Optimal Tracking Current Control of Switched Reluctance Motor Drives Using Reinforcement Q-Learning Scheduling

HAMAD ALHARKAN\textsuperscript{1,2},\ (Student Member, IEEE), SEPEHR SAADATMAND\textsuperscript{1},\ (Student Member, IEEE), MEHDI FERDOWSI\textsuperscript{1},\ (Member, IEEE), AND POURYA SHAMSI\textsuperscript{1},\ (Senior Member, IEEE)

\textsuperscript{1}Electrical Engineering Department, Missouri University of Science and Technology, Rolla, MO 65401, USA
\textsuperscript{2}Department of Electrical Engineering, College of Engineering, Qassim University, Unaizah 56452, Saudi Arabia

Corresponding author: Hamad Alharkan (haahx9@mst.edu)

ABSTRACT In this article, a novel Q-learning scheduling method for the current controller of a switched reluctance motor (SRM) drive is investigated. The Q-learning algorithm is a class of reinforcement learning approaches that can find the best forward-in-time solution of a linear control problem. An augmented system is constructed based on the reference current signal and the SRM model to allow for solving the algebraic Riccati equation of the current-tracking problem. This article introduces a new scheduled-Q-learning algorithm that utilizes a table of Q-cores that lies on the nonlinear surface of an SRM model without involving any information about the model parameters to track the reference current trajectory by scheduling the infinite horizon linear quadratic trackers (LQT) handled by Q-learning algorithms. Additionally, a linear interpolation algorithm is proposed to improve the transition of the LQT between trained Q-cores to ensure a smooth response as state variables evolve on the nonlinear surface of the model. Lastly, simulation and experimental results are provided to validate the effectiveness of the proposed control scheme.

INDEX TERMS Adaptive dynamic programming, current control, least square methods, motor drive, optimal control, reinforcement learning, switched reluctance motors.

I. INTRODUCTION

Lately, switched reluctance motor (SRM) has earned significant consideration for a wide range of transportation electrification and variable speed applications. This is because it has several inherences, such as a resilient and simple structure due to the lack of magnet, brushes, and rotor winding. Moreover, SRMs are efficient at high speed [1]. Based on the reduction in the power electronics costs, improved availability and performance of film capacitors to handle the pulse-type current of these machines, and the interest in the reduced utilization of rare-earth magnets, the utilization of SRMs for a variety of industrial and commercial applications has been on the rise [2]. This includes applications in traction drives as well as aeronautics where the high reliability, high temperature and vibration tolerance, and high-speed range of SRMs make them very competitive compared to more complex motors [3]–[7]. However, SRMs have suffered from certain drawbacks, including high acoustic noise production due to its torque ripple and flux paths and the expensiveness of drive due to a large number of semiconductor switches in its drive. Additionally, it has a highly nonlinear electromagnetic nature that is highly reliant on variations in the phase current and rotor positions. Many researchers have investigated SRMs to mitigate these issues by improving the SRM design to minimize torque ripples or developing a new converter topology using recently introduced and more affordable power electronics switches [8]–[11]. The highly nonlinear behavior of SRMs is the main challenge, which must be considered when designing an effective controller.

Unlike conventional sinusoidal motors, SRMs require pulse-type current that requires high variations of current (i.e. $di/dt$) and hence a high bandwidth drive system. To achieve a fast rate of current charge and discharge, a large dc-link voltage and low phase inductances are often needed. However, this dc-link voltage will make the regulation of phase currents more challenging, particularly during low speed operation modes. Traditionally, delta modulation or hystere-
sis current controllers have been used to regulate the phase current. Hysteresis-type controllers lead to a variable switching frequency, which is not of interest as managing the Electro-Magnetic Interferences (EMIs) becomes challenging. Additionally, power switches will impose an upper limit for the switching frequency and large current ripples will increase torque ripple and audible noises.

Many publications have investigated current control techniques for SRM, including enhanced hysteretic control, sliding-mode approaches, and fast PI controllers [12]–[17]. However, PI-based methods are slow, and methods such as delta-modulation will not be able to use the concept of duty-cycle to breakdown the switching cycle to shorter active periods. Therefore, a method is required to generate a duty cycle. Classical controllers such as PID controllers are not capable of controlling a system with such transients. Hence, researchers have investigated methods such as model predictive control and neural networks to cope with this issue [18]–[23]. To cope with the nonlinearities of the model, Ref [19] has introduced a Taylor expansion algorithm to approximate the variations of the model as a function of the rotor angle and current. Also, adaptive estimators are used to improve this approximation. However, the accuracy of the control is impacted by the Taylor expansion. As an improvement, Ref [24] has introduced a table-based inductance function that is used to form the model needed for the Model Predictive Control (MPC) in each cycle as oppose to a Taylor expansion. This table allows the MPC to have access to an accurate inductance value for a given rotor angle and current. Additionally, an adaptive estimator is used to update this table. Ref [24] has also introduced a linear interpolation technique for transitions between the models that will be incorporated in this article to introduce a novel scheduled Q-learning technique. In these literatures, a fundamental model is assumed, then an adaptive estimator is used to estimate the inductance of the phase as a function of rotor angle and current. Then, this value is used in a model predictive controller. The main drawbacks of the above works are the need for a separate estimator, a model predictive controller, and assumptions on the structure of the model.

