Physics-informed dual architecture neural networks for enhanced estimation of drilling dynamics

Jeanne Redaud, Maxime Darrin, Nasser Kazemi Nojadeh, Jean Auriol

PHYSICS-INFORMED DUAL ARCHITECTURE NEURAL NETWORKS FOR ENHANCED ESTIMATION OF DRILLING DYNAMICS

Jeanne Redaud	*Maxime Darrin*	*Nasser Kazemi*	*Jean Auriol*
Université Paris-Saclay, CentraleSupélec, L2S, Gif-sur-Yvette, France.	Université Paris-Saclay, CentraleSupélec, ILLS, Montreal, Canada.	UQAM, Earth Sciences Dept, Montreal, Canada.	Université Paris-Saclay, CentraleSupélec, L2S, Gif-sur-Yvette, France.

ABSTRACT

While drilling, the interaction with the rock can generate significant vibrations, resulting in an inefficient rate of penetration. Control strategies based on accurate distributed dynamics of the drill string have proven more efficient than classical approaches to prevent stick-slip apparition during operations. However, implementing such strategies requires knowledge of the physical parameters and state estimations and predictions. To overcome this limitation, we propose a dual architecture of transformer-based physics-informed neural networks. We obtain physics-guided estimations of the angular velocity and torque by adding physical constraints during the training. Experimental results are given on simulated data.

Index Terms— Physics-informed neural networks, parameter estimation, drilling optimization

1. INTRODUCTION

This study focuses on machine learning-based solutions for parameter estimation and drilling optimization in the context of closed-loop geothermal systems (CGS). Geothermal energy, offering significant potential as a renewable electricity source, has recently garnered increased attention. New closed-loop geothermal power plants have been developed to address the limitations of conventional open-loop systems. To maximize their efficiency, these plants often require a deviated well path with an extensive horizontal section at depths greater than 2.5 km [1]. Advanced drilling control strategies are essential for designing such complex wells, as undesirable vibrations can occur even during transient off-bottom phases. Notably, torsional oscillations, which cause damage and lower the rate of penetration, must be mitigated. This phenomenon, known as *stick-slip*, arises from significant friction between the drill string and the borehole [2]. To avoid the apparition of such oscillations before the beginning of the drilling operation, advanced control approaches have been developed. Incorporating the underlying distributed dynamics into controller design is highly beneficial [3]. More precisely, the field-validated model given in [4] accurately describes the evolution of the angular velocity and torque $(\omega(t,x),\tau(t,x))$ along the drilling device, using a set of coupled hyperbolic partial differential equations (PDE). Utilizing this model, a motor torque control input designed through a *recursive dynamics interconnection framework* in [3, 5] has demonstrated a faster convergence rate to a reference trajectory compared to state-of-the-art controllers while preventing the apparition of stick-slip oscillations. However, such a control law requires the knowledge of the distributed state along the drill string. On the field, the available measurements mainly consist of surface data, for instance, surface rotation per minute (RPM) or motor torque. Estimation strategies using predictors based on the aforementioned distributed model require a high computational effort, such that the estimated state may not be obtained in real-time, preventing these approaches from being used in the field. Moreover, these estimation methods require all parameters of the system (and, in particular, subsurface physical properties) to be known [6, 7]. In Geoscientific applications, such as seismic imaging and full-waveform inversion or drilling optimization, we usually deal with a physical system that requires solving partial differential equations with boundary constraints. Also, in real-world applications, it is rare to find solutions or labels to train the physical network based on pure data-centric methods. Hence, we take advantage of the model-based solutions, especially in the backpropagation stage of the neural network, to honor the physical constraints. In other words, the theory or physics-guided neural networks use the hybrid approach and incorporate the physics of the system and the available datasets to train the physics-consistent networks [8, 9].

In this paper, we tackle these issues by proposing an innovative *dual architecture of physics-informed transformer-based neural networks* [10]. Indeed, recent advances have shown that deep neural networks can be used to learn distributed dynamics from measurements [11, 12, 13]. Physics-informed neural networks have improved the generalization capacity of the different solutions [14, 15]. They considered the underlying physics as a priori knowledge in the training process. Therefore, such physics-informed neural networks can take advantage of the available field-validated model describing the dynamics of the drilling device [2] and lead to better performance compared to classical neural network algorithms, as the ones proposed in [16]. In our architecture, the first branch estimates the evolution of the physical parameters used in the physical model from the available surface measurements. Then, the second branch estimates the distributed states $(\omega(t,x),\tau(t,x))$, using a recurrent network trained using the model from [2].

