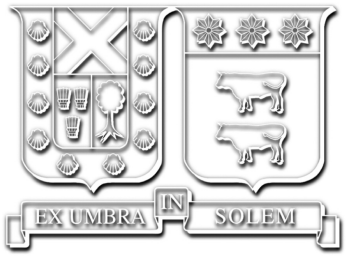
Universidad Técnica Federico Santa María
Departamento de Ingeniería Eléctrica
Valparaíso, Chile



Validación experimental del control de una máquina de inducción mediante inteligencia artificial basada en *Deep Q-Network*

Ignacio Enrique Araya Garcés

2026

Requisito parcial para obtener el título de:
Ingeniero Civil Electricista

Profesor Guía:
Dr. Ing. Margarita Norambuena Valdivia (UTFSM)

Comisión:
Dr. Ing. Margarita Norambuena Valdivia (UTFSM)
Dr. Ing. Jorge Juliet Avilés (UTFSM)

Valparaíso, Chile. Septiembre de 2026.

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Validación experimental del control de una máquina de inducción mediante inteligencia artificial basada en *Deep Q-Network*

Nombre del candidato(a): Ignacio Enrique Araya Garcés

Carrera / Grado: Ingeniería Civil Eléctrica / Ingeniero Civil Electricista

Campus: Casa Central Valparaíso **Departamento:** Ingeniería Eléctrica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Margarita Norambuena Valdivia, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

_____

_____

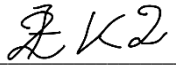
_____

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

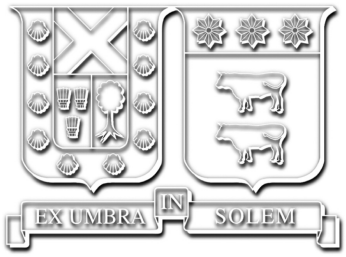
Fecha: 26.10.2026 **Firma:** 

Estudiante o Candidato(a):

Fecha: 23 de octubre de 2026 **Firma:** 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Universidad Técnica Federico Santa María
Departamento de Ingeniería Eléctrica
Valparaíso, Chile



Validación experimental del control de una máquina de inducción mediante inteligencia artificial basada en *Deep Q-Network*

Ignacio Enrique Araya Garcés

2026

"El corazón del hombre traza su camino,
pero el Señor dirige sus pasos."

— Proverbios 16:9

"Lo excelente es enemigo de lo bueno."

— Denis Araya Gutiérrez

"Encomienda al Señor tu camino;
confía en él y él actuará."

— Salmo 37:5

Agradecimientos

Gracias a Dios por acompañarme en este camino y por los padres que me dio. A ellos les agradezco el esfuerzo y los sacrificios que hicieron para que pudiera estudiar. De ustedes aprendí a enfrentar las dificultades y a seguir adelante, una enseñanza que me acompañó durante toda la carrera.

Papá, gracias por todo tu esfuerzo y por encontrar tiempo para compartir conmigo, incluso cuando era escaso. Gracias por hacerme reír, escucharme y tenerme paciencia cuando estaba frustrado.

Mamá, gracias por enseñarme a no rendirme. Te admiro por haber seguido adelante a pesar de las dificultades y por trabajar para que yo tuviera otras oportunidades. Siempre confiaste en mí, incluso cuando yo dudaba. Llegar hasta aquí también es una forma de demostrarte cuánto aprendí de ti y que, de haber tenido las mismas oportunidades, tú también habrías podido llegar hasta donde estoy.

A mi pareja, María Paz, gracias por estar a mi lado durante los momentos más difíciles de la carrera, los problemas de salud y las frustraciones. Tus abrazos y tus palabras me ayudaron a continuar. A tu familia, gracias por recibirme con cariño y hacerme sentir parte de ella. Gracias también por acercarme a Dios y ayudarme a encontrar en la fe tranquilidad y una manera distinta de mirar la vida. Desde entonces muchas cosas también comenzaron a cambiar para mejor.

A mis amigos de Iquique, gracias por las reuniones de verano, los chistes sin sentido y la amistad que hemos mantenido a pesar de la distancia. Volver a encontrarnos me ayudaba a dejar atrás las dificultades del año. Gracias por estar pendientes de mi regreso y por hacerme sentir siempre considerado.

A los amigos que encontré durante la carrera, gracias por acompañarme dentro y fuera de la universidad. A Francis y Carlos, por ayudarme y esperar cuando necesitaba más tiempo para avanzar. A Axel, el Nano y el Peter, por quedarse hasta la madrugada mientras terminaba los ensayos, incluso los fines de semana. Valoro también nuestras conversaciones y el tiempo que compartimos fuera de lo académico. Al grupo Blackout, gracias por considerarme siempre en sus actividades, por los campeonatos, las risas y la ayuda cuando la necesité.

Quiero recordar a quienes ya no están y fueron importantes en mi vida: a mi abuelo, que siempre quiso verme llegar hasta aquí; a Tito, Bruno, Nona, Luis Maureira, tío Vladi y Lala. También a mi perro Navidad, que me acompañó durante tantos años.

A la profesora Margarita Norambuena, gracias por confiar en mí, por su cercanía y por ayudarme a seguir avanzando cuando cometía errores. Agradezco su paciencia durante este proceso y la oportunidad de desarrollar un trabajo que pude presentar con orgullo en la Jornada Técnica del AC3E.

Al profesor Jorge Juliet, gracias por su vocación docente, por ayudarme incluso cuando no era alumno de sus cursos y por explicarme cada tema con paciencia, las veces que fuera necesario, hasta que lograra aprenderlo.

Al profesor Mel Grunert, gracias por ayudarme a fortalecer mi confianza y a transmitir de buena forma mis conocimientos. Muchas gracias por todo.

Al profesor Álvaro Varas, por el entusiasmo y la pasión con que enseña. A los profesores que confiaron en mí como ayudante y, en especial, al profesor Andrés Mora, por darme esa oportunidad en Máquinas Eléctricas. Asumir esa responsabilidad fue uno de mis primeros grandes desafíos y una experiencia muy valiosa en mi formación.

Al profesor Esteban Pulido, gracias por estar siempre dispuesto a ayudarme. A los profesores Gastón Fournies y Roger Schurch, por darme una segunda oportunidad cuando la necesité. A todos los profesores que formaron parte de mi carrera, gracias por contribuir a mi formación como ingeniero y como persona.

Finalmente, agradezco el apoyo de ANID a través del proyecto AC3E CIA250006, *Advanced Center for Electrical and Electronics Engineering*, y del proyecto ACT250059, en cuyo marco se desarrolló este trabajo.

Índice de contenidos

Índice de contenidos **I**

Índice de figuras **IV**

Índice de tablas **VI**

Nomenclatura **VIII**

Resumen **1**

1. Introducción **2**
1.1. Problema y contribución 2
1.2. Hipótesis y estrategia de transferencia 3
1.3. Objetivos 3
1.4. Alcance y organización 4

2. Fundamentos del accionamiento y del aprendizaje por refuerzo **5**
2.1. Dinámica de la máquina y orientación de campo 5
2.2. Tensiones de salida y modulación del inversor 7
2.3. Fundamentos del aprendizaje por refuerzo 8
2.4. Antecedentes y conexión con el método 12

3. Entorno y entrenamiento del agente DDQN **13**
3.1. Dinámica y secuencia de ejecución del entorno $R{-}L$ 14
3.2. Robustecimiento del entorno $R{-}L$ 15
3.3. Dominio final de entrenamiento 16
3.4. Observación, acción y recompensa 16
3.5. Agente, entrenamiento y selección 21
3.6. Alcance y límites de transferencia 25

4. Verificación sobre la carga $R{-}L$ **26**
4.1. Implementación y condiciones de ensayo 26
4.2. Verificación en bornes 27
4.3. Seguimiento de corriente en el marco dq 29
4.4. Respuesta transitoria 31
4.5. Resultado de la verificación 32

5. La referencia convencional **33**
5.1. El controlador y los criterios de ajuste 33
5.2. Ajuste del lazo de corriente 34
5.3. Ajuste de los lazos externos 35

5.4. Implementación de la referencia convencional 36
5.5. Alcance de la referencia convencional 37

6. Transferencia a la máquina de inducción 38
6.1. Integración del Agente 2 refinado en el IFOC 38
6.2. Comparación en simulación con el control convencional 42
6.3. Reflexión del eje de cuadratura 48
6.4. Resultado de la transferencia en simulación 50

7. Ensayos experimentales en la máquina de inducción 51
7.1. El banco, las maniobras y la procedencia de los registros 51
7.2. Resultados experimentales 53
7.3. Alcance de la validación experimental 64
7.4. Contraste con la simulación 65

8. Mejora del transitorio del agente 66
8.1. El problema inicial 66
8.2. La configuración final del agente 67
8.3. Formalización de las tres correcciones 67
8.4. Resultado conjunto a ±0,30 pu 70
8.5. Funcionamiento a 1050 rpm 78

9. Conclusiones 85
9.1. Respuesta a la hipótesis 85
9.2. Cumplimiento de los objetivos 85
9.3. Resultados principales 86
9.4. Limitaciones 87
9.5. Trabajo futuro 88

A. Parámetros y configuración de la plataforma 89
A.1. Banco experimental y cadena de señales 89
A.2. Parámetros de la máquina y bases de normalización 92
A.3. Configuración del entrenamiento 93
A.4. Ajuste de los controladores 102
A.5. La red dentro del procesador y su tiempo de inferencia 109
A.6. Cadena de actuación del agente 110
A.7. Tiempo muerto medido en el banco 112
A.8. La modulación en un período de portadora 113
A.9. Circuito de frenado 114

B. Validez de las campañas y límites experimentales 115
B.1. Cómo se miden las métricas dinámicas 115
B.2. Intervalos de medición durante las maniobras 116
B.3. Inicio de las maniobras en las capturas 119
B.4. Validez de captura: modo sombra y perfiles de monitorización 119
B.5. Comprobación de las mediciones externas 119
B.6. Tensión del *DC link* y circuito de frenado 121
B.7. Bajada de la corriente del eje directo 122
B.8. Perfil completo de las maniobras 122
B.9. Procedencia de las métricas 125

C. Trazabilidad de agentes y registros **126**
C.1. Cómo se identifica un agente 126
C.2. Procedencia de los registros 129
C.3. Material para reproducir y ampliar el trabajo 131

D. Comprobaciones complementarias **132**
D.1. Escalones de torque y rechazo de la perturbación 132
D.2. El Agente 2 con las perturbaciones compensadas 134
D.3. Barrido de la ganancia adaptativa en el gemelo 136
D.4. Secuencia de cada período de control 137
D.5. El umbral del primer paso grande 137

Bibliografía **139**

Índice de figuras

2.1. Estructura del control por campo orientado indirecto 6
2.2. Convertidor de dos niveles y hexágono de tensiones 7
2.3. Modulador PWM con inyección Min–Max 8
2.4. De la neurona biológica al perceptrón multicapa 10
2.5. Estructura común de DQN y DDQN 11
2.6. Valor objetivo en DQN y en DDQN 12
3.1. Arquitectura del agente DDQN 13
3.2. Acción incremental del agente 18
3.3. Por qué se acumulan los incrementos 19
3.4. Lazo de corriente del entorno $R-L$ con el agente 20
3.5. Evaluación periódica de los puntos de control 23
3.6. Obtención del Agente 2 24
3.7. Selección del Agente 2 refinado 25
4.1. Ensayos sobre la carga $R-L$ 27
4.2. Respuesta simulada en bornes 28
4.3. Respuesta experimental en bornes 28
4.4. Respuesta simulada en el marco dq 29
4.5. Respuesta experimental en el marco dq 30
5.1. Lazo cerrado de corriente de estator 34
5.2. Los dos lazos externos de la referencia convencional 35
6.1. Qué ve cada controlador 39
6.2. Esquema funcional del IFOC con el agente DDQN 41
6.3. La maniobra completa con el control convencional 42
6.4. La maniobra completa con el Agente 2 refinado 43
6.5. Transitorios de la Tabla 6.2 46
6.6. Errores de corriente en las ventanas de la Tabla 6.3 47
6.7. Inversión con y sin reflexión del eje q 48
7.1. Estación de trabajo de los ensayos sobre la máquina 52
7.2. Flujo de rotor estimado en las dos campañas 53
7.3. Velocidad mecánica en las dos campañas 54
7.4. Corriente del eje directo en las dos campañas 55
7.5. Corriente del eje de cuadratura en las dos campañas 56
7.6. Modulantes del marco sincrónico en las dos campañas 57
7.7. Corrientes de fase en bornes, en las dos campañas 58
7.8. Tensión fase–neutro en las dos campañas 59
7.9. Tensión del *DC link* en las dos campañas 60
7.10. Seguimiento transitorio de las referencias de corriente 62
8.1. Efecto medido de la corrección angular 68
8.2. La compensación del retardo en el banco 69