In this article, the controller has been formulated based on an infinite-horizon linear quadratic tracker (LQT). To eliminate the need for a known model, a reinforcement Q-learning scheme is used to learn and apply the best course of action at each control cycle. Q-learning is inherently a linear controller [25], and on the other hand, the model of an SRM has nonlinearities to rotor angle and current (i.e. saturation). Hence, this article proposes a scheduled Q-learning algorithm that utilizes a table of Q cores, each containing a linear controller for a given rotor angle and current. By transitioning between these Q cores using a linear interpolation mechanism, this article introduces a nonlinear tracking controller capable of handling SRM drives.

The specific contributions of this article include i) introduction of Q-learning LQT for SRMs, ii) scheduling a table of Q-cores to achieve nonlinear control capabilities out of traditional Q-learning techniques, and iii) introducing a linear interpolation technique for transitioning between Q-cores to achieve a smooth Q scheduling.

The article is organized as follows: Section II reviews the Q-learning algorithm and introduces the proposed controller. Section III proposes the Q scheduling algorithm and table interpolation. Sections IV and V verify the effectiveness of the proposed controller through simulations and experimental results.

II. Q-LEARNING CONTROL OF SRM DRIVE

As discussed in Section I, there are several approaches to current control SRMs with their own drawbacks. For instance, the hysteresis technique downside is its rippled current, and torque, and consequently the acoustic noises and low efficiency. The conventional linear controllers such as PI/PID lack the ability to cope with nonlinear systems. The nonlinear techniques such as model predictive controller rely on an accurate system model, which can be changed during the time. Therefore, model-free reinforcement learning is a powerful tool to tackle all these drawbacks. Considering the expensive computation costs for neural network-based reinforcement learning such as online adaptive optimal control problem of a class of continuous-time Markov jump linear systems (MJLSs) [26] or Online policy iterative-based H∞ optimization algorithm [27], it is infeasible to implement them into power electronics control circuits. On the other hand, the Q-learning technique based on the Q-table only requires enough memory spaces that are available in most of the power electronics microcontrollers.

The primary target of the Q-learning algorithm in the current control of SRMs is to solve the LQT problem, which allows the system output to track a specific reference signal. The implementation of the algorithm in this format minimizes the predetermined value function associated with the cost of the policy and the difference between the output current and the reference signal. The classic solutions to the LQT can be found by solving the feedback part using the algebraic Riccati equation (ARE) and a feedforward part using the noncausal difference equation [28]. However, these approaches are not applicable for SRMs or most of the industrial applications since they are solved offline and they need accurate information on system dynamics. Adaptive dynamic programming is a part of the Reinforcement Learning (RL) methods and has been used to solve infinite-horizon LQT problems online without knowing system dynamics [29]. Two major assumptions have been made for this scheme: full state feedback is observable for the controller and the full reference trajectory is known. The LQT is a special case of model predictive controllers were the performance index is quadratic and no further constraints are applied to the optimizer. Deriving the quadratic form of the performance index for the LQT has been proved in [29]. The benefit of quadratic forms is the availability of algorithms that can solve Bellman equations online. To cope with the reference trajectory, an augmented system is generated by incorporating the reference current
trajectory into the state space model of SRM. This augmented system leads to the development of ARE, which provides the optimal solution for the LQT. By solving ARE, the feedback and feedforward parts of the policy for the classic solution of the LQT are solved at the same time [30]. The main drawback of using the LQT Bellman equation to solve this problem is that the accurate model of SRM is required [31].

To cope with this issue, Q-learning is utilized to learn and adapt itself to the optimal solution of this LQT online. The LQT Bellman equation and Q-learning algorithm for the SRM drive system are introduced in this section.

A. THE LQT BELLMAN EQUATION ALGORITHM OF SRM DRIVE

A schematic diagram of a switched reluctance motor is depicted in Fig. 1, and Fig. 2 shows the circuit diagram of an SRM driver in a 3-phase and detailed single phase. Driving an SRM requires a train of current pulses applied with respect to the rotor position. Due to the negligible mutual inductance between phases and as done traditionally, the mutual inductances are neglected to achieve a phase model that is independent of other phases. Taking into consideration the reference current, an augmented system can be formed by discretizing the SRM model using the forward approximation as

\[
X_{k+1} = \begin{bmatrix} A & 0 \\ 0 & F \end{bmatrix} X_k + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k \equiv A_x X_k + B_u u_k \quad (1)
\]

\[
Y_k = \begin{bmatrix} C & 0 \end{bmatrix} X_k \equiv C_x X_k \quad (2)
\]

