à aperiodicidade do comportamento caótico. Portanto, associamos uma curva média para cada transição possível como a média entre várias curvas geradas com condição inicial na mesma região. Denominamos esta curva média de curva de transição típica e esta é utilizada no demodulador conforme descrito na Seção III.

III. O SISTEMA DE COMUNICAÇÃO

Vamos considerar o FSE na Fig. 3 para projetar um sistema de comunicação de dois estados com N_T antenas transmissoras e N_R antenas receptoras. O canal possui desvanecimento uniforme quasi-estático com ruído aditivo Gaussiano branco. Os coeficientes de desvanecimento h_{ij} da antena transmissora i para a antena receptora j são variáveis aleatórias Gaussianas complexas, constantes durante cada bloco de comprimento l e independentes para cada percurso. Estes coeficientes variam independentemente em blocos consecutivos. A cada intervalo de sinalização, o sinal caótico $w_{\beta\gamma}^\alpha(\theta)$, $0 \leq \theta \leq 4\pi$, é mapeado no sinal transmitido $s(t)$. O sinal recebido na j -ésima antena

receptora é dado pela superposição linear

$$r_j(t) = \sum_{i=1}^{N_T} h_{ij} s_i(t) + n(t), \quad j = 1, 2, \dots, N_R \quad (2)$$

em que $n(t)$ é um ruído aditivo Gaussiano branco com densidade espectral de potência $N_0/2$ e $|h(t)|$ possui densidade de probabilidade Rayleigh com segundo momento unitário. Admitindo que o decodificador conhece perfeitamente os coeficientes h_{ij} , a decodificação de máxima verossimilhança é baseada na minimização de

$$\sum_{j=1}^{N_R} \int (r_j(t) - \sum_{i=1}^{N_T} h_{ij} \bar{s}_i(t))^2 dt \quad (3)$$

em que $\bar{s}(t)$ denota uma curva típica de transição. A relação sinal ruído (SNR, *signal-to-noise ratio*) por antenna receptora é definida por $\sum_{i=1}^{N_T} E_i/N_0$, em que E_i é a energia média do sinal transmitido na i -ésima antenna.

Para o canal com desvanecimento quasi-estático, o projeto do CB-STTC segue os critérios apresentados em [5]–[7]. Seja $\mathbf{c} = (s_1^1(t) \dots s_1^{N_T}(t) s_2^1(t) \dots s_2^{N_T}(t) \dots s_\ell^1(t) \dots s_\ell^{N_T}(t))$ uma palavra-código em um bloco, em que $s_k^i(t)$ é o sinal caótico transmitido no k -ésimo intervalo de sinalização pela antenna i . Considere outra palavra-código $\mathbf{e} = (e_1^1(t) \dots e_1^{N_T}(t) e_2^1(t) \dots e_2^{N_T}(t) \dots e_\ell^1(t) \dots e_\ell^{N_T}(t))$. A matriz de distâncias entre o par de palavras-código (\mathbf{c}, \mathbf{e}) definida em [5] é uma matriz quadrada de dimensão $N_T \times N_T$ e é denotada por $\mathbf{A}(\mathbf{c}, \mathbf{e})$. O projeto de um sistema STTC deve seguir dois critérios apresentados em [5]. O primeiro é maximizar o posto mínimo de $\mathbf{A}(\mathbf{c}, \mathbf{e})$ para todos os pares possíveis de palavras código. Este critério está associado ao ganho de diversidade do sistema. O segundo critério é maximizar o determinante mínimo de $\mathbf{A}(\mathbf{c}, \mathbf{e})$ para todos os pares distintos de palavras código com o posto mínimo, e este critério está associado ao ganho de codificação. O ganho de diversidade do STTC é dado por rN_R , em que r é o posto de $\mathbf{A}(\mathbf{c}, \mathbf{e})$. Consequentemente, o ganho de diversidade possui valor máximo $N_T N_R$ e o código deve ser projetado de forma que o posto mínimo de $\mathbf{A}(\mathbf{c}, \mathbf{e})$ seja N_T .