8.3. Seguimiento de la configuración final a ±0,30 pu 72
8.4. Flujo y velocidad a ±0,30 pu 73
8.5. Corrientes del marco a ±0,30 pu 74
8.6. Modulantes y *DC link* a ±0,30 pu 75
8.7. Corrientes de fase a ±0,30 pu 76
8.8. Tensión fase–neutro a ±0,30 pu 77
8.9. Flujo y velocidad a 1050 rpm 80
8.10. Corrientes del marco a 1050 rpm 81
8.11. Modulantes y *DC link* a 1050 rpm 82
8.12. Corrientes de fase de la configuración final a 1050 rpm 83
8.13. Tensión fase–neutro a 1050 rpm 84
A.1. Etapa de potencia del banco 89
A.2. Plataforma BRAIn y ejecución del lazo 90
A.3. Tarjeta de disparo por fibra óptica 91
A.4. Tarjeta acondicionadora del *encoder* 92
A.5. Verificación experimental de la sintonía teórica del lazo de velocidad . . . 106
A.6. Detalle del ciclo límite con la sintonía teórica 106
A.7. Lazos externos con el Agente 2 refinado 107
A.8. Ciclo del agente dentro del período de control 110
A.9. Un paso de actuación del agente 111
A.10. Tiempo muerto medido a la portadora de operación 112
A.11. La modulación en un período de portadora 113
B.1. Permanencia de i_{sd} durante la magnetización 116
B.2. Las ventanas de medida sobre las dos maniobras 117
B.3. Bajada del eje directo sobre la carga $R-L$ 122
B.4. Perfil completo sobre la carga $R-L$ 123
B.5. Perfil completo sobre la máquina 124
D.1. Escalones de torque en simulación 134
D.2. El Agente 2 con las perturbaciones compensadas 136
D.3. El inversor y la secuencia de cada período de control 138

Índice de tablas

3.1. Procedencia de los tres agentes 24
4.1. Parámetros de la carga $R-L$ 26
4.2. Condiciones principales de los ensayos sobre la carga $R-L$. 27
4.3. Error de seguimiento sobre la carga $R-L$ 31
4.4. Respuesta transitoria sobre la carga $R-L$ 32
6.1. Elementos añadidos alrededor del agente 40
6.2. Respuesta transitoria en simulación 44
6.3. Resultados de la maniobra simulada 45
6.4. Entradas y acción bajo la reflexión del eje de cuadratura 49
6.5. Efecto de la reflexión del eje de cuadratura 49
7.1. Condiciones de los ensayos experimentales. 52
7.2. Indicadores transitorios de flujo, velocidad y corriente. En i_{sq}, los tiempos y el error $e = i^* - i$ se calculan después de promediar, en cada instante, los valores de los 10 ms anteriores; el RMSE y el sesgo corresponden a los primeros 100 ms. Criterios: Sección B.1. 61
7.3. Error de corriente en los últimos 200 ms de cada registro, sin suavizar: RMSE, sesgo y desviación estándar del error $e = i^* - i$. 63
7.4. Contraste de indicadores seleccionados entre simulación y banco. El RMSE de corriente corresponde a los últimos 200 ms de cada registro. 65
8.1. Las tres correcciones de la configuración final 67
8.2. Resultado conjunto a ±0,30 pu 71
8.3. El ruido a mayor velocidad y el filtro de 25 Hz 78
8.4. Resultado a 1050 rpm 79
9.1. Objetivos, respaldo y alcance 86
A.1. Parámetros de la máquina de inducción 93
A.2. Configuración de entrenamiento de los tres agentes 94
A.3. Definición de las quince entradas en cada implementación 96
A.4. Constantes del entorno de entrenamiento 97
A.5. Sorteo de cada escenario del Agente 2 refinado 99
A.6. Tensión estimada para cada velocidad 100
A.7. Escenarios de selección del punto de control 101
A.8. Referencias de los escenarios de selección 101
A.9. Resultado del episodio 675 en los escenarios de selección 102
A.10. Configuración efectiva del entrenamiento, las simulaciones y el banco . . . 108
A.11. Ejecución del agente en la unidad de control 109
A.12. Cadena de actuación del agente 110
A.13. Ajuste del circuito de frenado 114
B.1. Seguimiento de corriente por tramo 117
B.2. Corriente media de cuadratura en los últimos 200 ms de cada registro . . 118
B.3. Contraste de las mediciones durante el estado estacionario del arranque . 120
B.4. Velocidad que usa cada parte del accionamiento 121
B.5. Procedencia de las métricas principales 125
C.1. Cadena de entrenamiento de los agentes 127

C.2. Archivos de pesos, modelos y ejecutables de los ensayos 128
C.3. Correspondencia del procedimiento con el programa 128
C.4. Versión del programa de entrenamiento de cada agente 129
C.5. Programas y versiones recuperadas 131
D.1. Procedencia de las comprobaciones complementarias 132
D.2. Escalones de torque en simulación 133
D.3. El Agente 2 compensado frente al Agente 2 refinado 135
D.4. Barrido de la ganancia adaptativa en el gemelo 137

Nomenclatura

abc Coordenadas trifásicas de fase.

$\alpha\beta$ Marco bifásico estacionario.

dq Marco sincrónico orientado al campo.

$\boldsymbol{x}_s$ Fasor espacial en negrita; módulo y componentes dq en cursiva.

i_d, i_q Corrientes de la carga $R{-}L$ en los ejes d y q.

i_{sd}, i_{sq} Corrientes de estator de la máquina en los ejes d y q.

i_d^*, i_q^* Referencias de corriente de la carga R–L.

i_{dq} Par de componentes d y q escrito de forma compacta.

i_{sd}^*, i_{sq}^* Referencias de corriente de estator.

e_d, e_q Errores de seguimiento de corriente.

m_d, m_q Modulantes en el marco dq.

$S_x, \bar{S}_x$ Estado de conmutación de la pierna x y su complementario.

$\Delta m_d, \Delta m_q$ Incrementos de modulante seleccionados por el agente.

V_{dc} Tensión del *DC link*.

R, L Resistencia e inductancia de la carga de entrenamiento.

R_s, R_r' Resistencias de estator y rotor referida al estator.

L_s, L_r', L_m Inductancias de estator, rotor referida al estator y magnetización.

ψ_r' Enlace de flujo de rotor referido al estator.

$v_{r,d}, v_{r,q}$ Contribuciones de tensión del flujo de rotor en cada eje.

$\sigma, \sigma L_s$ Coeficiente de dispersión total e inductancia transitoria.

τ_r Constante de tiempo del rotor.

τ_{eq} Constante de tiempo equivalente del lazo cerrado de corriente.

τ_{disp} Retardo del convertidor en el lazo del entorno $R{-}L$.

k_{RL}, τ_{RL} Ganancia y constante de tiempo de la carga $R{-}L$ por eje.

$\tau_{\text{ref}}, t_{\text{ref}}$ Tiempo normalizado y físico desde el cambio de referencia.

τ_Δ Ventana sin castigo al movimiento de la modulante.

i_{mr} Corriente de magnetización estimada.

ω_m, ω_s Velocidades mecánica y sincrónica.

ω Velocidad eléctrica del rotor; en el entorno $R{-}L$, la del marco.

ω_g, ω_{sl} Velocidades del marco genérico y de deslizamiento.

$\omega_{\text{pu}}, \omega_b$ Velocidad en por unidad y su base eléctrica; en el entorno $R{-}L$, la frecuencia angular del marco.

I_b, V_b Bases de corriente y tensión, de cresta.

T_e Torque electromagnético.

T_s, T_{sw} Períodos de actualización del controlador y de conmutación.

o_k, a_k, r_k Observación, acción y recompensa en el instante k.

$Q(o, a; \theta)$ Valor de acción aproximado por la red neuronal.

γ, ε Factor de descuento y probabilidad de exploración.

$\mathcal{C}, \Delta\mathcal{C}$ Índice de evaluación y costo sobre el seguimiento perfecto, en amperes.

pu Sistema por unidad.

FOC Control por campo orientado.

IFOC Control por campo orientado indirecto.

PI Controlador proporcional–integral.

OM Óptimo del módulo.

OS Óptimo simétrico.

OSR Óptimo simétrico realista.

PWM Modulación por ancho de pulso.

DQN *Deep Q-Network*.

DDQN *Double Deep Q-Network*.

RMS Raíz cuadrática media.

RMSE Raíz del error cuadrático medio.

Resumen

Esta memoria evalúa la transferencia de un agente *Double Deep Q-Network* (DDQN) al control de corriente de una máquina de inducción de 5,5 kW. El agente se entrena con una carga resistiva–inductiva ($R-L$) trifásica simulada y un conjunto representativo de parámetros; durante el refinamiento se añade una perturbación adicional de tensión en el eje de cuadratura. El entorno no incorpora las ecuaciones del flujo de rotor ni la dinámica mecánica de la máquina. La carga $R-L$ se controla con el Agente 2, y la máquina, en simulación y en el banco, con el Agente 2 refinado; en ambos casos, con los pesos fijos y sin entrenamiento adicional. Sí se ajustan los lazos externos del accionamiento, y la perturbación del refinamiento se dimensionó con información nominal de la máquina.

La red selecciona incrementos de las modulantes dq y se implementa en la plataforma BRAIn. En el entrenamiento, las decisiones se actualizan a 3 kHz sin representar explícitamente la conmutación. En el banco, el lazo de corriente también opera a 3 kHz, mientras que las señales de disparo del inversor se generan mediante PWM con inyección Min–Max y una portadora triangular de 3 kHz. El accionamiento conserva la orientación de campo, el estimador de flujo y los lazos externos de flujo y velocidad. En la máquina se compara su desempeño con el del control convencional, que usa controladores PI de corriente.

La verificación experimental sobre una carga $R-L$ muestra seguimiento de las referencias de corriente, con una raíz del error cuadrático medio (RMSE) de 164,4 mA en el eje directo y 264,7 mA en el de cuadratura. En simulación y en el banco de la máquina, el accionamiento completa la magnetización, el arranque hasta +0,30 pu y la inversión de marcha hasta −0,30 pu. Ambas estrategias alcanzan la referencia de flujo estimado de 1 pu después de la magnetización y siguen la referencia de velocidad, limitada por una rampa de 1 pu/s.

En los últimos 200 ms de cada maniobra sobre la máquina, el agente presenta mayor RMSE que el control convencional en ambos ejes. En i_{sd}, el RMSE queda entre 310,8 y 502,5 mA con el agente y entre 18,3 y 106,3 mA con el control convencional; en i_{sq}, entre 126,7 y 627,3 mA con el agente y entre 16,6 y 297,5 mA con el control convencional. El valor RMS de los seis sesgos es de 253 mA con el agente y 5,94 mA con el control convencional. Los lazos externos compensan parte de estas desviaciones ajustando las referencias de corriente. La evaluación de la red tarda 227,9 μs y el ciclo completo del agente dentro de la interrupción, 239,2 μs, frente a un período de control de 333,3 μs.

Los ensayos respaldan la transferencia desde el entorno $R-L$ al accionamiento con velocidad de ±0,30 pu y sin carga mecánica aplicada al eje. El *DC link* se ajustó a 300 V antes de la magnetización y, en estado estacionario, quedó entre 282 y 285 V con el control convencional y entre 281 y 284 V para el agente. Estos resultados muestran que una política aprendida en ese entorno simplificado puede utilizarse en el lazo de corriente de la máquina bajo las condiciones evaluadas. Con tres correcciones externas a la red, el ensayo se extendió hasta unos 1050 rpm. Allí el agente completó la magnetización, el arranque y la inversión de marcha, y su mayor error en estado estacionario apareció tras la inversión.

Capítulo 1

Introducción

1.1. Problema y contribución

La máquina de inducción es ampliamente utilizada en accionamientos eléctricos y, en aplicaciones que requieren un mayor desempeño dinámico, puede ser controlada mediante técnicas de control por campo orientado. Este tipo de control permite regular el comportamiento de la máquina mediante una estructura de lazos de control en la que el seguimiento de corriente constituye una etapa fundamental para la regulación del flujo de rotor y de la velocidad [1, 2].

En esta memoria se estudia una alternativa para realizar el seguimiento de corriente de dicho lazo interno mediante aprendizaje por refuerzo profundo. La pregunta principal es si un agente entrenado únicamente sobre una planta distinta y más simple puede conservar la capacidad de controlar corriente cuando se transfiere posteriormente a una máquina de inducción.

Para estudiar esta pregunta, el agente se entrena sobre una carga resistiva–inductiva ($R-L$) trifásica simulada. Desde el punto de vista del control, la relación entre la tensión aplicada y la corriente de esta carga puede representarse mediante una planta de primer orden, caracterizada por una ganancia y una constante de tiempo. Durante el entrenamiento se utilizan distintos valores de resistencia e inductancia para que el agente aprenda a controlar cargas $R-L$ con diferentes ganancias y constantes de tiempo. El refinamiento mantiene las ecuaciones de estado de la carga $R-L$ y añade una perturbación de tensión en el eje q, sin incorporar el modelo dinámico de la máquina. Una vez finalizado el aprendizaje, el agente se implementa primero sobre la carga $R-L$ experimental y posteriormente se transfiere al accionamiento de una máquina de inducción de 5,5 kW sin realizar entrenamiento adicional sobre ella.

La contribución de este trabajo consiste, por tanto, en evaluar experimentalmente si un agente entrenado sobre la planta de primer orden de una carga $R-L$ puede transferirse al control de corriente de una máquina de inducción, donde aparecen fenómenos que no estuvieron presentes durante el entrenamiento, entre ellos la dinámica del flujo de rotor, la tensión inducida y la dinámica mecánica. Como referencia, la estrategia se compara con el control convencional utilizado en el mismo accionamiento.