2. METHODOLOGY

2.1. Distributed model for torsional dynamics

In this work, we consider that the torsional motion of a drilling device of length L corresponds to the dominating dynamic behavior. The bit is assumed to be off-bottom, i.e., there is no bit-rock interaction. An off-bottom bit occurs in transient phases when the bit is not in contact with the rock.It was observed that stick-slip oscillations may appear during this transient phase [4]. For the sake of clarity, we briefly present the field-validated distributed torsional model given in [4]. For $(t,x) \in [0,T] \times [0,L]$ (where $T > 0$ is the chosen time window and $L > 0$ is the length of the drilling device), we denote $\omega(t,x)$ the angular velocity and $\tau(t,x)$ the torque at any point of the drill string. The states satisfy the following equations

$$\frac{\partial \tau(t,x)}{\partial t} + JG\frac{\partial \omega(t,x)}{\partial x} = 0, \tag{1}$$

$$J\rho\frac{\partial \omega(t,x)}{\partial t} + \frac{\partial \tau(t,x)}{\partial x} = -\mathcal{F}(t,x), \tag{2}$$

where J is the polar moment for inertia, G is the shear modulus, and ρ is the mass density, averaged for a drill string section and supposedly known. The source terms $\mathcal{F}$ is due to frictional contact with the borehole. After neglecting the viscous shear stresses, the source term can be modeled using the following differential inclusion (equation (3))

$$\begin{cases} \mathcal{F}(t,x) = \text{sign}(\omega(t,x))F_k(x), & |\omega(t,x)| > \omega_c, \\ \mathcal{F}(t,x) \in \pm r_o(x)\mu_s F_N(x), & |\omega(t,x)| \leqslant \omega_c, \end{cases} \tag{3}$$

where the function $\mathcal{F}(\omega) = -\frac{\partial \tau(t,x)}{\partial x} \in \pm r_o(x)\mu_s F_N(x)$ denotes the inclusion. The functions $F_k(x) \doteq r_o(x)\mu_k F_N(x)$ (resp. $F_s(x) \doteq r_o(x)\mu_s F_N(x)$) corresponds to the dynamic (resp. static) Coulomb torque. The expression of the normal force acting between the drill string and the borehole wall $F_N(x)$ depends on the well geometry [17]. While the outer drill string radius $r_o(x)$ is known, the static and kinetic friction coefficients (μ_s,μ_k) and the angular velocity threshold ω_c are not. At the surface level, the top drive is actuated by an electrical motor that imparts torque to the drill string. The evolution of the angular velocity at the surface level and the downhole torque follows

$$\frac{d}{dt}\omega(t,0) = \frac{1}{I_{TD}}(\tau_m(t) - \tau(t,0)), \quad \tau(t,L) = 0 \tag{4}$$

where I_{TD} corresponds to the top-drive inertia and τ_m to the motor torque, which may be expressed in closed-loop as a PI feedback [18] or through more advanced control strategies [2, 3, 5]. However, such advanced mitigation laws require the knowledge of the whole distributed angular state. On the field, we usually have access to the top-drive angular velocity and to the motor torque $(\omega(t,0),\tau_m(t))$ measured with a frequency of $1Hz$. It is, therefore, necessary to reconstruct the whole state from these measurements, as proposed in [19]. Nevertheless, reconstructing the state may require the knowledge of unknwon physical parameters (as the static and kinetic friction coefficients) In the following, we denote $Y = (P, \{(\omega(t_i,0),\tau_m(t_i))_{i\in[1,N]}\}) \in \mathcal{Y} = \mathbb{R}^{N_p} \times \mathbb{R}^{2\times N}$ an input data containing a set of N_p known physical parameters (depth of bit L, collar and pipes mechanical properties) concatenated in the vector P and N sequential pairs of surface measurements.

2.2. Problem setup

From surface measurements, our objective is now twofold:

1. Estimate the physical parameters (μ_s,μ_k) of the model from the surface data Y,
2. Estimate the distributed state $(\omega(t,x),\tau(t,x))$ at time t using sequences of past measurements.

Several state and parameter estimation procedures have been presented in [16], emphasizing their advantages and drawbacks. Here, we propose a new learning methodology based on a *dual architecture*. The network takes the sequences Y as input. A first transformer-based neural network estimates the physical parameters (μ_s,μ_k). We choose the transformers [10] over recurrent neural networks to increase speed and reduce the computational cost of the training. Then the estimated physical parameters (μ_s,μ_k) and original input Y are fed to the second transformer-based neural network, aiming to provide the distributed state $(\omega(t,x),\tau(t,x))$ along the drill string. A schematic representation is given in Figure 1.