where \(X_k = [x_k T^T]^T\) and \(A = 1 - TR/L_k, B = T/L_k, x_k\) is the phase current, \(u_k\) is the DC-bus voltage, \(R\) is the phase resistance, and the original output of the system \(y_k\) is the phase current, meaning that \(C = 1\). The parameter \(L_k\) is the nonlinear phase inductance as a function of both phase current and rotor position. Parameter \(T\) is the sampling time and \(F\) is the model of the reference trajectory (i.e. \(F = 1\) for a flat current). Due to the actual mechanical design of the machine, the nature of the inductance surface with respect to a rotor angle is periodic, starting from an unaligned position between rotor and stator poles until they are aligned. To solve the LQT problem and achieve tracking, the reference current generator pulses are assumed to be incorporated in the augmented system as in (1). It is expressed as

\[
r_{k+1} = Fr_k \quad (3)
\]

where \(r_k\) is the reference current trajectory and \(F \in R^n\) is the reference current generator. This can generate different types of waveforms, including a sequence of square waveforms, the reference current for SRM. Even the command generator \(r_k\) does not need to be constant. The value of \(\gamma\) should be less than 1 to attain a stable value function as the reference current in SRM is generated as a train of pulses and therefore has a positive dc average [32]. Based on (5), the value function relies on the current augmented state and an infinite horizon of the control inputs. By initializing the state of the value function with fixed control input, that infinite sum can be written as

\[
V(x_k) = \frac{1}{2} \{r_k - y_k\}^T Q \{r_k - y_k\} + u_k^T Ru_k + \gamma V(X_{k+1}) \quad (5)
\]

Equation (4) is equivalent to the LQT bellman equation. As it has been proved in [29] that the value function can be derived in a quadratic form and \(V(x_k) = \frac{1}{2} X_k^T PX_k\), the LQT Bellman equation with respect to a kernel \(P\) matrix is generated as

\[
X_k^T PX_k = x_k^T Q x_k + u_k^T Ru_k + \gamma X_{k+1}^T PX_{k+1} \quad (6)
\]

where \(P\) matrix is the optimum solution of ARE with elements derived in [29]. By obtaining the Hamiltonian function of the LQT and applying the stationary condition to obtain the optimal control policy (8), the solution of ARE that allows the
matrix $P$ to converge to its optimal values can be generated as

$$P = Q_{q} + \gamma A_a^T P A_a - \gamma^2 A_a^T P B_b (R + \gamma B_b^T P B_b)^{-1} B_b^T P A_a \quad (7)$$

Now, one may construct the algorithm based on the policy iteration method to solve the LQT problem by iterating the Bellman equation until convergence using data measured during the operation of the machine as in algorithm 1 as follows.

### B. THE Q-LEARNING ALGORITHM OF SRM DRIVE

Let’s assume that $L_k$ and hence the model of the machine is linear. For instance, the controller is operating while the variations of the current and angle of the rotor are negligible. This is due to the fact that the Q-learning algorithm utilized in this section can only operate on linear systems. In the next section, the nonlinearity is addressed through scheduling.

In Algorithm 1, Policy Integration (PI) is applied to LQT Bellman equation to acquire the optimum solution for ARE.

This algorithm requires all SRM dynamic parameters (i.e. $A_a$) to solve the LQT problem online. Q-learning is among the RL control methods that offer an adaptive tuning algorithm to track the reference signal online without requiring the system dynamic [33]. By extracting sets of data during the operation, including the reference current and augmented states, the algorithm can train Q-function until convergence at each iteration. The Q-function of LQT can be provided in matrix form by substituting the augmented model (1) and reference current in the LQT Bellman equation as

$$Q(X_k, u_k) = \frac{1}{2} \begin{bmatrix} X_k \end{bmatrix}^T \begin{bmatrix} Q_a + \gamma A_a^T P A_a & \gamma A_a^T P B_b \\ B_b^T P A_a & R + \gamma B_b^T P B_b \end{bmatrix} \begin{bmatrix} X_k \\ u_k \end{bmatrix} \quad (8)$$

which can be written as

$$Q(X_k, u_k) = \frac{1}{2} \begin{bmatrix} X_k \end{bmatrix}^T \begin{bmatrix} G_{XX} & G_{Xu} \\ G_{uX} & G_{uu} \end{bmatrix} \begin{bmatrix} X_k \\ u_k \end{bmatrix} \quad (9)$$

The Q-learning algorithm can be designed based on the policy iteration method to solve the LQT online in a way that ensures the system model parameters do not appear in the algorithm processes [34]. This process improves the control input until the system converges to the optimal level, which allows the output current in the SRM to follow the reference current. Algorithm 2 demonstrates the procedure of finding the solution to the Q-learning. In this algorithm, $M$ is defined as $M = [X_k \ u_k]^T$. Optimizing the Q-function in Algorithm 2 can be achieved as G matrix trains and converges to the optimum solution. The policy evaluation step for both algorithms 1 and 2 requires the solver to achieve convergence before updating the policy [31].