1.2. Hipótesis y estrategia de transferencia

El entrenamiento mediante aprendizaje por refuerzo requiere que el agente interactúe repetidamente con un entorno y explore distintas acciones antes de alcanzar un comportamiento de control adecuado [3, 4]. Aunque este proceso podría plantearse directamente sobre una máquina de inducción, en este trabajo el aprendizaje se realiza sobre una carga $R-L$ simulada, lo que permite estudiar inicialmente el problema de seguimiento de corriente en un entorno de menor complejidad.

La carga $R-L$ no se considera equivalente a la máquina de inducción. Su utilización responde a que permite representar, desde el punto de vista del control, una relación de primer orden entre tensión y corriente mediante una ganancia y una constante de tiempo. La máquina incorpora, además, fenómenos que no forman parte de este entorno de entrenamiento, asociados al flujo de rotor, a la tensión inducida y a la dinámica mecánica [1, 2].

A partir de esta diferencia se plantea la hipótesis central del trabajo: un agente entrenado para realizar seguimiento de corriente sobre un dominio de cargas $R-L$ puede conservar esa capacidad cuando se incorpora al lazo interno de una máquina de inducción, aun cuando durante el entrenamiento haya estado expuesto a la dinámica de la carga $R-L$ y no a la de la máquina. La transferencia se realiza sin reentrenar el agente sobre la máquina, en el sentido definido en la Sección 3.6, de manera que los ensayos posteriores permiten evaluar directamente si la capacidad de seguimiento aprendida puede conservarse en el accionamiento.

1.3. Objetivos

1.3.1. Objetivo general

Validar experimentalmente una estrategia de control basada en un agente *Deep Q-Network* para una máquina de inducción.

1.3.2. Objetivos específicos

1. Revisar los fundamentos del control por campo orientado (FOC) de máquinas de inducción, identificando el rol del flujo de rotor, las corrientes i_{sd} e i_{sq}, el torque electromagnético y la actuación del inversor para establecer las bases del control clásico de una máquina.
2. Implementar el esquema de control de referencia para la máquina de inducción, considerando el control por campo orientado, lazos de corriente, lazo de velocidad y generación de pulsos de disparo mediante modulación en PLECS.
3. Formular y diseñar un entorno de entrenamiento para la estrategia de control *Deep Q-Network* (DQN), basado en una carga *R–L* trifásica en coordenadas *dq*, utilizando parámetros representativos para controlar posteriormente una máquina de inducción sin reentrenar la red neuronal.
4. Implementar la estrategia de control diseñada, basada en DQN, en un entorno de simulación en PLECS, comprobando su desempeño en una carga $R-L$ y en una máquina de inducción.

5. Comparar el desempeño de la estrategia basada en DQN con el control clásico (FOC), considerando el seguimiento de corriente, la estabilidad del flujo de rotor, el torque electromagnético, la velocidad, el esfuerzo de conmutación y la calidad de las señales de disparo.
6. Validar experimentalmente el control diseñado en una carga $R-L$ y en una máquina de inducción.

Los objetivos corresponden a la propuesta de memoria aprobada, en la cual se utiliza la denominación DQN. La implementación final emplea DDQN, una variante de esta familia que separa la selección y la evaluación de la acción durante el aprendizaje. El alcance experimental de estos objetivos se establece a partir de las campañas y condiciones presentadas en esta memoria.

1.4. Alcance y organización

Esta memoria estudia el alcance de la transferencia de una red neuronal entrenada mediante aprendizaje por refuerzo profundo a una planta distinta de la utilizada durante el aprendizaje. La validación experimental se desarrolla en dos etapas. Primero, el agente se evalúa sobre la carga $R-L$ del banco mediante cuatro maniobras de seguimiento de corriente. Los resultados medidos se contrastan con los obtenidos en simulación, porque en esta etapa no se implementó un controlador PI experimental de referencia.

En segundo lugar, el agente se incorpora al lazo interno de corriente de la máquina de inducción. Se compara su desempeño con el control convencional mediante campañas experimentales que incluyen magnetización, arranque hasta +0,30 pu e inversión de marcha hasta −0,30 pu.

El trabajo no busca establecer una superioridad general frente al control convencional ni demostrar la optimalidad de la observación, la arquitectura de la red neuronal o el espacio de acciones utilizado. Los ensayos sobre la máquina se realizaron con las condiciones disponibles en el banco, sin carga mecánica aplicada al eje y con una tensión en el *DC link* inferior a la necesaria para operar cerca de la condición nominal. También se midió el tiempo de inferencia de la red neuronal para verificar que su ejecución se mantuviera dentro del período de control disponible. Por esta razón, las conclusiones se limitan a las condiciones ensayadas.

La memoria se organiza de la siguiente manera. El Capítulo 2 presenta los fundamentos necesarios para el desarrollo del trabajo; el Capítulo 3 define el entorno de entrenamiento, la observación, las acciones y el agente DDQN; el Capítulo 4 presenta la evaluación sobre la carga $R-L$; el Capítulo 5 establece el control convencional utilizado como referencia; el Capítulo 6 estudia la transferencia del agente a la máquina de inducción en simulación; el Capítulo 7 presenta la validación experimental sobre el banco; el Capítulo 8 mejora el transitorio del agente con tres correcciones fuera de la red y extiende el ensayo hasta unos 1050 rpm; y el Capítulo 9 reúne las conclusiones, las limitaciones del trabajo y las líneas de desarrollo futuro.

Capítulo 2

Fundamentos del accionamiento y del aprendizaje por refuerzo

Los controladores PI y el Agente 2 refinado se comparan como dos soluciones del lazo interno de corriente dentro de la misma estructura de campo orientado. Este capítulo presenta las ecuaciones de la máquina que determinan ese lazo, las tensiones que puede generar el inversor mediante la modulación PWM y los fundamentos del aprendizaje por refuerzo utilizados para definir la observación, la acción y la estructura de aprendizaje del agente. Se incluyen sólo las expresiones empleadas en el diseño y en la interpretación de los ensayos.

2.1. Dinámica de la máquina y orientación de campo

2.1.1. Modelo de la máquina en el marco sincrónico

Para un sistema trifásico equilibrado, el fasor espacial de una magnitud x se define como

$$\boldsymbol{x}_s = \frac{2}{3}\left(x_a + \underline{a}\, x_b + \underline{a}^2\, x_c\right), \qquad \underline{a} = e^{j2\pi/3}.$$

Los fasores espaciales se escriben en negrita; sus módulos y sus componentes en el marco dq, en cursiva normal. En un marco genérico que gira a velocidad eléctrica ω_g, el modelo de la máquina de inducción de jaula puede escribirse como [1, 2]

$$\begin{aligned} \boldsymbol{v}_s &= R_s \boldsymbol{i}_s + \frac{d\boldsymbol{\psi}_s}{dt} + j\omega_g \boldsymbol{\psi}_s, & 0 &= R'_r \boldsymbol{i}'_r + \frac{d\boldsymbol{\psi}'_r}{dt} + j(\omega_g - \omega)\boldsymbol{\psi}'_r, \\ \boldsymbol{\psi}_s &= L_s \boldsymbol{i}_s + L_m \boldsymbol{i}'_r, & \boldsymbol{\psi}'_r &= L_m \boldsymbol{i}_s + L'_r \boldsymbol{i}'_r, \end{aligned}$$

donde R_s y R'_r son las resistencias de estator y rotor referido al estator; L_s, L'_r y L_m son las inductancias propias y de magnetización; $\boldsymbol{\psi}_s$ y $\boldsymbol{\psi}'_r$ son los enlaces de flujo, y ω es la velocidad eléctrica del rotor. La inductancia transitoria asociada al lazo de corriente es σL_s, con

$$\sigma = 1 - \frac{L_m^2}{L_s L'_r}. \tag{2.1}$$

El marco sincrónico dq se obtiene rotando el fasor mediante

$$x_d + j x_q = \boldsymbol{x}_s e^{-j\varphi},$$

donde φ es el ángulo del marco. La transformación vuelve continuas las magnitudes sinusoidales que giran sincrónicamente y facilita su regulación; no garantiza por sí sola error nulo. En IFOC, el eje d se orienta con el flujo de rotor,

$$\psi'_{rd} = |\boldsymbol{\psi}'_r|, \qquad \psi'_{rq} = 0,$$

por lo que i_{sd} regula la magnetización e i_{sq} participa en la producción de torque. Bajo esta condición [1, 2],

$$T_e = \frac{3}{2} p \frac{L_m}{L'_r} |\boldsymbol{\psi}'_r| i_{sq} = \frac{3}{2} p \frac{L_m^2}{L'_r} i_{mr} i_{sq}, \tag{2.2}$$

$$J \frac{d\omega_m}{dt} = T_e - T_c - B\omega_m, \qquad \omega = p\omega_m, \tag{2.3}$$

donde p es el número de pares de polos, $i_{mr} = |\boldsymbol{\psi}'_r|/L_m$, J la inercia, B el coeficiente de fricción viscosa y T_c el torque de carga.

2.1.2. Estructura IFOC y estimador

En el control indirecto el ángulo de orientación no se mide. El modelo de corriente del rotor estima i_{mr} a partir de i_{sd} y obtiene el deslizamiento a partir de i_{sq}:

$$\tau_r \frac{di_{mr}}{dt} + i_{mr} = i_{sd}, \qquad \tau_r = \frac{L'_r}{R'_r}, \tag{2.4}$$

$$\omega_{sl} = \frac{i_{sq}}{\tau_r i_{mr}}, \qquad \hat{\varphi} = \int (\omega + \omega_{sl})\, dt, \qquad i_{mr} \neq 0.$$

La implementación debe proteger la división durante la inicialización. La Figura 2.1 muestra los lazos anidados: los lazos externos generan i^*_{sd} e i^*_{sq}, mientras el lazo interno genera la referencia de tensión de estator. El estimador, las transformaciones y la PWM no cambian; la sintonía de los lazos externos se declara en el Capítulo 5.

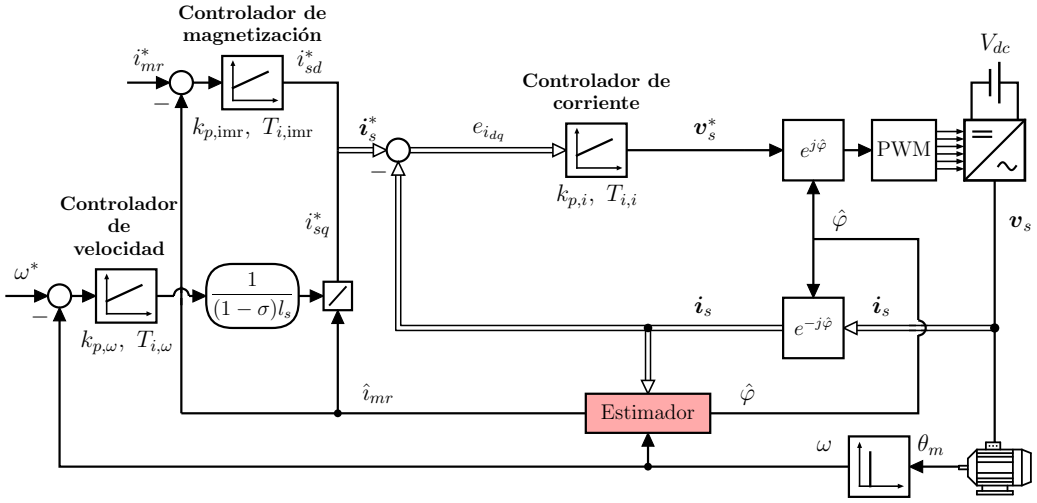
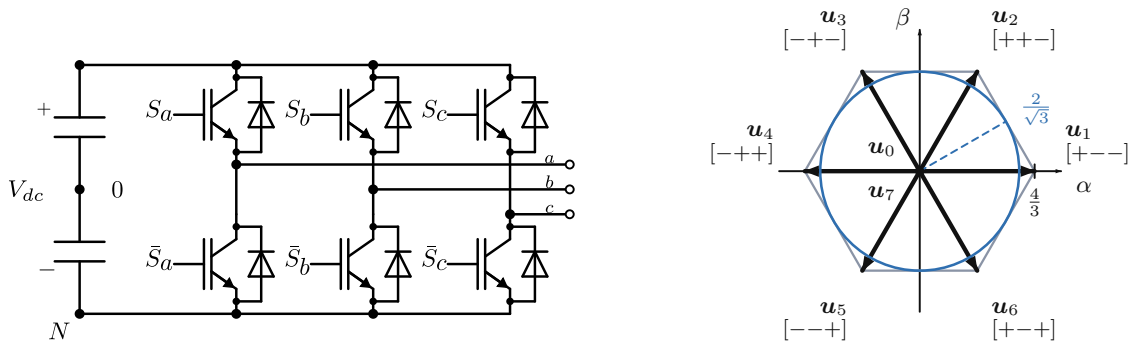


Figura 2.1: Estructura IFOC convencional, con un controlador PI en el lazo interno de corriente. Bajo cada controlador se indican su ganancia y su tiempo integral. Elaboración propia a partir de [1, 2].