Following [20, 21], we incorporate the physical laws in the training loss functions to improve the performance of the proposed estimation algorithms. To compute them, the output of the first network is necessary. Both networks are trained simultaneously using an exhaustive training dataset of simulated data, as detailed in Section 3.

2.3. Transformer-based dual architecture network

In this section, we present our network architecture for estimating the physical parameters and the distributed state for the drill string system. The proposed network is schematically illustrated in Figure 2.

2.3.1. Estimation of physical parameters

The first neural network aims at estimating the physical parameters $M = (\mu_k,\mu_s) \in \mathbb{R}^2$. This neural network is a

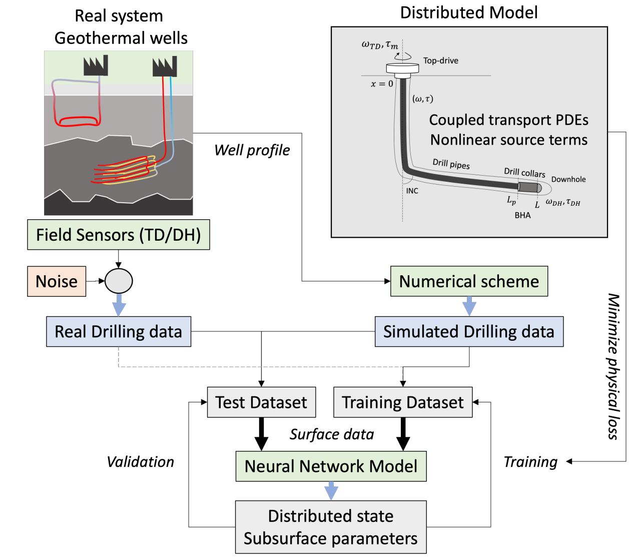

Fig. 1: Schematic representation of the overall approach

transformer [10] characterized by trainable parameters θ, and denoted $T_\theta(.)$. It aggregates the sequence of inputs $Y_i \in \mathbb{R}^{N_p+2N_t}$ and gives as an output $\hat{M} \in \mathbb{R}^2$. We obtain the trainable parameters θ by minimizing the following L_2- loss function $\mathcal{L}_{L_2}(\theta) = \frac{1}{N_b}\sum_{i=1}^{N_b} \|M_i - T_\theta(Y_i)\|_2^2$.

2.3.2. *Estimation of the distributed state*

Inspired by [11], we propose to tackle the second objective using a two-branch architecture. Let us denote $\mathcal{X} = [0, L]$, the space domain, and $\mathcal{T} = [0, T]$ the temporal domain where the state is defined. We aim to select the most appropriate state representation concerning physical constraints using sets of discrete inputs Y_i. Thus, we ought to approximate

$$S: \begin{array}{ccc} \mathcal{Y} & \longrightarrow & \mathcal{C}^\infty(\mathcal{T}\times\mathcal{X}, \mathbb{R}^2) \\ Y & \mapsto & S_{Y,\Theta}(\cdot,\cdot) \end{array},$$

where $\mathcal{C}^\infty$ is the set of infinitely differentiable functions from $\mathcal{T}\times\mathcal{X}$ to $\mathbb{R}^2$ (the regularity could be leveraged). This defines a class of parametric functions $S_{Y,\Theta}$, which we use for approximating the real states $(\omega(t,x), \tau(t,x))$. The corresponding approximations is denoted as $(\hat{\omega}(t,x), \hat{\tau}(t,x))$. In our design, the first branch of the proposed architecture relies on a transformer encoder to aggregate the input sequences Y_i augmented with $\hat{M}$ (obtained following the procedure described in Section 2.3.1). It produces an abstract representation of the system when combined with an abstract representation of the requested coordinates. Inspired by Fourier Neural Operators [13], it is then used to output the intensity, frequency, and phase of the Fourier decomposition representing the distributed state $X(t,x) \doteq (\omega(t,x), \tau(t,x))$ along the drill string. The second branch builds the spatiotemporal grid mesh $(t,x) \in \mathcal{T}\times\mathcal{X}$, where the estimation is evaluated.