Neste trabalho consideramos $N_T = 2$ com taxa 2 bits/s/Hz, logo há 4 ramos divergentes de cada estado no FSE do CB-STTC e cada ramo é rotulado com dois bits de informação. O codificador de dois estados é projetado de forma que as matrizes $\mathbf{A}(\mathbf{c}, \mathbf{e})$ associadas a pares de palavras-código possuam posto mínimo $r = N_T = 2$ para quaisquer pares de palavras-código. Portanto, o ganho de diversidade do código é igual a $2N_R$.

As formas de ondas caóticas transmitidas em cada antenna seguem o esquema da Fig. 3 em que variáveis distintas são transmitidas em cada antenna. Desta forma a matriz $\mathbf{A}(\mathbf{e}, \mathbf{c})$ tem posto mínimo 2. Com o intuito de aumentar o determinante mínimo desta matriz, buscas computacionais indicam que as variáveis x e z são transmitidas em cada antenna e inverte-se um dos sinais. Neste caso, a inversão é associada ao bit de informação 1. Portanto, a antenna 1 transmite o sinal $\pm x(t)$ e a antenna 2 transmite $\pm z(t)$. O FSE para o CB-STTC proposto é detalhado na Tabela I. A Fig. 4 ilustra uma realização das formas de onda $x(\theta)$ e $z(\theta)$ para quatro transições no FSE.

TABELA I
FSE DE DOIS ESTADOS PARA O CB-STTC.

Transição no FSE	Par transmitido	Antena 1	Antena 2
$A \rightarrow A$	00	$x_{AA}^A(t)$	$z_{AA}^A(t)$
$A \rightarrow A$	11	$-x_{AA}^A(t)$	$-z_{AA}^A(t)$
$A \rightarrow B$	01	$x_{AB}^A(t)$	$-z_{AB}^A(t)$
$A \rightarrow B$	10	$-x_{AB}^A(t)$	$z_{AB}^A(t)$
$B \rightarrow A$	01	$x_{AA}^B(t)$	$-z_{AA}^B(t)$
$B \rightarrow A$	10	$-x_{AA}^B(t)$	$z_{AA}^B(t)$
$B \rightarrow B$	00	$x_{AB}^B(t)$	$z_{AB}^B(t)$
$B \rightarrow B$	11	$-x_{AB}^B(t)$	$-z_{AB}^B(t)$

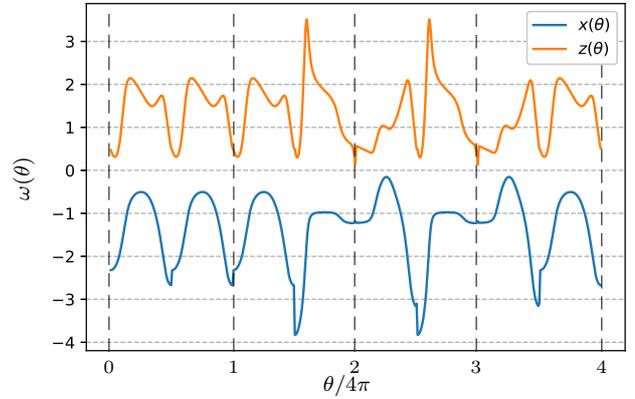


Fig. 4. Trajetórias caóticas $x(\theta)$ e $z(\theta)$ para quatro transições do FSE detalhado na Tabela I.

IV. APRENDIZAGEM DE MÁQUINA

Nesta seção é detalhada a arquitetura da rede neural empregada na decodificação do CB-STTC proposto. É utilizada a biblioteca Tensorflow versão 2.1 juntamente com sua API Keras para modelagem e treinamento da rede neural. O serviço *Google Colaboratory* foi utilizado para treinamento acelerado da rede neural [13]. O desenvolvimento da arquitetura e treinamento da rede neural consiste em obter um modelo capaz de realizar inferências a respeito do esquema proposto sem a necessidade de desenvolver algoritmos específicos para a situação proposta.