### III. Q-LEARNING SCHEDULING

In the previous section, the adaptive Q-learning algorithm controller for SRM was proposed to solve LQT and enable the current of the SRM drive to track a reference trajectory assuming that $L_k$ was constant. The inductance profile of the SRM is a nonlinear function of the current and the rotor angle. For instance, the inductance profile of the motor utilized later in the experimental section is shown in Fig. 3. In addition to this function, in the long term, effects such as aging of bearings and changes in the airgap, chemical degradation of the core such as rust, which can lead to changes in the airgap length, and temperature expansion can cause further variation in the inductance profile. Also, common manufacturing related variations such as variations in the airgap length, permeability, and even number of turns can cause some differences between the expected model and the actual inductance profile. To mitigate these effects, adaptive estimation approaches to update the dynamic parameters of the machine are of interest. Various methods have been utilized to estimate the inductance profile of SRM and update the nonlinear model of the SRM [35], [36]. However, these methods are unlike the proposed Q-learning approach, which can perform both tracking reference and adaptive estimation at the same time. The Q-learning by itself is not feasible or applicable to a nonlinear system such as an SRM. To address this issue, one can incorporate a proper local linearization scheme for the nonlinear inductance surface of SRM to allow the Q matrix to train in its locally linearized region.
A. TRAINING LOCAL Q MATRICES

In this article, the least square approach has been utilized to solve the tracking problem and learn the Q matrix by using enough data packets measured through the operating of the machine. The least square method does not require a system identification model. In practice, an observer is required to observe that states online. To implement policy evaluation, no less than $H = (m_x + m_y + m_y) \times (m_x + m_y + m_y + 1)/2$ data tuples are needed to perform LS method while $Q(X_k, u_k) = 1/2 M_k^T G M_k$ and the number of elements in G matrix are $(m_x + m_y + m_y) \times (m_x + m_y + m_y)$. This can be solved using the Kronecker product. The Kronecker product of $G_{m \times n}$ and $Z_{p \times q}$ can be defined as $G \otimes Z = \begin{bmatrix} g_{11} \cdot Z & \ldots & g_{1n} \cdot Z \\ \vdots & \ddots & \vdots \\ g_{m1} \cdot Z & \ldots & g_{mn} \cdot Z \end{bmatrix}$.

The Kronecker product enables the Q matrix to appear as columns of stacking vectors as

$$A \left( \text{vec}(G^T) \right) = B \quad (14)$$

The definition of $A_k$ and $B_k$ are expressed as

$$A = \begin{bmatrix} M_k \otimes M_k - \gamma M_{k+1} \otimes M_k + 1 \\ \vdots \\ M_k+2 \otimes M_k+2 - \gamma M_{k+3} \otimes M_{k+3} + 1 \\ (X_k^T) Q_k (X_k + (u_k)^T R(u_k)^T) \\ \vdots \\ (X_{k+2}^T) Q_{k+2} (X_{k+2} + (u_{k+2})^T R(u_{k+2})^T) \end{bmatrix}$$

$$B = \begin{bmatrix} (X_k^T) Q_k (X_k + (u_k)^T R(u_k)^T) \\ \vdots \\ (X_{k+2}^T) Q_{k+2} (X_{k+2} + (u_{k+2})^T R(u_{k+2})^T) \end{bmatrix}$$

where $z \geq H$ is the number of samples for each iteration. Then, the batch least square equation for solving Q matrix is provided as

$$\text{vec} \left( G^{i+1} \right) = (A^T A)^{-1} A^T B \quad (17)$$

By maintaining the persistence condition, least square may be applied iteratively by solving recursive least square (RLS) equations as

$$e_k(t) = Q(X_k, u_k) - A_k^T \overline{G}_k (t-1) \quad (18)$$

$$\overline{G}_k (t) = \overline{G}_k (t-1) + \frac{\eta_k (t-1) A_k e_k}{1 + A_k^T \eta_k (t-1) A_k} \quad (19)$$

$$\eta_k (t) = \eta_k (t-1) - \frac{\eta_k (t-1) A_k^T \eta_k (t-1)}{1 + A_k^T \eta_k (t-1) A_k} \quad (20)$$

where $t$ is the index of iterations of the RLS, $e$ is the error, and $\eta$ is the covariance matrix whereas $\eta_k (0) = \tau I$ for a big positive number $\tau$ while $I$ is an identity matrix.

B. TABLE DATA EXTRACTION AND LINEAR INTERPOLATION

Table readout algorithm is important to enabling extracting the knowledge from the Q-cores table and utilizing the data to improve policy. A table of Q-learning has been computed and formed in the previous section that contains the locations for current-rotor position points selected from the surface of the inductance profile. The typical current pulse for each SRM
phase placed on the Q-learning table is shown in Fig. 4-a. 
One method of implementing the Q-learning table is to use 
the optimal Q matrix that is located at near the current path. 
In this case, the algorithm will read the value of the current 
and measure the distance to neighboring matrices to find the 
nearest Q matrix. This process solves the problem of using 
only one learned Q matrix in the locally linearized region. 
Although, in practice, this method is relatively simple, it leads 
to transients in the current waveform every time the controller 
watches between two table elements.