2.2. Tensiones de salida y modulación del inversor

El controlador actualiza la referencia de tensión en cada período de control. Esta referencia se mantiene hasta la siguiente actualización, mientras el convertidor de dos niveles la reproduce mediante la modulación PWM. La Figura 2.2 muestra (a) las tres piernas del convertidor y (b) el hexágono normalizado de tensiones. Sus seis vértices corresponden a los estados activos; los dos estados restantes producen un fasor espacial de tensión nulo.



(a) (b)

Figura 2.2: Convertidor trifásico de dos niveles: (a) topología y estados complementarios en cada pierna; (b) hexágono normalizado de tensiones, con seis fasores espaciales activos, los fasores nulos $\boldsymbol{u}_0 = [---]$ y $\boldsymbol{u}_7 = [+++]$ en el origen, y la circunferencia inscrita de radio $2/\sqrt{3}$. Adaptada de [2, 5].

2.2.1. Estados y tensiones

Cada pierna se representa mediante $S_x \in \{0, 1\}$, $x \in \{a, b, c\}$. Su tensión de salida referida al punto medio del *DC link*, que en adelante se escribe v_{x0}, vale

$$u_x = 2S_x - 1 \in \{-1, +1\}, \qquad v_{x0} = \frac{V_{dc}}{2} u_x.$$

Esa misma tensión de salida, referida ahora al negativo del *DC link*, se escribe v_{xN} y difiere de la anterior en la constante $V_{dc}/2$. Para una carga equilibrada con neutro flotante, la tensión fase–neutro se obtiene mediante [2, 5]

$$\begin{aligned} v_{xN} &= S_x V_{dc}, \qquad v_{nN} = \frac{v_{aN} + v_{bN} + v_{cN}}{3}, \\ v_{an} &= v_{aN} - v_{nN}. \end{aligned} \tag{2.5}$$

Al recorrer los ocho estados, la tensión reconstruida toma cinco niveles,

$$v_{an} \in \left\{ 0,\ \pm\frac{V_{dc}}{3},\ \pm\frac{2V_{dc}}{3} \right\}. \tag{2.6}$$

2.2.2. PWM e inyección Min–Max

La PWM compara cada modulante con una portadora triangular. Mientras $|m_x| \le 1$, el valor medio de v_{x0} durante un período de conmutación es [5, 6]

$$\bar{v}_{x0} = \frac{1}{T_{\mathrm{sw}}} \int_t^{t+T_{\mathrm{sw}}} v_{x0}(\lambda)\, d\lambda = \frac{V_{dc}}{2} m_x, \qquad m_x = \frac{v_x^*}{\hat{v}_{\mathrm{mod}}}. \tag{2.7}$$

Aquí $\hat{v}_{\mathrm{mod}}$ es la amplitud de la portadora, en volt. Para aprovechar el hexágono del inversor se suma a las tres modulantes una componente común Min–Max,

$$m_0 = -\frac{\mathrm{mín}(m_a, m_b, m_c) + \mathrm{máx}(m_a, m_b, m_c)}{2}, \qquad m'_x = m_x + m_0,$$

esta inyección extiende la región lineal de modulación hasta

$$m_{\mathrm{máx}} = \frac{2}{\sqrt{3}} \simeq 1{,}1547,$$

mientras la modulante desplazada m'_x que llega al comparador permanece acotada por la unidad. El límite $m_{\mathrm{máx}} = 2/\sqrt{3}$ coincide con el radio de la circunferencia inscrita de la Figura 2.2(b). La Figura 2.3 muestra la incorporación de la componente Min–Max al modulador PWM.

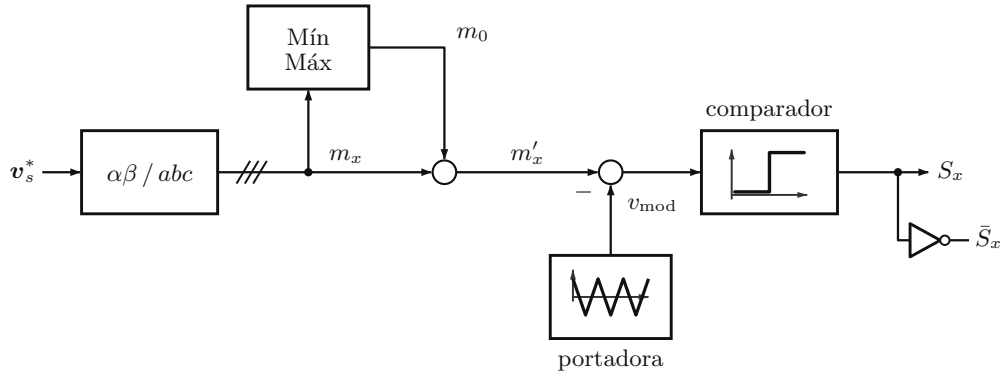


Figura 2.3: Modulador PWM basado en portadora con inyección Min–Max. La componente común m_0 se suma a cada modulante y m'_x se compara con la portadora para obtener el estado de conmutación S_x de la pierna y su complementario $\bar{S}_x$. Adaptada de [2].

En un sistema trifásico sin conexión de neutro y dentro de la zona lineal, la componente común m_0 no modifica las tensiones de línea ni las corrientes [5, 7].

2.2.3. Escalas de tiempo

El período de control T_s y el período de conmutación T_{sw} son magnitudes distintas; la implementación determina su relación. La tensión $\bar{v}_{x0}$ definida en (2.7) corresponde al valor promedio de v_{x0} durante un período de conmutación. La acción del agente es un incremento discreto de las modulantes y no la selección directa de un estado de conmutación, como en el control predictivo de conjunto finito [8, 9].

2.3. Fundamentos del aprendizaje por refuerzo

En aprendizaje por refuerzo, un agente interactúa con un entorno mediante acciones y recibe como respuesta una nueva observación y una recompensa. A partir de estas interacciones, el agente aprende una política que busca maximizar la recompensa acumulada [3].

2.3.1. MDP y *Q-learning*

Un proceso de decisión de Markov se describe mediante $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma)$, donde $\mathcal{S}$ representa el conjunto de estados, $\mathcal{A}$ el conjunto de acciones, $\mathcal{P}$ las transiciones del entorno, r la recompensa y γ el factor de descuento. En el paso k, el agente observa el estado s_k, selecciona una acción a_k, recibe una recompensa r_k y transita al estado s_{k+1}.

El objetivo es maximizar el retorno esperado,

$$G_k = \sum_{i=0}^{\infty} \gamma^i r_{k+i}, \qquad 0 \leq \gamma < 1. \tag{2.8}$$

La función óptima de valor de acción satisface [3]

$$Q^*(s,a) = \mathbb{E}\left[r + \gamma \max_{a'} Q^*(s',a') \mid s, a \right], \tag{2.9}$$

donde $\mathbb{E}[\cdot]$ representa el valor esperado sobre las posibles transiciones y recompensas del entorno.

Q-learning aproxima la función $Q^*(s,a)$ a partir de las transiciones obtenidas durante la interacción con el entorno. Para una transición, la actualización puede escribirse como

$$Q(s_k, a_k) \leftarrow Q(s_k, a_k) + \alpha \left[r_k + \gamma \max_{a'} Q(s_{k+1}, a') - Q(s_k, a_k) \right], \tag{2.10}$$

donde α es la tasa de aprendizaje y el término entre corchetes corresponde al error de diferencia temporal [10].

2.3.2. DQN y *Double DQN*

En la formulación del MDP, s representa el estado del sistema. En DQN, la red recibe una observación o, correspondiente al conjunto de variables disponibles que el agente utiliza para seleccionar una acción.

DQN reemplaza la representación tabular de la función Q por una red neuronal que entrega un valor $Q(o, a; \theta)$ para cada acción, donde θ representa los parámetros de la red en línea. Una memoria de repetición reduce la correlación temporal entre las muestras utilizadas durante el aprendizaje y una red objetivo, de parámetros θ^-, estabiliza el valor objetivo [4]. Para una transición (o, a, r, o'), el valor objetivo y la pérdida son

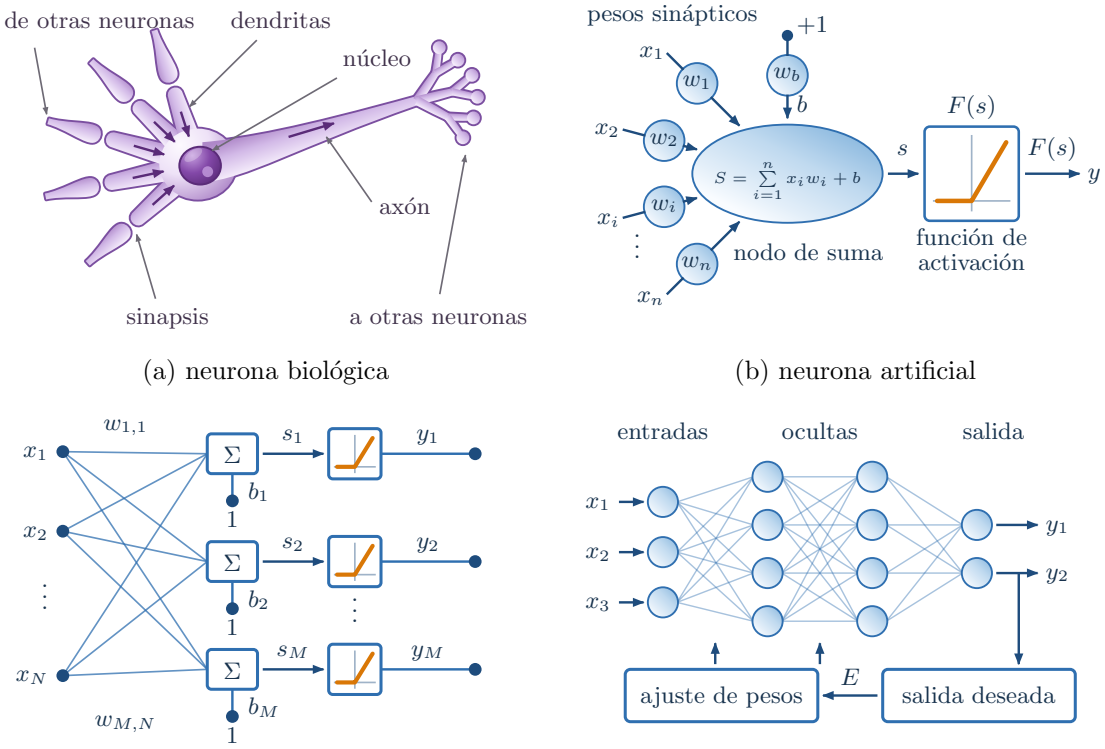
$$y^{\mathrm{DQN}} = r + \gamma \max_{a'} Q(o', a'; \theta^-), \qquad \mathcal{L}(\theta) = \mathbb{E}\left[\ell\left(y^{\mathrm{DQN}} - Q(o, a; \theta) \right) \right].$$

En este trabajo ℓ es la pérdida de Huber [11],

$$\ell_{\mathrm{H}}(\delta) = \begin{cases} \frac{1}{2}\,\delta^2, & |\delta| \leq \kappa, \\ \kappa\left(|\delta| - \frac{1}{2}\kappa \right), & |\delta| > \kappa, \end{cases} \tag{2.11}$$

donde δ es el residuo entre el valor objetivo y el valor estimado, y κ el umbral que separa sus dos tramos. La pérdida no está acotada, sino que crece linealmente fuera de ese umbral; lo acotado es su derivada respecto del residuo, lo que limita la influencia de errores grandes

durante la actualización de los pesos. La función Q se aproxima mediante un perceptrón multicapa. La Figura 2.4 muestra el paso desde la neurona artificial hasta esa estructura.



(a) neurona biológica

(b) neurona artificial

(c) capa totalmente conectada

(d) perceptrón multicapa

Figura 2.4: De la neurona biológica al perceptrón multicapa: (a) estructura biológica; (b) suma ponderada, sesgo y activación; (c) capa totalmente conectada; y (d) composición en capas. Elaborada en base a las figuras 1(a), 1(b), 3(a) y 4 de [12].

La neurona artificial del panel (b) calcula

$$z = \sum_{i=1}^{n} x_i w_i + b, \qquad y = F(z). \tag{2.12}$$

Una capa totalmente conectada agrupa varias neuronas descritas por (2.12); el perceptrón multicapa se obtiene al conectar varias de estas capas [12, 13]. Las capas ocultas emplean $\mathrm{ReLU}(z) = \max(0, z)$. Sin una activación no lineal, varias capas totalmente conectadas se reducirían a una sola transformación afín; ReLU permite aproximar Q mediante regiones afines. En su región activa su derivada es uno, por lo que no atenúa el gradiente por saturación, y evaluarla sólo requiere comparar con cero [14]. La salida es lineal y admite valores Q de ambos signos. Estas propiedades justifican el uso de ReLU, sin establecer que sea la mejor activación para este problema. La configuración completa se presenta en el Capítulo 3.