Seja $\mathbf{y} = (y_1, y_2, \dots, y_\ell)$ um vetor de dimensão ℓ sobre o alfabeto $\{0, 1, 2, 3\}$, em que cada elemento de \mathbf{y} constitui uma classe representando um par de bits de informação $\{00, 01, 10, 11\}$. A partir de \mathbf{y} , calcula-se a palavra código correspondente que são transmitidas pelo canal e calcula-se $r_j(t)$ conforme (2). Em seguida, calcula-se as métricas definidas em (3) com todas as possíveis curvas típicas de transição do sistema que são transmitidas pelas duas antenas simultaneamente (em cada intervalo de sinalização existem oito pares de curvas possíveis, como indicado nas linhas da Tabela I). Em seguida, é construída a matriz \mathbf{X} de dimensão $8 \times \ell$ com as possíveis métricas em cada bloco. A matriz \mathbf{X} é

obtida para diferentes valores de SNR e o par (\mathbf{X}, \mathbf{y}) simboliza o sistema de comunicação. Desta forma, todas as informações apresentadas à rede neural sobre o sistema de comunicação são definidas por meio de seus dados de entrada. Espera-se que por meio exclusivamente destes dados o modelo consiga aprender uma forma de mapear \mathbf{X} em \mathbf{y} que generalize para casos não previamente apresentados à rede.

A. Arquitetura da Rede

A arquitetura da rede utiliza três tipos de camadas: convolucionais, recorrentes e totalmente conectadas [14]. O uso simultâneo dessas camadas se beneficia da extração de atributos proporcionada pelas camadas convolucionais e preserva a dinâmica sequencial do conjunto de dados por meio das camadas recorrentes enquanto a camada totalmente conectada é utilizada para a classificação final da sequência de entrada.

As camadas iniciais da rede utilizam filtros convolucionais unidimensionais causais com taxas de dilatação dobradas a cada etapa [15]. Esta estruturação é usualmente utilizada a fim de obter atributos em dados sequenciais. Os parâmetros desta camada, apresentados na Tabela II, são a quádrupla (*número de filtros, tamanho do kernel, stride, taxa de dilatação*). As camadas seguintes implementam a estrutura recorrente da rede por meio de camadas GRU (*Gate Recurrent Unit*) bidirecionais com funções de ativação *tanh*, estrutura já utilizadas na decodificação de códigos convolucionais em canais AWGN [16]. A arquitetura recorrente é a responsável pelo processamento sequencial dos dados recebidos e deve realizar estimativas em cada instante de tempo a respeito dos dados de entrada. Camadas de *Batch Normalization* são empregadas intercaladamente em toda a rede buscando estabilidade nos gradientes de treinamento da rede [17]. A camada totalmente conectada faz a estimativa final da classificação de \mathbf{X} nos possíveis símbolos transmitidos por meio da camada *softmax*, a qual normaliza as funções de ativação da camada anterior buscando obter uma distribuição de probabilidades sobre as classes (00, 01, 10, 11). A arquitetura completa da rede é apresentada na Tabela II.

B. Treinamento

Para o treinamento são utilizados dois conjuntos de dados: os conjuntos de treinamento e validação. O conjunto de treinamento é formado por 10000 blocos compostos por pares (\mathbf{X}, \mathbf{y}) . Cada elemento da matriz \mathbf{X} é normalizado subtraindo a média dos elementos de \mathbf{X} e dividindo pelo desvio padrão. Consideramos que a rede conhece exatamente os coeficientes h_{ij} e o objetivo desta é aprender a estrutura dinâmica gerada pelo sistema Sprott D. O desempenho do sistema é sensível à escolha da SNR para treinar a rede [9]. Para um canal AWGN, o treinamento deve ser feito para a pior SNR em que se deseja testar a rede neural [16]. Na presença de desvanecimento, este procedimento gera um conjunto de treinamento composto de formas de onda de energia variadas.