The bilinear interpolation algorithm provides a smoother 
and more accurate scheduling than does the nearest Q matrix method. This algorithm divides the Q matrix among its four 
closest Q matrices neighbors in the opposite proportion of the 
distance, which means if the state of the system is located at 
equal distance from four Q neighbors, the values of scheduled 
Q matrix are divided equally; if it is a near one of the four 
matries, most of the scheduled Q matrix data are transmitted 
from that adjacent Q matrix. Observing the four neighboring 
Q matrices points \( Q_{11}, Q_{12}, Q_{21} \) and \( Q_{22} \), which are the 
four closest neighbors of scheduled Q matrix \( Q_s \), then \( Q_s \) is obtained as

\[
Q_s = \beta_0 + \beta_1 \theta + \beta_2 \theta i + \beta_3 \theta i^2
\]

(21)

where the coefficient of bilinear scheduling \( \beta_0, \beta_1, \beta_2 \) and \( \beta_3 \) 
are obtained by solving

\[
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\end{bmatrix} =
\begin{bmatrix}
1 & \theta_1 & i_1 & \theta_1 i_1 \\
1 & \theta_2 & i_2 & \theta_2 i_2 \\
1 & \theta_1 & i_1 & \theta_1 i_1 \\
1 & \theta_2 & i_2 & \theta_2 i_2 \\
\end{bmatrix}^{-1}
\begin{bmatrix}
Q_{11} \\
Q_{21} \\
Q_{12} \\
Q_{22} \\
\end{bmatrix}
\]

(22)

In practical implementation, to avoid solving systems of 
equations and performing matrix inversions that are not fea-
sible in a digital controller, and since the scheduled Q matrix 
lies on a square grid of four Q matrices, one can use a 
simplified algorithm based on a unit square, \( Q_s \) is computed as

\[
Q_s = [1 - l_2] Q_{11} + [1 - l_1] Q_{21} + l_1 [1 - l_2] Q_{12} + l_2 [1 - l_1] Q_{22}
\]

(23)

where \( l_1 \in [0, 1] \) and \( l_2 \in [0, 1] \) are the lengths 
between \( Q_s \) and the nearest Q matrix in the rotor angels 
and current axis, respectively. These lengths are calculated as 
shown in Fig. 4-b as \( l_1 = (\theta - \theta_1 / \theta_2 - \theta_1) \) and \( l_2 = 
(i - i_1 / i_2 - i_1) \). Implementing this method drastically min-
izes the computational burden of the scheduling process 
and the number of cycles required for scheduling.

IV. SIMULATION RESULTS
The Q-learning algorithm integrated with the bilinear 
scheduling approach has been simulated to study the per-
fomance of the proposed current controller and verify the 
effectiveness of the controller. The control scheme is depicted 
in Fig. 5. This controller has been applied to a 500 W 
12/8 SRM, which has a phase resistance of 2 \( \Omega \) and a nominal 
current of 5 A. The inductance profile of the controller begins 
from the aligned position at 16 mH and gradually decreases 
until it reaches the unaligned position at 6 mH. The available 
dc voltage is 100V. The simulation sampling time is defined 
by the switching frequency. In other words, for this simulation 
that the switching frequency is 10kHz, the sampling time is 
1/10kHz equal to 0.1 millisecond. The control cycle in 
which the Q-learning/hysteresis performs is also defined by 
the switching frequency. To smoothen the simulation results, 
the time step for electrical parts 0.1 of the control cycle (10 
 microseconds). Algorithm 2 has been utilized for training all 
Q-cores pre-located on the nonlinear surface of the machine. 
In this case, the algorithm should initialize the process using 
a stable control policy and an augmented state. The initial 
augmented state and initial control policy have been selected 
to be \( X_0 = [0\ 0]^T \) and \( K_0 = [100\ -100]^T \), respectively. 
The cost function has been applied with the weights of
Q = 100 and R = 0.001. The discount factor in this function is \( \gamma = 0.9 \).

The ratio between Q and R is essential for training the local Q matrix. If the value of R is large, an extremely high cost associated with the control input will occur, which prevents the linear quadratic tracker from tracking the reference. Moreover, if there is a huge Q/R ratio or R=0, the controller will track the reference in the first step due to a huge applied control input. This means that the duty cycles switch between two values, either 0% or 100%, which allow the controller to act as a delta-modulation controller that causes a remarkably high pulsation on the current. Therefore, selecting an effective Q/R ratio for tracking controller is of interest that permits the control input to vary freely as well as preventing the controller from tracking the reference from the first cycle. Hence, we have chosen the weights to be Q = 100 and R=0.001 as they are the best selection based on a design technique.