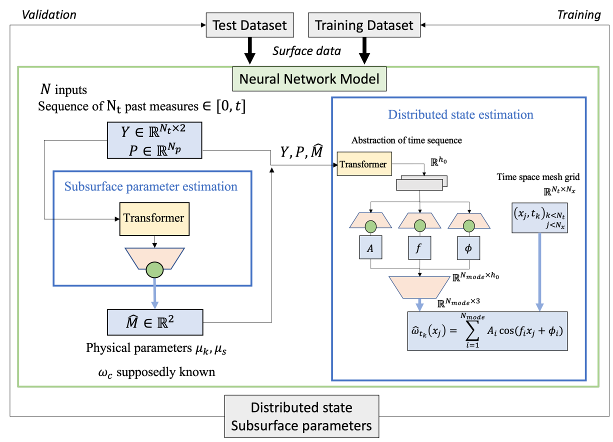

Fig. 2: Detailed view of the neural network architecture

2.4. Physic-informed neural networks

Following [21, 20, 14], we define a composite loss that takes into account experimental data ($\mathcal{L}_\mathcal{D}$) as well as a physical error. While the former is the usual error term when training a neural network on a dataset, the latter ensures that the solution verifies theoretical PDEs ($\mathcal{L}_{\text{PDE}}$) and boundary conditions ($\mathcal{L}_{\text{BC}}$).

$$\mathcal{L}(\Theta) = \mathcal{L}_\mathcal{D}(\Theta) + \mathcal{L}_{\text{PDE}}(\Theta) + \mathcal{L}_{\text{BC}}(\Theta) \qquad (5)$$

It should be seen as data and physics-driven losses. This adds *a priori* knowledge of the underlying dynamics (1)-(4) during training. As before, the first loss term corresponds to the state estimation residual in the squared L_2-norm: $\mathcal{L}_\mathcal{D}(\Theta) = \frac{1}{N_x}\frac{1}{N_t}\frac{1}{N_b}\sum_{i=1}^{N_b}\sum_{j=1}^{N_x}\sum_{k=1}^{N_t}\|X^i(t_k,x_j) - S_{Y_i,\Theta}(t_k,x_j)\|_2^2$. The following term ensures that (1)-(2) are satisfied $\mathcal{L}_{\text{PDE}}(\Theta) = \frac{1}{N_x}\frac{1}{N_t}\frac{1}{N_b}\sum_{i=1}^{N_b}\sum_{j=1}^{N_x}\sum_{k=1}^{N_t}\|\mathcal{O}(f_{Y_i,\Theta}(t_k,x_j))\|_2^2$ with $\mathcal{O}(S_{Y_i,\Theta}(t,x)) = \begin{pmatrix} \frac{\partial\hat{\tau}^i}{\partial t}(t,x) + JG\frac{\partial\hat{\omega}^i}{\partial x}(t,x) \\ \frac{\partial\hat{\tau}^i}{\partial x}(t,x) + J\rho\frac{\partial\hat{\omega}^i}{\partial t}(t,x) + \mathcal{F}(\hat{\omega}^i,x) \end{pmatrix}$. Finally, we want the solution to meet the boundary condition (4), which induces the error term

$$\mathcal{L}_{\text{BC}}(\Theta) = \frac{1}{N_t}\frac{1}{N_b}\sum_{i=1}^{N_b}\sum_{k=1}^{N_t}\|\mathcal{B}(S_{Y_i,\Theta}(t_k))\|_2^2 + |\hat{\tau}^i(t_k,L)|^2,$$

with $\mathcal{B}(S_{Y_i,\Theta}(t)) = \frac{\partial\hat{\tau}^i}{\partial t}(t,0) - \frac{1}{I_{TD}}(\tau_m^i(t) - \hat{\tau}^i(t,0))$.

3. EXPERIMENTAL SETTING

3.1. Generation of dataset

To train and validate our estimation algorithm, we generate a wide dataset following the numerical scheme[1] presented in

[1]The Matlab implementation can be found on https://github.com/Open-Source-Drilling-Community/Aarsnes-and-Shor-Torsional-Model.

[4]. To obtain representative data, we use the real well geometry (J1), illustrated in Figure 3. We generated 1000 sequences of 100s of 20Hz surface measurements (motor torque and surface angular velocity), for $L \in [2500, 4000]$m, $\mu_s \in [0.2, 0.8]$, $\mu_k \in [0.06, 0.72]$.

Remark The first limitation is that this dataset contains only samples for a single value of angular velocity threshold $\omega_c = 1.5$rad.s^{-1} and zero initial conditions. The reference trajectory is constant (60RPM with slope). The implemented control input is a PI control with fixed gains.