O conjunto de validação testa iterativamente o desempenho da rede neural durante cada etapa do treinamento. A SNR utilizada para este conjunto deve ser diferente da utilizada para o treinamento, uma vez que busca-se verificar a capacidade de generalização da rede. Os valores de SNR utilizados para

ambos os conjuntos dependem da quantidade de antenas utilizadas na recepção uma vez que o ganho de diversidade afeta a taxa de erro de bloco, os valores foram estimados empiricamente por meio do treinamento para valores variados de SNR. Para uma antena (resp. quatro antenas) a SNR de treinamento e validação são 9 dB e 16 dB (resp. 4,5 dB e 7,5 dB), respectivamente. O objetivo do treinamento é minimizar a função custo [14]

$$J(\theta) = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{\ell} l(\hat{y}_i^{(k)}, y_i^{(k)}) \quad (4)$$

em que θ são os parâmetros treináveis da rede, o *batch size* K é a quantidade de tuplas (\mathbf{X}, \mathbf{y}) utilizadas durante o cálculo de J , $\hat{y}_i^{(k)}$ é a estimativa da rede sobre $y_i^{(k)}$ correspondente à tupla k no instante de tempo i e ℓ é a função entropia cruzada por amostra [14].

Foi utilizado o algoritmo de otimização Adam [18] para atualização dos pesos da rede θ . Foram utilizadas duas heurísticas para o controle e convergência do treinamento. A taxa de aprendizagem utilizada pelo otimizador Adam foi reduzida por um fator de 10 caso não ocorra uma melhora do desempenho do sistema durante 4 iterações consecutivas. Da mesma forma, a quantidade de iterações de treinamento é determinada pela estagnação da mesma métrica no conjunto de validação, isto é, o treinamento acontece enquanto a métrica de interesse continuar melhorando no conjunto de validação [19].

V. ANÁLISE DE DESEMPENHO

O desempenho do sistema CB-STTC é analisado via simulações computacionais. A decodificação de máxima verossimilhança é realizada pelo algoritmo de Viterbi [20] com uma treliça derivada do FSE apresentado na Tabela I com incremento de métrica dado por (3). Na decodificação pela rede neural, os incrementos de métrica são os dados de entrada.

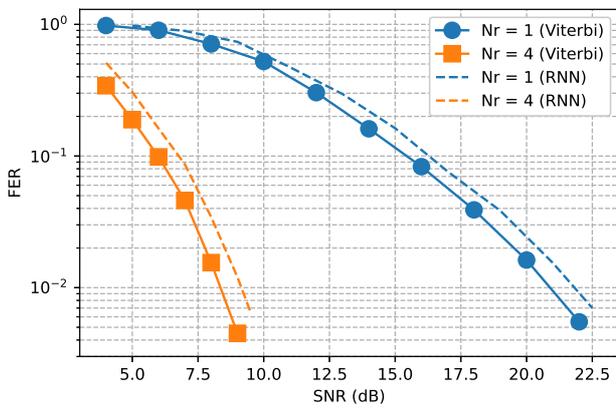
A Fig. 5 mostra as curvas da taxa de erro de bloco (FER), em que cada bloco tem comprimento $\ell = 130$, obtidas para $N_T = 2$ antenas transmissoras e $N_R = 1$ e 4 antenas receptoras. Observa-se o ganho de diversidade com o aumento de antenas receptoras, uma vez que as matrizes $\mathbf{A}(c, e)$ têm posto máximo. O desempenho obtido pela arquitetura de rede neural proposta segue o mesmo comportamento das curvas obtidas pelo algoritmo de Viterbi, indicando que a rede tem a capacidade de aprender a estrutura do CB-STTC, porém com desempenho inferior.

VI. CONCLUSÕES

Neste trabalho foi proposto o esquema CB-STTC com $N_T = 2$ antenas transmissoras e N_R antenas receptoras que apresenta ganho de diversidade máximo $2N_R$. Também foi proposta uma arquitetura de rede neural para decodificação do sistema CB-STTC que apresenta uma perda de desempenho de aproximadamente 1,2 dB em relação à decodificação de máxima verossimilhança. Portanto, a proposta é viável em aplicações envolvendo sistemas caóticos em canais sem fio. Uma possível continuação deste trabalho é considerar outros atratores caóticos bem como projetar esquemas CB-STTC com maior número de estados.

TABELA II
 ARQUITETURA DA REDE NEURAL EMPREGADA.