The reference model generates a train of square wave signals, the typical reference current for SRM. The Q-matrix at the unaligned position and a current of 4 A converges to its optimal values to allow tracking performance as shown below:

\[
G = \begin{bmatrix} 438100 & -253100 & 5729 \\ -253100 & 56630 & -2595 \\ 5729 & -2595 & 98.1 \end{bmatrix}
\] (24)

And, the optimal control gain K converges to

\[
K = [120 \quad -122]
\] (25)

The optimal values vary based on Q-core along with the states that are located in the domain of the system. For each Q-core, there are 6 data tubes collected per iteration to train the Q-matrices using the LS method. In this simulation, the speed of the SRM is constant and has been selected to be 60 RPM to demonstrate the result for the proposed controller. Fig. 6 shows how the SRM drive current tracks the reference of sequent pulses within a few time steps. Fig. 7 shows how the control gains K values that have converged to their optimal numbers change (considering the movement along the scheduling-table as well). The optimal voltage signal introduced to the motor to verify the best tracking performance is shown in Fig. 8. Fig. 9 illustrates the behavior of the current once the reference changed from 4A to 5.5A. This figure depicts that the Q-table starts re-learning as a result of any change in the reference current. Hence, even though one cycle requires currents of up to 18A, the system will soon learn the correct model (in a practical application, over currents will be eliminated using a supervisory hysteresis band).

V. EXPERIMENTAL RESULTS

In this section, the results and observations are presented to show the practical feasibility of the control method. The experimental components include a 3-phase 500 W 12/8 SRM, DC machine with DC power supply to control the field and hence loading of the machine as a mechanical load, H bridge converter, control board with a TI TMS320F28377D microcontroller, and a mixed domain oscilloscope. The experimental setup is shown in Fig. 10. The unaligned and aligned inductances for the machine used for validation are 6mH and 16mH, respectively, and the nominal current is 5 A per phase. The proposed Q-learning algorithm is implemented inside one of the two TMS320F28377D cores capable of operating at 200MFLOPS each. The available processing power in this controller is sufficient to control a 3-phase SRM at a 40 kHz control frequency.

To demonstrate the effectiveness of the proposed Q-learning technique compared to the conventional hysteresis controller, both techniques are applied, where the proposed scheduled Q-learning approach controls the first
phase, and the hysteresis technique controls the second and the third phase. The behavior of the current at different stages of the learning process is shown in Fig. 11. In this figure, the controller is set to run starting from the preloaded Q table to the point that the Q table is trained to the actual hardware online. In this figure, probe 1 shows the behavior of Phase A under the proposed control while probe 2 shows Phase B under the traditional Delta-modulation for comparison. During the learning process, when the Q matrices are not fully trained, the current tries to track the reference current (Fig. 11-a). The zoomed version of the current response is shown in Fig. 11-b. After a couple of cycles, the Q matrices are fully trained and the current can successfully track the reference current with almost no ripples on the current pulses (Fig. 11-c). Delta-modulation is not effective in minimizing the ripples for the current pulses. The Q-learning algorithm, once the Q-matrices are fully trained, are much more effective at minimizing the ripples for current pulses.

A. CHANGING THE REFERENCE CURRENT RESPONSE

In this test, the reference current is changed from 5.5 A to 4.5 A. The observations are illustrated in Fig. 12. This figure shows that the Q-matrices begin retraining once the reference current varies to 4.5 A. Fig. 12-a shows how the new reference is tracked. After a few cycles, the current tracks the reference effectively after the Q-matrices are fully trained (Fig. 12-b). When conventional delta-modulation is used, large ripples will be observed in phase current, and there is no way to mitigate them (Fig. 12-b).

To illustrate the novelty of the proposed technique, it can be seen that a model-free Q-learning control can regulate the current in only three cycles, which in general is less than 15 milliseconds. Moreover, the proposed technique reduces the ripples significantly compared to the conventional hysteresis
summation, a 200 MHz microcontroller can easily implement the algorithm. However, for more complicated reinforcement learning methods with multilayer neural network, forward computations, and back propagation training, alternative solutions are required, which are not the main concern of this article.

VI. CONCLUSION
The Q-learning scheduling algorithm for controlling the current of an SRM drive was studied in this article. By defining a Q-learning LQT, a table of Q-cores was generated to cover the nonlinear surface of the SRM model. Using this table, a scheduled Q-learning controller was derived, which is capable of controlling a nonlinear system, particularly, an SRM drive. Additionally, an online training mechanism was introduced capable of controlling the SRM without having any information regarding the system model parameters. This training mechanism updates each Q-core in the table as the state variables evolve over the domain of this table. Furthermore, a linear interpolation technique was used to ensure smooth transitions between these Q-cores. Lastly, simulation and experimental results demonstrated that the proposed algorithm is successful in controlling the current of a switched reluctance motor, minimizing its ripples, and adapting to the underlying SRM without any prior information regarding its parameters.