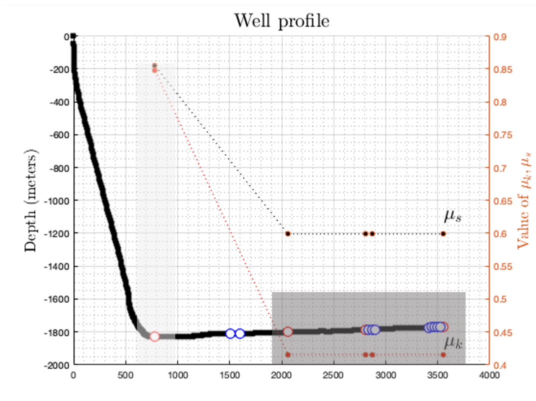

Fig. 3: Schematic profile of well J1.

3.2. Training parameters

The dataset is split $80\% - 20\%$ as training and validation datasets. To obtain the parameters (θ, Θ) minimizing the losses $\mathcal{L}_{L_2}(\theta), \mathcal{L}(\Theta)$, we use AdamW [22] with an initial learning rate of 10^{-3}. The training is done on 100 epochs, with a batch size $N_b = 16$. We use automatic differentiation techniques [23] to compute the derivatives.

3.3. Simulation results

Both networks are trained simultaneously using an exhaustive training dataset of simulated data generated using the proposed well geometry. The performance of the trained neural networks is evaluated on a different set of simulated data (validation dataset) as explained in [16]. We compare the results with the estimation obtained using the convolutional neural network-based strategy proposed in [16]. The friction coefficients were estimated on the validation dataset with an average relative error of $\delta(\mu_k) = 2.3\%$ (resp. $\delta(\mu_s) = 3.3\%$) and a standard deviation of $5.2e^{-3}$(resp. $1.8e^{-2}$) after 2500 steps. As illustrated in Figure 4, our proposed method outperformed the existing estimation methods presented in [16], as the standard deviation is reduced, and the average estimated value is closer to the true value. We also obtained promising results for the state estimation with an average absolute error of 2.3 on 200 validation examples. We have plotted in Figure 5 the obtained estimation for τ and ω and compared them

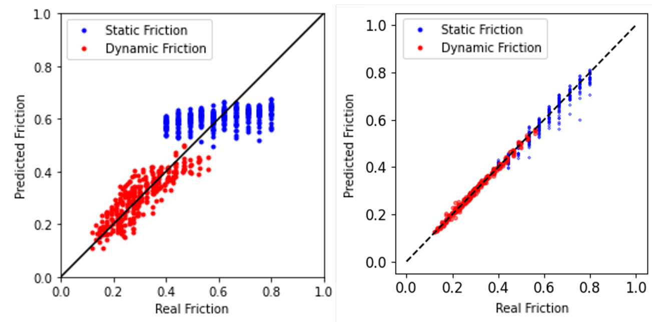

Fig. 4: Regression performance for the estimation of (μ_k, μ_s) using the algorithm from [16] (left) and our approach (right).

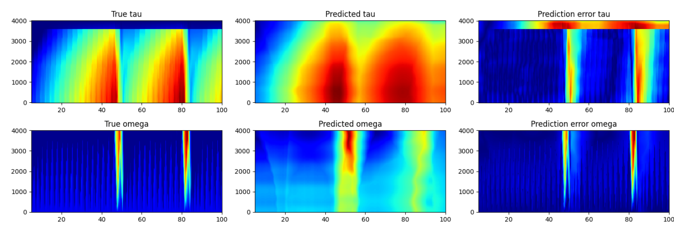

Fig. 5: Example of state prediction for the validation set after training and comparison with real values.

with their real values. We used color plots to picture these 2D data (red corresponding to higher values, the horizontal axis being time, and the vertical axis being the curvilinear abscissa). As expected, the solutions predicted by the proposed networks are consistent with the physics.

4. CONCLUSIONS

Using a combined transformer approach, we achieved real-time estimation of the distributed torsional states by utilizing available top-drive measurements. Our approach enabled accurate estimations of unknown underground physical parameters. Due to the limited availability of field data from similar subsurface environments and drilling assemblies, we relied on synthetic training data. However, with additional field data, it is feasible to construct a hybrid physics-data model. The current architecture includes numerous hyperparameters that can be optimized to enhance estimation performance, particularly by balancing the empirical loss of model predictions, model complexity, and physical loss. This strategy could improve generalization performance, especially since the field training data are limited and not fully representative. In future work, state estimations will be used to compute control laws that efficiently mitigate undesired stick-slip oscillations, with a primary focus on the robustness of the control law concerning state estimation errors. Additionally, this approach will be combined with *Fourier neural operators* to predict future values of the distributed state.