Tipo da Camada	Parâmetros	Dimensão da Saída (K = Batch Size = 32)
Causal Conv1D	(32, 3, 1, 1)	(K, ℓ , 32)
Causal Conv1D	(32, 3, 1, 2)	(K, ℓ , 32)
Causal Conv1D	(32, 3, 1, 4)	(K, ℓ , 32)
Causal Conv1D	(32, 3, 1, 8)	(K, ℓ , 32)
Batch Normalization	Sem Parâmetros	(K, ℓ , 32)
GRU Bidirecional	200 nós, <i>tanh</i>	(K, ℓ , 400)
Batch Normalization	Sem Parâmetros	(K, ℓ , 400)
GRU Bidirecional	80 nós, <i>tanh</i>	(K, ℓ , 160)
Batch Normalization	Sem Parâmetros	(K, ℓ , 160)
GRU Bidirecional	40 nós, <i>tanh</i>	(K, ℓ , 80)
Batch Normalization	Sem Parâmetros	(K, ℓ , 80)
TimeDistributed Totalmente Conectada	4 nós, <i>softmax</i>	(K, ℓ , 1)


 Fig. 5. FER versus SNR do sistema CB-STTC proposto com $N_T = 2$ e $N_R = 1$ e 4. Cada bloco possui comprimento $\ell = 130$.

REFERÊNCIAS

- [1] G. Kaddoum, "Wireless chaos-based communication systems: A comprehensive survey," *IEEE Access*, vol. 4, pp. 2621–2648, May 2016.
- [2] J. Yao, Y. Sun, H. Ren, and C. Grebogi, "Experimental wireless communication using chaotic baseband waveform," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 578–591, Jan. 2019.
- [3] H.-P. Ren, M. S. Baptista, and C. Grebogi, "Wireless communication with chaos," *Phys. Rev. Lett.*, vol. 110, pp. 184 101–1 – 184 101–5, April 2013.
- [4] J.-L. Yao, C. Li, H.-P. Ren, and C. Grebogi, "Chaos-based wireless communication resisting multipath effects," *Phys. Rev. E*, vol. 96, p. 032226, Sept. 2017.
- [5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, March 1998.
- [6] R. S. Blum, "Some analytical tools for the design of space-time convolutional codes," *IEEE Transactions on Communications*, vol. 50, no. 10, pp. 1593–1599, Oct. 2002.
- [7] Jinhong Yuan, Zhuo Chen, B. Vucetic, and W. Firmanto, "Performance and design of space-time coding in fading channels," *IEEE Transactions on Communications*, vol. 51, no. 12, pp. 1991–1996, Dec 2003.
- [8] C. E. C. Souza, C. Pimentel, and D. P. B. Chaves, "A symbolic dynamics approach to trellis-coded chaotic modulation," *IEEE Transactions on Circuits and Systems II: Express Briefs*, pp. 1–1, Nov. 2019.
- [9] T. Gruber, S. Cammerer, J. Hoydis, and S. t. Brink, "On deep learning-based channel decoding," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, March 2017, pp. 1–6.
- [10] I. Be'Ery, N. Raviv, T. Raviv, and Y. Be'Ery, "Active deep decoding of linear codes," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 728–736, Feb. 2020.
- [11] J. C. Sprott, "Some simple chaotic flows," *Phys. Rev. E*, vol. 50, pp. R647–R650, Aug. 1994.
- [12] R. Gilmore and M. Lefranc, *The Topology of Chaos: Alice in Stretch and Squeezeland*, 2nd ed. Wiley-VCH, 2012.
- [13] T. Carneiro, R. V. Medeiros Da Nóbrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access*, vol. 6, pp. 61 677–61 685, 2018.
- [14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: The MIT Press, Jan. 2016.
- [15] A. V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," *arXiv:1609.03499*, Sept. 2016.
- [16] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication Algorithms via Deep Learning," in *Proc. Int. Zurich Seminar Inf. Commun. (IZS)*, Feb. 2018, pp. 48–50.
- [17] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167*, March 2015.
- [18] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980*, Jan. 2017.
- [19] G. Raskutti, M. J. Wainwright, and B. Yu, "Early stopping for non-parametric regression: An optimal data-dependent stopping rule," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept. 2011, pp. 1318–1325.
- [20] T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*, 1st ed. Wiley, 2005.