La diferencia entre DQN y DDQN se observa directamente en el valor objetivo. En DQN,

$$y^{\mathrm{DQN}} = r + \gamma \underbrace{\max_{a'} Q(o', a'; \theta^-)}_{\text{la red objetivo selecciona y evalúa}} .$$

En DDQN se separan ambas operaciones [15]:

$$\underbrace{a^* = \arg\max_{a'} Q(o', a'; \theta)}_{\text{la red en línea selecciona}}, \tag{2.13}$$

$$y^{\mathrm{DDQN}} = r + \gamma \underbrace{Q(o', a^*; \theta^-)}_{\text{la red objetivo evalúa}} . \tag{2.14}$$

Esta separación modifica sólo el cálculo del valor objetivo y busca reducir la sobreestimación de los valores Q; la memoria de repetición, la red objetivo y la política de exploración se conservan. La Figura 2.5 muestra la estructura compartida por ambos algoritmos, pero el máximo que dibuja corresponde a DQN; el Agente 2 refinado emplea el cálculo DDQN anterior.

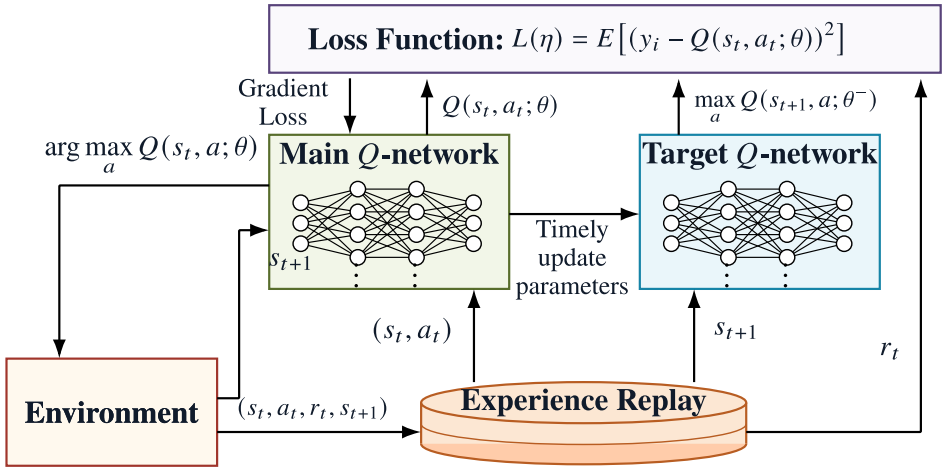


Figura 2.5: Estructura común de los lazos de decisión y aprendizaje de DQN y DDQN. El valor objetivo dibujado corresponde a DQN y no al utilizado por el Agente 2 refinado: en DDQN, la red en línea selecciona la acción y la red objetivo la evalúa, según (2.13)–(2.14). La fuente representa una pérdida cuadrática; en esta memoria se emplea la pérdida de Huber de (2.11). Reproducida de [16].

Ejemplo del valor objetivo. Sean $r = -0{,}40$, $\gamma = 0{,}99$ y tres acciones en el estado siguiente, con valores hipotéticos $Q(o', a'; \theta) = 2{,}0;\ 3{,}0;\ 1{,}5$ y $Q(o', a'; \theta^-) = 2{,}5;\ 2{,}2;\ 2{,}8$. La red en línea selecciona $a^* = 2$ y la red objetivo la evalúa, de modo que $y^{\mathrm{DDQN}} = -0{,}40 + 0{,}99 \cdot 2{,}2 = 1{,}778$. DQN usa el máximo de la red objetivo, 2,8, de la acción 3, y obtiene $y^{\mathrm{DQN}} = -0{,}40 + 0{,}99 \cdot 2{,}8 = 2{,}372$. El máximo tiende a escoger los errores de estimación positivos, de modo que separar la selección de la evaluación reduce esa sobreestimación.

La Figura 2.6 compara los dos valores objetivo y la actualización común del Algoritmo A.1.

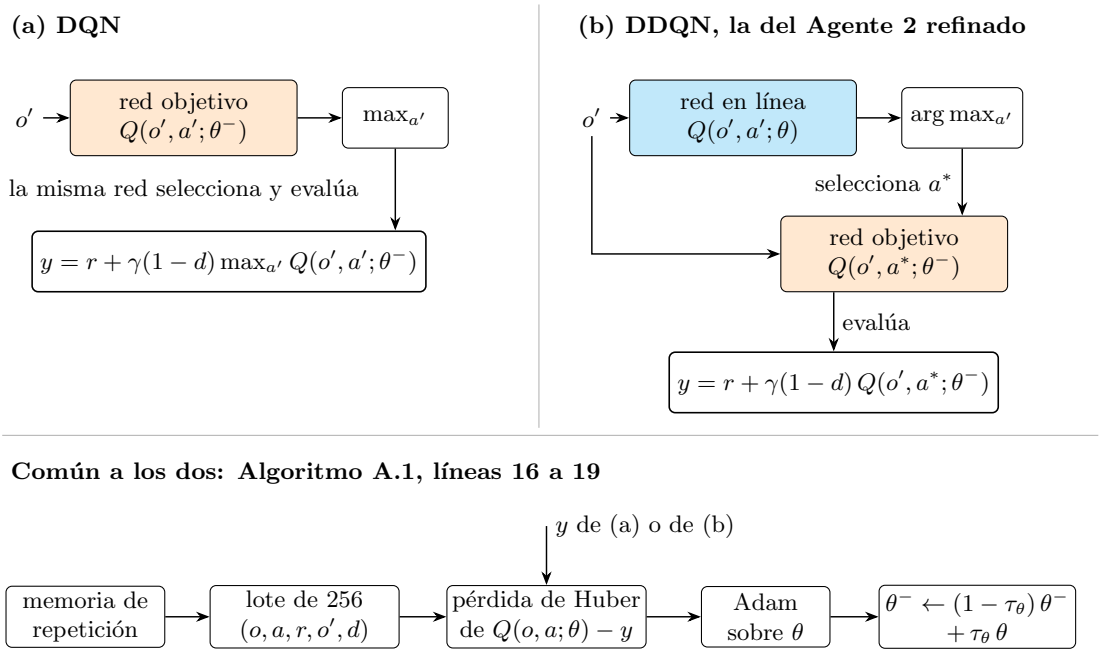


Figura 2.6: Valor objetivo en DQN y en DDQN. (a) En DQN, la red objetivo selecciona y evalúa la acción del estado siguiente. (b) En DDQN, la red en línea la selecciona y la red objetivo la evalúa, según (2.13) y (2.14). Abajo, la actualización común, que corresponde a las líneas 16 a 19 del Algoritmo A.1. Elaboración propia.

2.4. Antecedentes y conexión con el método

Las redes neuronales tienen antecedentes en electrónica de potencia y accionamientos [12], incluidos convertidores [13].

El aprendizaje por refuerzo profundo también se ha aplicado directamente al control de accionamientos. Schenke y coautores [17] lo propusieron como método de diseño de controladores para accionamientos eléctricos, y Traue y coautores [18] publicaron un entorno de simulación de motores para entrenar y comparar agentes. Schenke y Wallscheid [19] aplicaron *deep Q-learning* al control directo de torque de una máquina sincrónica de imanes permanentes, con acciones iguales a los estados de conmutación de un inversor de dos niveles. Book y coautores [20] llevaron a un banco de ensayos el aprendizaje *online*, en el que el agente aprende mientras controla el motor, y señalaron como dificultades el tiempo real, la seguridad, los retardos y los límites del *hardware* embebido.

Kaemling [21] es el antecedente directo: formuló el seguimiento de corriente mediante *Q-learning* y DQN con acciones de conmutación. Aquí el Agente 2 refinado actúa mediante incrementos de las modulantes dq y conserva la PWM. La política se aprende *offline*, en simulación, sobre una carga R–L y se refina en ese entorno con una perturbación de tensión en q. Después se transfiere al accionamiento sin reentrenar, a diferencia del aprendizaje *online* de [20].

Capítulo 3

Entorno y entrenamiento del agente DDQN

Este capítulo describe el entorno con que se entrena al agente DDQN. Representa una carga $R-L$ en coordenadas dq y, en su configuración final, incorpora una exigencia adicional de tensión en el eje q sin introducir un modelo de la máquina de inducción. La transferencia de la política a la máquina se verifica por separado. La Figura 3.1 resume el procedimiento completo.

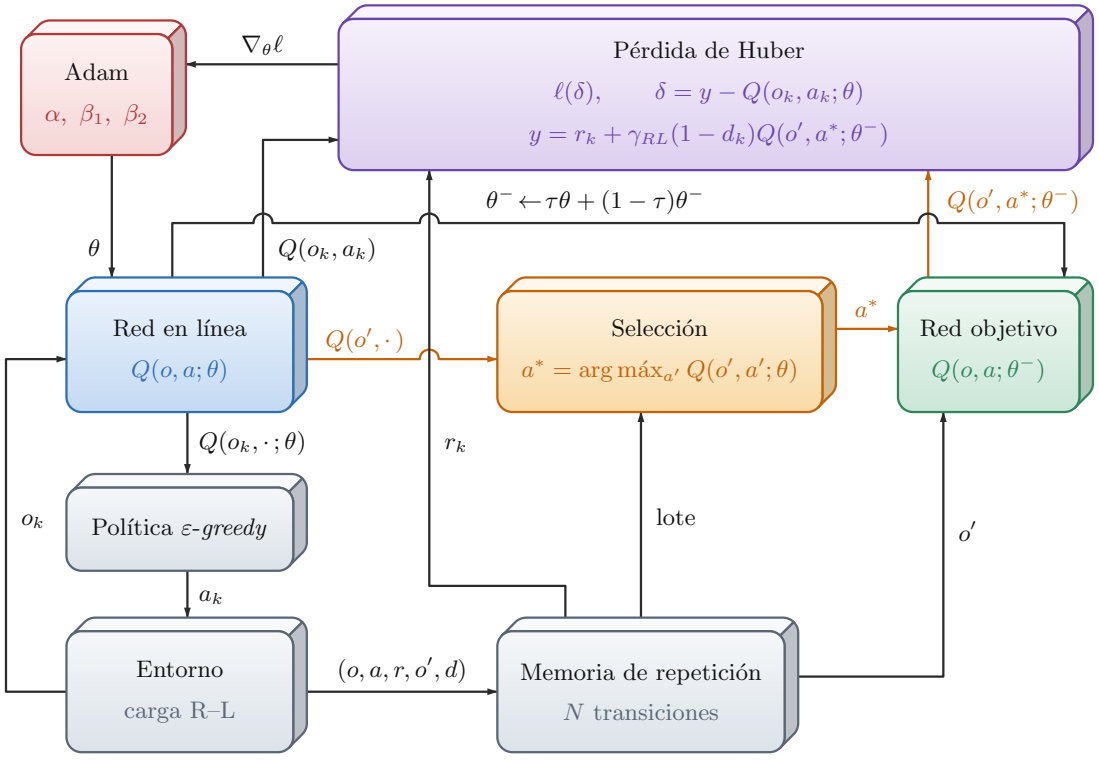


Figura 3.1: Arquitectura del agente DDQN utilizada durante el entrenamiento. Elaboración propia.

3.1. Dinámica y secuencia de ejecución del entorno $R-L$

Cada fase de la carga R–L tiene una relación de primer orden entre tensión y corriente. Para el entrenamiento se expresa en coordenadas dq, donde su modelo continuo es

$$v_d = Ri_d + L\frac{\mathrm{d}i_d}{\mathrm{d}t} - \omega L i_q, \qquad v_q = Ri_q + L\frac{\mathrm{d}i_q}{\mathrm{d}t} + \omega L i_d. \tag{3.1}$$

Definiendo las variables de estado

$$\boldsymbol{x} = \begin{bmatrix} i_d \\ i_q \end{bmatrix}, \qquad \boldsymbol{v} = \begin{bmatrix} v_d \\ v_q \end{bmatrix}, \tag{3.2}$$

la dinámica puede escribirse como

$$\dot{\boldsymbol{x}} = \underbrace{\begin{bmatrix} -\dfrac{R}{L} & \omega \\ -\omega & -\dfrac{R}{L} \end{bmatrix}}_{\mathbf{A}} \boldsymbol{x} + \underbrace{\frac{1}{L}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{B}} \boldsymbol{v}. \tag{3.3}$$

Los términos con ω representan el acoplamiento entre los ejes d y q introducido por el cambio de coordenadas. Cuando $\omega = 0$, ambos ejes quedan desacoplados.

Durante cada episodio de entrenamiento, R, L y ω se mantienen constantes. Además, la tensión aplicada se considera constante durante cada intervalo de muestreo T_s. Bajo esta hipótesis, la dinámica se discretiza mediante un retenedor de orden cero (*Zero-Order Hold*, ZOH), obteniéndose

$$\boldsymbol{x}_{k+1} = \mathbf{A}_d \boldsymbol{x}_k + \mathbf{B}_d \boldsymbol{v}_{\mathrm{ef},k}, \tag{3.4}$$

donde

$$\mathbf{A}_d = e^{\mathbf{A}T_s}, \qquad \mathbf{B}_d = \int_0^{T_s} e^{\mathbf{A}\tau}\mathbf{B}\,\mathrm{d}\tau. \tag{3.5}$$

La tensión efectiva utilizada para actualizar la carga se expresa como

$$\boldsymbol{v}_{\mathrm{ef},k} = \boldsymbol{v}_{k-1} + \boldsymbol{v}_{\mathrm{dt},k} + p_{q,k}\begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{3.6}$$

El término $\boldsymbol{v}_{k-1}$ aparece porque la acción calculada a partir de la muestra k se aplica en la actualización siguiente; por tanto, durante el intervalo $[t_k, t_{k+1})$ la carga evoluciona con la tensión correspondiente al período anterior.