REFERENCES
[1] N. Yan, X. Cao, and Z. Deng, “Direct torque control for switched reluctance motor to obtain high Torque–Ampere ratio,” IEEE Trans. Ind. Electron., vol. 66, no. 7, pp. 5144–5152, Jul. 2019, doi: 10.1109/TIE.2018.2870355.
[2] Q. Sun, J. Wu, C. Gan, C. Shi, and J. Guo, “DSSRM design with multiple pole arc optimization for high torque and low torque ripple applications,” IEEE Access, vol. 6, pp. 27166–27175, 2018, doi: 10.1109/ACCESS.2018.2834901.
[3] M. Kawa, K. Kiyota, J. Furqani, and A. Chiba, “Acoustic noise reduction of a high-efficiency switched reluctance motor for hybrid electric vehicles with novel current waveform,” IEEE Trans. Ind. Appl., vol. 55, no. 3, pp. 2519–2528, May 2019, doi: 10.1109/TIA.2018.2888847.
[4] V. Valdivia, R. Todd, F. J. Bryan, A. Barrado, A. Lázaro, and A. J. Forsyth, “Behavioral modeling of a switched reluctance generator for aircraft power systems,” IEEE Trans. Ind. Electron., vol. 61, no. 6, pp. 2690–2699, Jun. 2014, doi: 10.1109/TIE.2013.2276768.
[5] H. Chen, W. Yan, J. J. Gu, and M. Sun, “Multiobjective optimization design of a switched reluctance motor for low-speed electric vehicles with a Taguchi–CSO algorithm,” IEEE/ASME Trans. Mechatronics, vol. 23, no. 4, pp. 1762–1774, Aug. 2018, doi: 10.1109/TMECH.2018.2839619.
[6] K. Kiyota and A. Chiba, “Design of switched reluctance motor competitive to 60-kW IPMSM in third-generation hybrid electric vehicle,” IEEE Trans. Ind. Appl., vol. 48, no. 6, pp. 2303–2309, Nov. 2012, doi: 10.1109/TIA.2012.2227091.
[7] M. Krishnamurthy, C. S. Edrington, A. Emadi, P. Asadi, M. Ehsani, and B. Fahimi, “Making the case for applications of switched reluctance motor technology in automotive products,” IEEE Trans. Power Electron., vol. 21, no. 3, pp. 659–675, May 2006, doi: 10.1109/TPEL.2006.872571.
[8] P. Shamsi and B. Fahimi, “Single-bus star-connected switched reluctance drive,” IEEE Trans. Power Electron., vol. 28, no. 12, pp. 5578–5587, Dec. 2013, doi: 10.1109/TPEL.2013.2251475.
[9] Y. Yang, N. Schofield, and A. Emadi, “Double-rotor switched reluctance machine (DRSRM),” IEEE Trans. Energy Convers., vol. 30, no. 2, pp. 671–680, Jun. 2015, doi: 10.1109/TEC.2014.2378211.
[10] W. Wang, M. Luo, E. Cosorobia, B. Fahimi, and M. Kiani, “Rotor shape investigation and optimization of double stator switched reluctance machine,” IEEE Trans. Magn., vol. 51, no. 3, Mar. 2015, Art. no. 8103304, doi: 10.1109/TMAG.2014.2356573.
[11] A. H. Isfahani and B. Fahimi, “Comparison of mechanical vibration between a double-stator switched reluctance machine and a conventional switched reluctance machine,” *IEEE Trans. Magn.*, vol. 50, no. 2, pp. 293–296, Feb. 2014, doi: 10.1109/TMAG.2013.2286569.

[12] R. Gobbi and K. Ramar, “Optimisation techniques for a hysteresis current controller to minimise torque ripple in switched reluctance motors,” *IET Elect. Power Appl.*, vol. 3, no. 5, pp. 453–460, Sep. 2009, doi: 10.1049/iet-epa.2008.0191.

[13] B. Shao and A. Emadi, “A digital PWM control for switched reluctance motor drives,” in *Proc. IEEE Vehicle Power Propuls. Conf.*, Sep. 2010, pp. 1–6, doi: 10.1109/VPPC.2010.5790103.

[14] S. E. Schulz and K. M. Rahman, “High-performance digital pi current regulator for ev switched reluctance motor drives,” *IEEE Trans. Ind. Appl.*, vol. 39, no. 4, pp. 1118–1126, Jul. 2003, doi: 10.1109/28.814580.

[15] F. Blaabjerg, P. C. Kjaer, P. O. Rasmussen, and C. Cossar, “Improved digital current control methods in switched reluctance motor drives,” *IEEE Trans. Power Electron.*, vol. 14, no. 3, pp. 563–572, May 1999, doi: 10.1109/63.761700.