5. REFERENCES

[1] C. Wei, L. Mao, C. Yao, and G. Yu, "Heat transfer investigation between wellbore and formation in U-shaped geothermal wells with long horizontal section," *Renewable energy*, vol. 195, pp. 972–989, 2022.

[2] U. J. F. Aarsnes and R. J. Shor, "Stick-slip and torsional friction factors in inclined wellbores," in *MATEC Web of Conferences*, Les Ulis, 2018, vol. 148, p. 16002, EDP Sciences.

[3] J. Auriol, I. Boussaada, R. J. Shor, H. Mounier, and S.-I. Niculescu, "Comparing advanced control strategies to eliminate stick-slip oscillations in drillstrings," *IEEE access*, vol. 10, pp. 10949–10969, 2022.

[4] U. J. F. Aarsnes and R. J. Shor, "Torsional vibrations with bit off bottom: Modeling, characterization and field data validation," *Journal of Petroleum Science and Engineering*, vol. 163, pp. 712–721, apr 2018.

[5] J. Redaud, J. Auriol, and S.-I. Niculescu, "Recursive dynamics interconnection framework applied to angular velocity control of drilling systems," in *2022 American Control Conference (ACC)*, 2022, pp. 5308–5313.

[6] Jean Auriol, Nasser Kazemi, Roman J Shor, Kristopher A Innanen, and Ian D Gates, "A sensing and computational framework for estimating the seismic velocities of rocks interacting with the drill bit," *IEEE Transactions on geoscience and remote sensing*, vol. 58, no. 5, pp. 3178–3189, 2019.

[7] Jean Auriol, Nasser Kazemi, and Silviu-Iulian Niculescu, "Sensing and computational frameworks for improving drill-string dynamics estimation," *Mechanical Systems and Signal Processing*, vol. 160, pp. 107836, 2021.

[8] Jian Sun, Zhan Niu, Kristopher A Innanen, Junxiao Li, and Daniel O Trad, "A theory-guided deep-learning formulation and optimization of seismic waveform inversion," *Geophysics*, vol. 85, no. 2, pp. R87–R99, 2020.

[9] Siwei Yu and Jianwei Ma, "Deep learning for geophysics: Current and future trends," *Reviews of Geophysics*, vol. 59, no. 3, pp. e2021RG000742, 2021.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[11] L. Lu, P. Jin, and G. E. Karniadakis, "DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 218–229, Mar. 2021, arXiv:1910.03193 [cs, stat].

[12] K. Bhattacharya, B. Hosseini, N. Kovachki, and A. Stuart, "Model reduction and neural networks for parametric pdes," 2020.

[13] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, and A. Stuart, "Fourier neural operators for parametric partial differential equations," p. 16, 2021.

[14] S. Wang, H. Wang, and P. Perdikaris, "Learning the solution operator of parametric partial differential equations with physics-informed DeepONets," *Science Advances*, vol. 7, no. 40, pp. 8605, Oct. 2021.

[15] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, "Physics-informed neural operator for learning partial differential equations," 2021.

[16] J. Auriol, R. Shor, S.-I. Niculescu, and N. Kazemi, "Estimating drill string friction with model-based and data-driven methods," in *2022 American Control Conference (ACC)*, 2022, pp. 3464–3469.

[17] M. C. Sheppard, C. Wick, and T. Burgess, "Designing Well Paths To Reduce Drag and Torque," *SPE Drilling Engineering*, vol. 2, pp. 344–350, 1987.

[18] K. J. Åström and R. M. Murray, *Feedback systems: an introduction for scientists and engineers*, Princeton university press, 2nd edition, 2010.

[19] U. J. F. Aarsnes, J. Auriol, F. Di Meglio, and R. Shor, "Estimating friction factors while drilling," *Journal of Petroleum Science and Engineering*, vol. 179, pp. 80–91, 2019.

[20] J. Sun, K. A. Innanen, and C. Huang, "Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis," *GEOPHYSICS*, vol. 86, no. 3, pp. R303–R317, May 2021.

[21] M. Raissi, P. Perdikaris, and G.E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, Feb. 2019.

[22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017.

[23] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," 2015.