El término de tiempo muerto se construye a partir del error promedio por pierna en PWM [22]. Al conservar su componente fundamental y expresarla en dq se obtiene la forma de primer armónico de [23, ec. (3.34)]; la deducción se desarrolla en el Apéndice A.3.1.

$$\boldsymbol{v}_{\mathrm{dt},k} = -\frac{4}{\pi}\frac{T_d}{T_{\mathrm{sw}}} V_{dc}\,\rho(\|\boldsymbol{i}_k\|)\frac{\boldsymbol{i}_k}{\|\boldsymbol{i}_k\|}, \qquad \rho(I) = \mathrm{mín}\left(1, \frac{I}{I_{\mathrm{zc}}}\right), \tag{3.7}$$

para $\|\boldsymbol{i}_k\| > 0$; en corriente nula se toma $\boldsymbol{v}_{\mathrm{dt},k} = \mathbf{0}$. Se define $\boldsymbol{v}_{\mathrm{dt},k} = \boldsymbol{v}_{\mathrm{ap},k} - \boldsymbol{v}_{\mathrm{ord},k}$; por tanto, $\boldsymbol{v}_{\mathrm{ap},k} = \boldsymbol{v}_{\mathrm{ord},k} + \boldsymbol{v}_{\mathrm{dt},k}$ y el término se suma en (3.6). Su signo negativo representa la reducción de tensión en la dirección de la corriente. La transición ρ suaviza el cruce por cero. Se utilizan $T_d = 1{,}0$ μs, $T_{\mathrm{sw}} = 1/3000$ s e $I_{\mathrm{zc}} = 0{,}05 I_b$; T_{sw} es el período de la

portadora PWM, mientras T_s es el período con que el agente observa y decide. En este trabajo ambos valen 1/3000 s, pero representan operaciones distintas. El primer valor es consistente con los 1,003 μs medidos en el Apéndice A.7. El entorno adopta $t_{\text{on}} = t_{\text{off}}$ y omite las caídas de conducción; bajo estas hipótesis, (3.7) representa la componente fundamental aproximada, no cada conmutación. La perturbación $p_{q,k}$ se añade sólo en q y ninguno de los dos términos modifica las matrices $\mathbf{A}$ y $\mathbf{B}$ de la carga $R-L$.

Antes de construir la observación, las corrientes se perturban con ruido gaussiano y se cuantizan. Luego el agente selecciona la acción, actualiza las modulantes y la actuación se aplica en el período siguiente.

3.2. Robustecimiento del entorno $R-L$

La carga $R-L$ reproduce la dinámica resistiva–inductiva y el acoplamiento entre los ejes d y q, pero no contiene la tensión inducida por el flujo de rotor de la máquina. En un sistema orientado al flujo, la magnitud de esta contribución puede expresarse por ejes como

$$|v_{r,d}| = \frac{k_r}{\tau_r}|\psi'_r|, \qquad |v_{r,q}| = k_r|\omega||\psi'_r|, \tag{3.8}$$

donde $v_{r,d}$ y $v_{r,q}$ son las contribuciones de tensión de cada eje, k_r el factor de acoplamiento, τ_r la constante de tiempo del rotor, $|\psi'_r|$ el módulo del enlace de flujo de rotor referido al estator y, en estas expresiones, ω corresponde a la velocidad angular eléctrica del rotor [1,2].

La principal diferencia aparece en el eje q. Mientras la contribución del eje d no aumenta con la velocidad del rotor, la del eje q crece proporcionalmente con su magnitud. En efecto,

$$\frac{|v_{r,q}|}{|v_{r,d}|} = |\omega|\tau_r. \tag{3.9}$$

Para la máquina utilizada, $\tau_r = 193{,}39$ ms. Considerando $|\omega| = 0{,}30\,\omega_b$, con $\omega_b = 2\pi 50$ rad/s, esta relación resulta aproximadamente igual a 18. Por tanto, al aumentar la velocidad del rotor, el eje q presenta una exigencia de tensión considerablemente mayor que la representada por la carga $R-L$ utilizada inicialmente durante el entrenamiento.

Esta diferencia motivó el refinamiento del entorno. En lugar de incorporar el modelo de la máquina, se mantuvo la dinámica $R-L$ y se añadió p_q a la tensión efectiva del eje q. Conservando el error de tiempo muerto y omitiendo sólo los índices para facilitar la lectura, (3.6) da $v_{q,\text{ef}} = v_q + v_{\text{dt},q} + p_q$. Sustituyendo en (3.1) y despejando la tensión ordenada al inversor se obtiene

$$\begin{aligned} v_q + v_{\text{dt},q} + p_q &= Ri_q + L\frac{\mathrm{d}i_q}{\mathrm{d}t} + \omega L i_d, \\ v_q &= Ri_q + L\frac{\mathrm{d}i_q}{\mathrm{d}t} + \omega L i_d - v_{\text{dt},q} - p_q. \end{aligned} \tag{3.10}$$

Aquí ω es la frecuencia angular del marco dq, como en (3.1). El agente debe seguir la referencia de corriente frente a esta exigencia adicional de tensión.

Para la misma corriente, su derivada y el mismo error de tiempo muerto, la diferencia respecto del caso $p_q = 0$ es $\Delta v_q = v_q(p_q) - v_q(0) = -p_q$. Por ello, un p_q negativo exige una tensión ordenada v_q mayor. El refinamiento conserva las matrices $\mathbf{A}$ y $\mathbf{B}$ de la carga y añade la perturbación a la entrada del eje q. El origen de su escala se desarrolla en el Apéndice A.3.

3.3. Dominio final de entrenamiento

El entrenamiento utiliza distintos valores de resistencia e inductancia para cubrir cargas $R-L$ con diferentes ganancias y constantes de tiempo. También varían la tensión del *DC link* y la frecuencia del marco dq. Incluye tres condiciones: magnetización con $\omega = 0$, el punto de operación de la carga $R-L$ del banco y un muestreo general con frecuencias de ambos signos y perturbación en q. Los intervalos y probabilidades se detallan en el Apéndice A.3.

En el muestreo general, p_q se añade sólo a la tensión del eje q. Su amplitud aumenta con la frecuencia del marco y con la tensión del *DC link*, e incorpora una variación acotada entre episodios. Con el marco detenido, $p_q = 0$; al invertir el signo de su frecuencia, también cambia el signo de la perturbación. Este refinamiento mantiene la planta $R-L$: no integra el flujo de rotor ni modifica la inductancia de la carga. El sorteo, los límites y el cálculo de la escala se desarrollan en el Apéndice A.3.1.

3.4. Observación, acción y recompensa

3.4.1. Observación

Las corrientes y los errores se expresan en por unidad de $I_b = 15{,}6$ A. La observación entregada al agente contiene quince componentes, agrupadas según la información que aportan:

$$\boldsymbol{o}_k = \Big[\underbrace{e_{d,k},\, e_{q,k},\, \Delta e_{d,k},\, \Delta e_{q,k}}_{\text{error y tendencia}},\, \underbrace{i^*_{d,k},\, i^*_{q,k},\, \omega_{\mathrm{pu},k}}_{\text{punto de operación}},$$

$$\underbrace{m_{d,k-1},\, m_{q,k-1},\, \sin\theta_{M,k-1},\, \cos\theta_{M,k-1}}_{\text{memoria de actuación}},$$

$$\underbrace{\bar{e}_{d,k},\, \bar{e}_{q,k},\, \tau_{\mathrm{ref},k}}_{\text{memoria temporal}},\, \underbrace{V_{dc,k}/V_{dc,\mathrm{ref}}}_{\text{margen de tensión}} \Big]^{\mathsf{T}}.$$

El primer grupo informa el error instantáneo de seguimiento en ambos ejes y su variación entre muestras. Los términos e_d y e_q indican cuánto falta para alcanzar la referencia y en qué sentido debe realizarse la corrección, mientras que Δe_d y Δe_q permiten distinguir si el error está disminuyendo o aumentando. La diferencia es $\Delta e_{x,k} = e_{x,k} - e_{x,k-1}$ en cada eje.

El segundo grupo define la condición en que está operando el entorno. Las referencias i^*_d e i^*_q indican las corrientes que se desean seguir, mientras que ω_{pu} informa la frecuencia angular normalizada del sistema de referencia dq. Como $e_x = i^*_x - i_x$, la corriente medida puede recuperarse mediante $i_x = i^*_x - e_x$ y no es necesario incorporarla nuevamente como una entrada independiente.

El tercer grupo proporciona una memoria explícita de la actuación previa. Debido a que la acción del agente corresponde a un incremento de modulante, conocer $m_{d,k-1}$ y $m_{q,k-1}$ permite saber desde qué valor acumulado se aplicará la nueva corrección. La orientación de esa modulante se incorpora mediante $\theta_{M,k-1} = \operatorname{atan2}(m_{q,k-1}, m_{d,k-1})$, representada por su

seno y coseno para evitar la discontinuidad angular entre $-\pi$ y π. Si las dos modulantes son nulas, el entrenamiento toma $\theta_{M,k-1} = 0$, y la simulación y el banco conservan el ángulo anterior.

El cuarto grupo entrega información temporal que no está contenida en el error instantáneo. Las variables $\bar{e}_d$ y $\bar{e}_q$ corresponden a versiones filtradas de los errores absolutos y permiten identificar si un error es momentáneo o se ha mantenido durante varias muestras. Se calculan con $\bar{e}_{x,k} = 0{,}98\,\bar{e}_{x,k-1} + 0{,}02\,|e_{x,k}|$, desde cero, y su constante de tiempo es de 16,5 ms. La entrada temporal es $\tau_{\mathrm{ref},k} = \mathrm{mín}(t_{\mathrm{ref},k}/\tau_\Delta, 1)$, con $t_{\mathrm{ref},k} = n_{\mathrm{ref},k} T_s$ en segundos y $\tau_\Delta = 20$ ms. Así, la red recibe un tiempo normalizado entre cero y uno desde el último cambio de referencia. En el entrenamiento, $n_{\mathrm{ref},k}$ vuelve a cero con cualquier cambio de referencia. En la simulación y en el banco vuelve a cero sólo cuando la variación filtrada de la referencia alcanza 0,05 pu. En la simulación de la máquina eso no ocurrió, y τ_{ref} permaneció en uno.

Finalmente, $V_{dc}/V_{dc,\mathrm{ref}}$ informa al agente del margen de tensión disponible. Esta entrada resulta necesaria porque una misma corrección de corriente puede requerir una acción distinta cuando cambia la tensión del *DC link*.

Las constantes utilizadas en los filtros y en las normalizaciones se presentan en el Apéndice A.3. La Tabla A.3 define cada entrada en el entrenamiento, la simulación y el banco.

3.4.2. Acción incremental

La red no entrega directamente una tensión ni un valor final de modulante. En cada muestra selecciona un incremento para las modulantes de los ejes d y q. En cada eje se dispone de siete incrementos positivos, sus correspondientes valores negativos y la acción nula,

$$\mathcal{D} = \{-\Delta m_7, \ldots, -\Delta m_1,\, 0,\, \Delta m_1, \ldots, \Delta m_7\}\,. \tag{3.11}$$

Por tanto, cada eje dispone de 15 posibilidades.

La acción seleccionada combina un incremento para cada eje,

$$\Delta \boldsymbol{m}_{dq}(k) = \begin{bmatrix} \Delta m_d(k) \\ \Delta m_q(k) \end{bmatrix}, \qquad \Delta m_d(k),\, \Delta m_q(k) \in \mathcal{D}. \tag{3.12}$$

De esta forma, las 15 posibilidades de cada eje generan $15 \times 15 = 225$ acciones discretas. La red posee una salida por cada acción y selecciona el par de incrementos asociado al mayor valor Q.

Los incrementos positivos se distribuyen geométricamente entre un valor mínimo y uno máximo,

$$\Delta m_j = \Delta m_{\mathrm{mín}}\, g_m^{j-1}, \qquad g_m = \left(\frac{\Delta m_{\mathrm{máx}}}{\Delta m_{\mathrm{mín}}}\right)^{1/6}, \qquad j = 1, \ldots, 7. \tag{3.13}$$

Los pasos pequeños permiten realizar correcciones finas de la actuación, mientras que los pasos mayores permiten modificarla con mayor rapidez cuando el error es elevado. El criterio utilizado para dimensionar $\Delta m_{\mathrm{mín}}$ y $\Delta m_{\mathrm{máx}}$ se presenta en el Apéndice A.3.

El incremento seleccionado no se aplica directamente como una tensión. Se acumula sobre la modulante utilizada en el período anterior,

$$\boldsymbol{m}_{dq}(k) = \mathrm{sat}_{\pm m_{\mathrm{lím}}} \Big[\underbrace{\boldsymbol{m}_{dq}(k-1)}_{\text{memoria de actuación}} + \underbrace{\Delta \boldsymbol{m}_{dq}(k)}_{\text{acción seleccionada}} \Big], \tag{3.14}$$

donde

$$\boldsymbol{m}_{dq}(k) = \begin{bmatrix} m_d(k) \\ m_q(k) \end{bmatrix}. \tag{3.15}$$

La Figura 3.2 representa esa selección y la acumulación de la acción escogida.