[16] J. Ye, P. Malysz, and A. Emadi, “A fixed-switching-frequency integral sliding mode current controller for switched reluctance motor drives,” *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 3, no. 2, pp. 381–394, Jun. 2015, doi: 10.1109/JESTPE.2014.2357717.

[17] S. M. Lukic and A. Emadi, “State-switching control technique for switched reluctance motor drives: Theory and implementation,” *IEEE Trans. Ind. Electron.*, vol. 57, no. 9, pp. 2932–2938, Sep. 2010, doi: 10.1109/TIE.2009.2038942.

[18] A. Ali Akcayol, “Application of adaptive neuro-fuzzy controller for SRM,” *Adv. Eng. Sofw.*, vol. 35, nos. 3–4, pp. 129–137, Mar. 2004, doi: 10.1016/j.advengsoft.2004.03.005.

[19] X. Li and P. Shamsi, “Model predictive current control of switched reluctance motors with inductance auto-calibration,” *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3934–3941, Jun. 2016, doi: 10.1109/TIE.2015.2497301.

[20] R. Mikail, I. Husain, Y. Sozer, M. S. Islam, and T. Sebastian, “A fixed switching frequency predictive current control method for switched reluctance machines,” *IEEE Trans. Ind. Appl.*, vol. 50, no. 6, pp. 3717–3726, Nov. 2014, doi: 10.1109/TIA.2014.2322144.

[21] S. S. Ahmad and G. Narayanan, “Predictive control based constant current injection scheme for characterization of switched reluctance machine,” *IEEE Trans. Ind. Appl.*, vol. 54, no. 4, pp. 3383–3392, Jul. 2018, doi: 10.1109/TIA.2018.2818646.

[22] Z. Lin, D. Reay, B. Williams, and X. He, “High-performance current control for switched reluctance motors based on on-line estimated parameters,” *IET Elect. Power Appl.*, vol. 4, no. 1, pp. 67–74, Jan. 2010, doi: 10.1049/iet-epa.2009.0016.

[23] S. Saadatmand, S. Azizi, M. Kavousi, and D. Wunsch, “Autonomous control of a line follower robot using a Q-learning controller,” in *Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2020, pp. 566–561, doi: 10.1109/CCWC47524.2020.9331160.

[24] X. Li and P. Shamsi, “Inductance surface learning for model predictive current control of switched reluctance motors,” *IEEE Trans. Transp. Electrific.*, vol. 1, no. 3, pp. 287–297, Oct. 2015, doi: 10.1109/TTE.2015.2468178.

[25] P. J. Werbos, “Approximate dynamic programming for real-time control and neural modeling,” in *Handbook of Intelligent Control*, D. A. White and D. A. Sofge, Eds. New York, NY, USA: Reinhold, 1992.

[26] S. He, M. Zhang, H. Fang, F. Liu, X. Luan, and Z. Ding, “Reinforcement learning and adaptive optimization of a class of Markov jump systems with completely unknown dynamic information,” *Neural Comput. Appl.*, vol. 32, no. 18, pp. 14311–14320, Sep. 2020, doi: 10.1007/s00521-019-04180-2.

[27] S. He, H. Fang, M. Zhang, F. Liu, X. Luan, and Z. Ding, “Online policy iterative-based H∞ optimization algorithm for a class of nonlinear systems,” *Inf. Sci.*, vol. 495, pp. 1–13, Aug. 2019, doi: 10.1016/j.ins.2019.04.027.

[28] F. L. Lewis, D. L. Vrabie, and V. L. Syrmos, *Optimal Control*, 3rd ed. Hoboken, NJ, USA: Wiley, 2012.

[29] B. Kiumarsi, F. L. Lewis, H. Modares, A. Karimpour, and M.-B. Naghibi-Sistani, “Reinforcement learning for optimal tracking control of linear discrete-time systems with unknown dynamics,” *Automatica*, vol. 50, no. 4, pp. 1167–1175, Apr. 2014, doi: 10.1016/j.automatica.2014.02.015.
MEHDI FERDOWSI (Member, IEEE) received the B.S. degree in electronics from the University of Tehran, Tehran, Iran, in 1996, the M.S. degree in electronics from the Sharif University of Technology, Tehran, in 1999, and the Ph.D. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2004.

In 2004, he joined the Missouri University of Science and Technology, Rolla, MO, USA, as a Faculty, where he is currently a Professor with the Electrical and Computer Engineering Department. His research interests include power electronics, energy storage, smart grid, vehicular technology, and wide bandgap devices.

Dr. Ferdowsi was a recipient of the National Science Foundation CAREER Award, in 2007. He is an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS.

POURYA SHAMSI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2007, and the Ph.D. degree in electrical engineering from The University of Texas at Dallas, Richardson, TX, USA, in 2012.

He is currently the Woodard Associate Professor of electrical engineering with the Missouri University of Science and Technology (formerly UMR), Rolla, MO, USA. His research interests are power electronics, microgrids, reachability of hybrid systems, networked control systems, and motor drives.