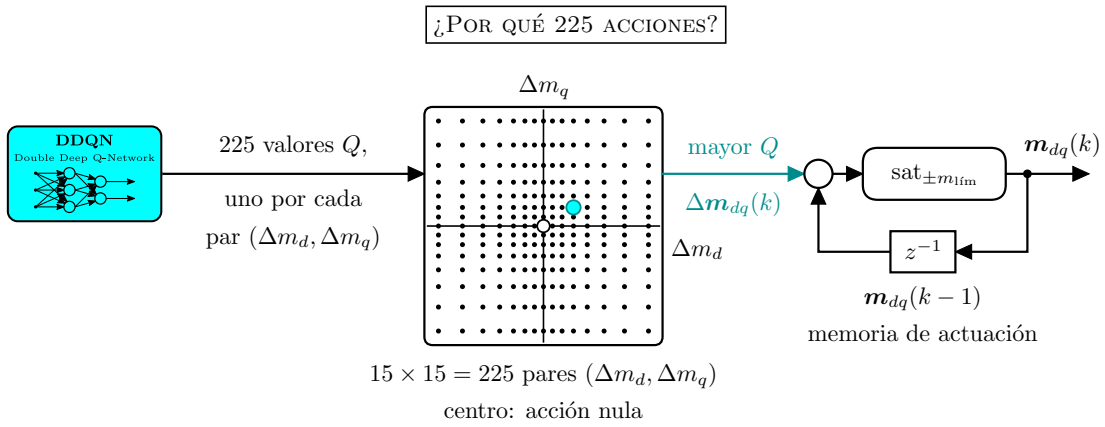


Figura 3.2: Acción incremental del agente. (a) La red entrega 225 valores Q, uno por cada par $(\Delta m_d, \Delta m_q)$. (b) Cada acción es un par $(\Delta m_d, \Delta m_q)$ con incrementos de (3.11); el centro es la acción nula y la separación no está a escala. (c) La acción con el mayor valor Q determina el par que se acumula según (3.14). Elaboración propia.

El límite utilizado en el entrenamiento depende de la tensión del *DC link*:

$$m_{\mathrm{lím}}(V_{dc}) = \mathrm{mín}\left(1, \frac{V_{dc}}{V_N}\right), \qquad V_N = 650 \text{ V}. \tag{3.16}$$

La saturación $\mathrm{sat}_{\pm m_{\mathrm{lím}}}$ actúa de manera independiente sobre cada componente, de modo que

$$|m_d(k)| \le m_{\mathrm{lím}} \le 1, \qquad |m_q(k)| \le m_{\mathrm{lím}} \le 1. \tag{3.17}$$

Para $V_{dc} = 50$ a 300 V, el límite por componente va de 0,077 a 0,462. La modulante se expresa sobre esa misma base: antes del tiempo muerto y de la perturbación, la carga recibe $\boldsymbol{v}_k = (V_N/2)\,\boldsymbol{m}_{dq}(k)$, de modo que $m_{\mathrm{lím}}$ corresponde a $V_{dc}/2$ por componente. Una acción nula conserva la actuación del período anterior, mientras que las demás acciones aumentan o disminuyen progresivamente m_d y m_q. La memoria de actuación es la modulante ya recortada: el valor $\boldsymbol{m}_{dq}(k)$ que entrega (3.14) se aplica en el período siguiente y es el que reciben las entradas $m_d(k-1)$ y $m_q(k-1)$ de la observación siguiente. En el entrenamiento éste es el único límite de la orden: no hay límite de módulo, paso a *abc* ni Min–Max.

Las 225 acciones corresponden, por tanto, a un conjunto discreto de incrementos y no a 225 valores finales de modulante. Los valores acumulados de m_d y m_q pueden adoptar

distintos valores dentro de su intervalo permitido. Los valores de $\Delta m_{\text{mín}}$, $\Delta m_{\text{máx}}$ y la correspondencia entre cada par de incrementos y las salidas de la red se detallan en el Apéndice A.3.

La Figura 3.3 compara esta elección con una salida formada por valores finales de modulante. Si las 225 salidas fueran valores finales de modulante, cada eje dispondría de 15 niveles fijos y un valor requerido entre dos de ellos no podría mantenerse constante. Con incrementos, la modulante acumulada alcanza valores intermedios, su cambio en cada período queda acotado por $\Delta m_{\text{máx}}$ y la acción nula conserva la actuación alcanzada.

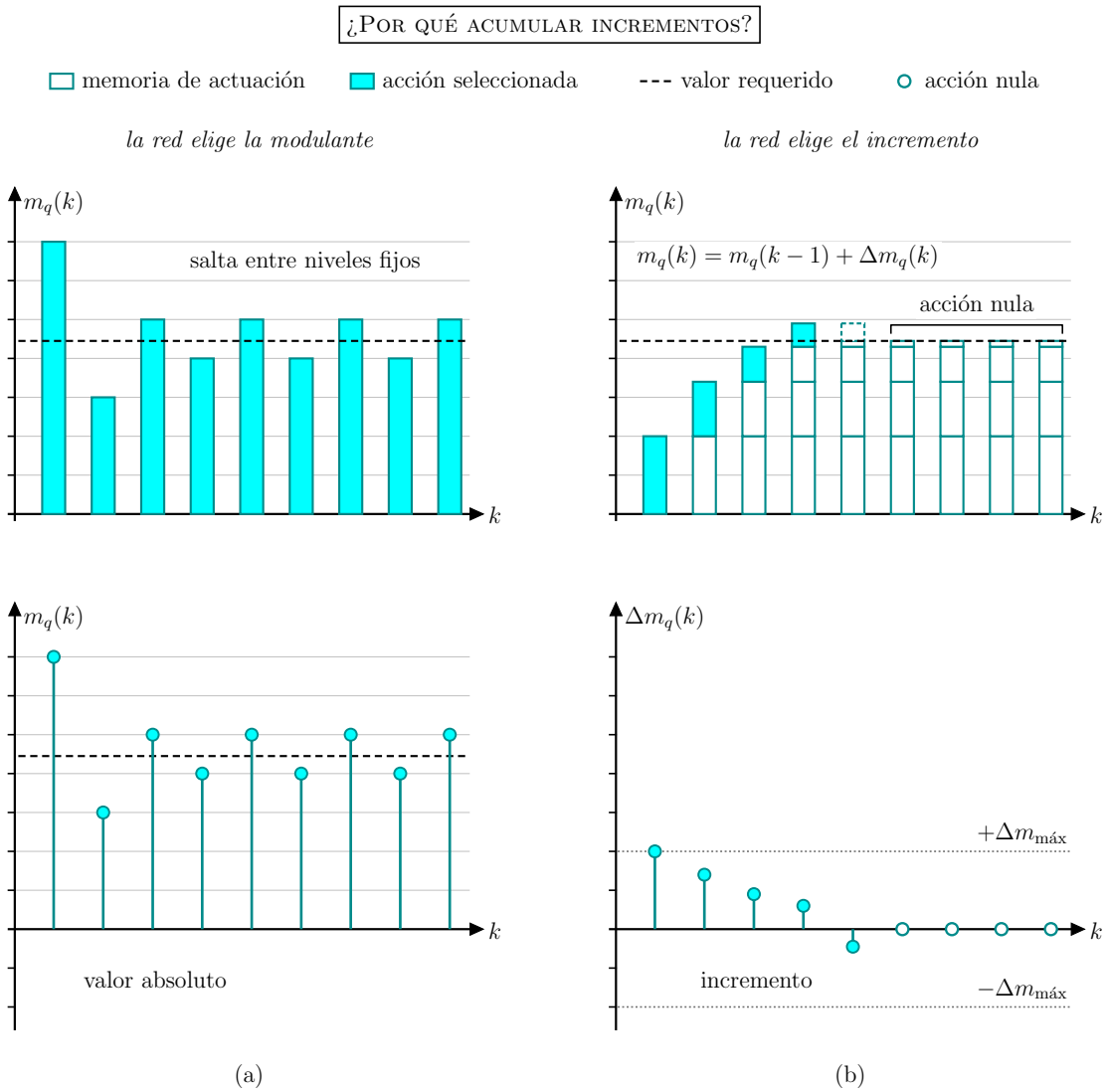


Figura 3.3: Por qué se acumulan los incrementos. (a) Modulante elegida directamente entre quince niveles fijos por eje. (b) Modulante formada por los incrementos almacenados y la acción de cada período; la acción nula conserva el valor alcanzado. Abajo, lo que elige la red en cada período: (a) el valor absoluto de la modulante; (b) el incremento, acotado por $\pm\Delta m_{\text{máx}}$. Se muestra el eje q, con alturas ilustrativas. Elaboración propia.

La Figura 3.4 resume el lazo completo.

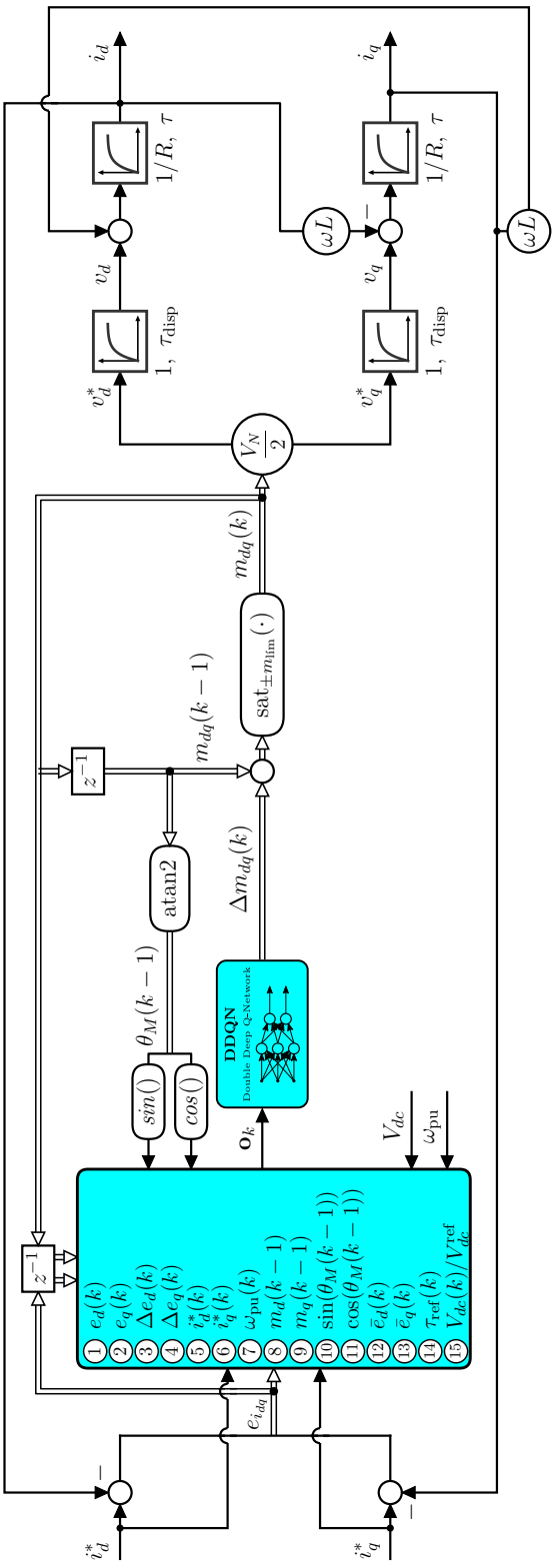


Figura 3.4: Lazo de corriente del entorno R–L controlado por el agente, desde la observación hasta la tensión aplicada a la carga. La acción incremental se acumula según (3.14), con el límite $\pm m_{\text{lim}}$ por componente, y la base $V_N/2$ la convierte en la tensión que se incorpora a la carga con el retardo discreto descrito en (3.6). El bloque τ_{disp} representa la aproximación continua empleada para ese efecto en el esquema equivalente, con $\tau_{\text{disp}} = \frac{3}{2}T_{\text{sw}}$ según [1]. Bajo los bloques de la carga, $1/R$ y $\tau = L/R$ son la ganancia y la constante de tiempo que (3.1) da a cada eje. La carga conserva el acoplamiento ωL en ambos ejes; ω_{pu} y V_{dc} permanecen constantes durante cada episodio, según la Sección 3.3. Las escalas son las del entrenamiento. En la máquina, la Figura 6.2 expresa la orden con $V_{dc}/2$ y su límite. La Tabla A.12 detalla los valores de cada implementación. Elaboración propia.

3.4.3. Recompensa

La recompensa combina tres términos: el error de seguimiento de corriente, el movimiento de la modulante y una penalización asociada a su magnitud. El término de seguimiento es

$$r_{e,k} = -c_d|e_{d,k}| - c_q|e_{q,k}|, \qquad e_{d,k} = i^*_{d,k} - i_{d,k}, \qquad e_{q,k} = i^*_{q,k} - i_{q,k}. \tag{3.18}$$

De esta forma, la recompensa disminuye a medida que aumenta el error de corriente en cualquiera de los dos ejes.

El segundo término penaliza el movimiento de la modulante,

$$r_{m,k} = \begin{cases} 0, & t_{\mathrm{ref},k} < \tau_\Delta, \\ -c_m \dfrac{\Delta m_k}{\Delta m_{\mathrm{máx}}}, & t_{\mathrm{ref},k} \geq \tau_\Delta, \end{cases} \qquad \Delta m_k = \sqrt{(\Delta m_{d,k})^2 + (\Delta m_{q,k})^2}. \tag{3.19}$$

Durante el intervalo inmediatamente posterior a un cambio de referencia, esta penalización se desactiva para permitir que el agente realice correcciones mayores. Una vez transcurrido τ_Δ, se favorecen acciones de menor magnitud cuando éstas son suficientes para mantener el seguimiento.

El tercer término penaliza la magnitud de la modulante cuando supera el umbral radial definido para el entrenamiento,

$$r_{s,k} = \begin{cases} 0, & m_k \leq m_{\mathrm{lim}}, \\ -c_s(m_k - m_{\mathrm{lim}}), & m_k > m_{\mathrm{lim}}, \end{cases} \qquad m_k = \sqrt{m_d^2(k) + m_q^2(k)}. \tag{3.20}$$

La recompensa total se obtiene sumando los tres aportes,

$$r_k = \underbrace{r_{e,k}}_{\text{seguimiento}} + \underbrace{r_{m,k}}_{\text{movimiento}} + \underbrace{r_{s,k}}_{\text{magnitud de la modulante}}. \tag{3.21}$$

El término de seguimiento permanece activo durante todo el episodio. La única contribución que se desactiva temporalmente después de un cambio de referencia es la penalización asociada al movimiento de la modulante. Los coeficientes c_d, c_q, c_m y c_s, junto con τ_Δ y m_{lim}, se presentan en el Apéndice A.3.

Debe distinguirse esta penalización radial del recorte definido en (3.14). El recorte actúa de manera independiente sobre cada eje y garantiza $|m_d| \leq m_{\mathrm{lím}}$ y $|m_q| \leq m_{\mathrm{lím}}$, mientras que (3.20) considera el módulo conjunto de la modulante. Esta forma radial se adopta para simplificar el entrenamiento: no requiere representar el hexágono del inversor ni transformar la orden a *abc*. La inyección Min–Max forma parte del modulador implementado: se aplica después y no modifica la referencia *dq* entregada por el controlador.

3.5. Agente, entrenamiento y selección

La red en línea tiene una arquitectura 15–128–128–225, con activaciones ReLU en las capas ocultas y salida lineal. Sus 15 entradas corresponden a la observación descrita en la sección anterior, mientras que las 225 salidas entregan un valor Q para cada una de las

acciones disponibles. La red contiene 47 585 parámetros y requiere 47 104 operaciones de multiplicación y acumulación por inferencia. Estas cifras caracterizan su carga de cálculo, pero el tiempo de ejecución debe verificarse directamente en la plataforma.

Durante el entrenamiento, cada transición almacenada en la memoria se utiliza para calcular el valor objetivo DDQN definido en la Sección 2.3.2. Cada episodio dura 0,5 s y empieza con un escenario nuevo. Su última transición se guarda siempre con $d_i = 1$, de modo que su objetivo es sólo la recompensa. En las demás, $d_i = 0$, y no hay otra condición de terminación. Ese final es una condición terminal artificial en una tarea de control continuo, no de horizonte finito, y no se comparó con conservar el valor del estado siguiente. La red no recibe el tiempo restante, porque τ_{ref} sólo mide el tiempo desde el último cambio de referencia, saturado a los 20 ms. El efecto de ese final no se aisló ni se presenta como causa del error en estado estacionario. La red objetivo se actualiza de forma gradual mediante

$$\boldsymbol{\theta}^{-} \leftarrow (1 - \tau_\theta)\boldsymbol{\theta}^{-} + \tau_\theta \boldsymbol{\theta},$$

donde $\boldsymbol{\theta}$ y $\boldsymbol{\theta}^{-}$ corresponden a los parámetros de las redes en línea y objetivo, respectivamente. La función de pérdida es la de Huber descrita en la Sección 2.3.2 y los parámetros de la red en línea se actualizan mediante Adam [24]. Los hiperparámetros utilizados se presentan en el Apéndice A.3.

El entrenamiento se organiza en episodios de 0,5 s. Al finalizar cada uno se realizan las actualizaciones de la red a partir de muestras extraídas de la memoria de repetición. La probabilidad de exploración ε disminuye durante la primera parte del entrenamiento y posteriormente se mantiene constante. De manera periódica, la política se evalúa sin exploración sobre ocho escenarios fijos y estas evaluaciones no incorporan nuevas transiciones a la memoria. El Algoritmo A.1 reúne el procedimiento completo.

Para comparar las políticas obtenidas durante el entrenamiento se utiliza el índice

$$\mathcal{C} = \frac{1}{8}\sum_{j=1}^{8}\left[I_b\left(|\mu_{e_d,j}| + |\mu_{e_q,j}|\right) + \frac{1}{2}\left(\sigma_{i_d,j} + \sigma_{i_q,j}\right)\right], \tag{3.22}$$

donde $\mu_{e_d,j}$ y $\mu_{e_q,j}$ son los errores medios en el último 40 % de cada escenario, expresados en pu, y $\sigma_{i_d,j}$ y $\sigma_{i_q,j}$ son las desviaciones estándar de las corrientes en amperios para el escenario j. La recompensa guía la actualización de la red y reúne los términos ponderados de seguimiento, movimiento y magnitud de la modulante. En cambio, $\mathcal{C}$ compara las políticas mediante el error y la variabilidad de corriente, ambos en amperios. Un menor valor representa un mejor resultado según este criterio de selección.

En esa ventana las referencias todavía cambian, de modo que un seguimiento perfecto, con $i_d = i_d^*$ e $i_q = i_q^*$, también tendría desviación estándar. Ese valor, $\mathcal{C}^*$, se obtiene al reemplazar en (3.22) cada corriente por su referencia. Depende sólo de los ocho escenarios y es el mismo para todas las evaluaciones de una etapa. Las figuras de este capítulo muestran el costo sobre el seguimiento perfecto,

$$\Delta\mathcal{C} = \mathcal{C} - \mathcal{C}^*, \tag{3.23}$$

que suma el sesgo y la desviación estándar que agrega el agente. Restar $\mathcal{C}^*$ no cambia la evaluación de menor costo. La Figura 3.5 resume este cálculo. El Apéndice A.3 los define.

Una vez terminado el entrenamiento, sólo se exportan los parámetros de la red en línea y éstos permanecen fijos durante los ensayos. En cada período, la observación genera 225

valores Q, uno por cada acción disponible, y se selecciona

$$a_k = \arg\max_{a \in \mathcal{A}} Q(\boldsymbol{o}_k, a; \boldsymbol{\theta}). \tag{3.24}$$

La acción seleccionada determina el par de incrementos de modulante descrito anteriormente. Durante esta etapa no intervienen la red objetivo, la memoria de repetición ni la exploración. La implementación de la red y la medición de su tiempo de inferencia se presentan en el Apéndice A.5.

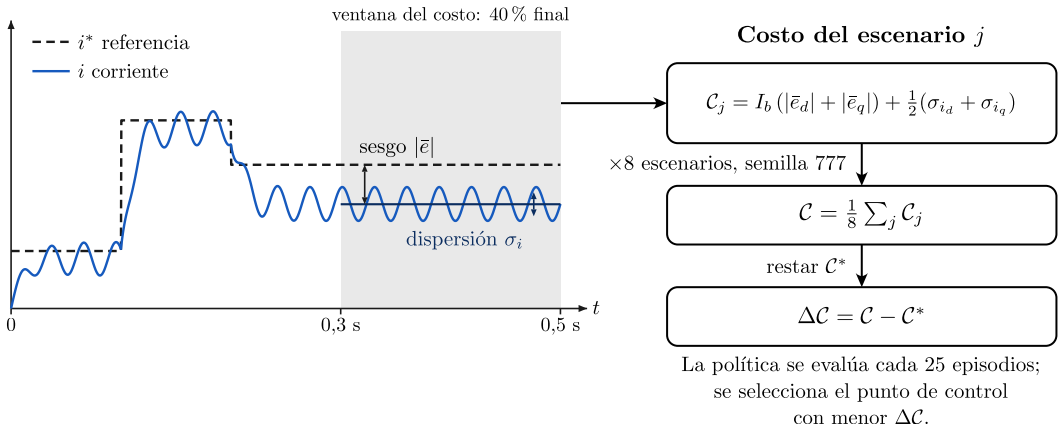


Figura 3.5: Evaluación periódica de los puntos de control. El costo de cada escenario reúne el sesgo y la desviación estándar de la corriente, σ_i, medidos en el 40 % final y suma los ejes d y q; $I_b = 15{,}6$ A expresa el sesgo en amperios. Como σ_i es de la corriente, un cambio de referencia dentro de la ventana también suma al costo. $\mathcal{C}$ promedia ocho escenarios reproducibles y $\Delta\mathcal{C}$ resta $\mathcal{C}^*$, el mismo índice con $i_d = i_d^*$ e $i_q = i_q^*$. Las curvas son ilustrativas y muestran un solo eje.

3.5.1. Evolución de los tres agentes

La Tabla 3.1 resume las tres etapas. El Agente 2 refinado procede del linaje del Agente 2 y continúa el aprendizaje desde un punto de control de refinamiento previo. La etapa mantiene la dinámica $R{-}L$, la perturbación en q y las condiciones indicadas. Los pesos se ajustan durante este refinamiento y quedan fijos para los ensayos.

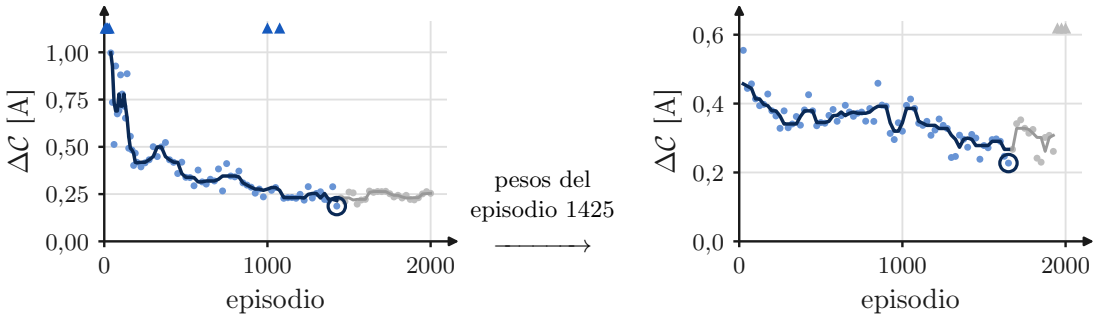
El Agente 2 no procede directamente del Agente 1. Primero se entrenó desde cero una red con resistencia baja y se seleccionó el episodio 1425. Sus pesos iniciaron una segunda etapa, en la que se ampliaron el intervalo del *DC link* y la corriente. El episodio 1650 de esa etapa se seleccionó como Agente 2. La Figura 3.6 muestra las dos etapas con el costo sobre el seguimiento perfecto de (3.23). En cada una se seleccionó la evaluación de menor costo: 0,19 A en el episodio 1425 y 0,23 A en el 1650, con 0,14 y 0,18 A de sesgo y 0,05 A de desviación estándar adicional en ambos.

El Agente 2 refinado parte del Agente 2, episodio 1650, y continúa el aprendizaje en cinco etapas sobre la misma dinámica $R{-}L$. La Figura 3.7 muestra la evolución del costo sobre el seguimiento perfecto y de la recompensa media durante la última de ellas, de la que se seleccionó el Agente 2 refinado.

Tabla 3.1: Procedencia de los tres agentes. Cada etapa modifica varias condiciones, por lo que la tabla no permite aislar el efecto de una única modificación.

Etapa y agente	Cambios introducidos	Punto de control de origen	Episodio seleccionado	Archivo de pesos	Ensayos en que se utilizó
Agente 1	Entrenamiento desde cero con R entre 12 y 18 Ω, *DC link* entre 100 y 210 V y frecuencia fija	Ninguno	1775	`ckpt_ep1775.mat` y `ddqn_e6_ep1775_obs15.h`	No se utilizó en los ensayos finales
Agente 2	Nueva acción incremental para resistencia baja y ampliación posterior de *DC link* y corriente	Etapa de resistencia baja, episodio 1425 (Figura 3.6)	1650	`ckpt_ep1650.mat` y `ddqn_e9_ep1650_obs15.h`	Capítulo 4: simulación y banco con la carga $R-L$
Agente 2 refinado	Refinamiento progresivo sobre la misma dinámica $R-L$, con perturbación de tensión en q, intervalos de frecuencia y corriente ampliados, mayor castigo al movimiento y 35 % de episodios con $\omega = 0$, $p_q = 0$ y referencia de corriente sostenida	Agente 2, episodio 1650	675	`ckpt_ep0675.mat` y `ddqn_e9_ep0675_obs15.h`	Capítulos 6 y 7: simulación y banco con la máquina

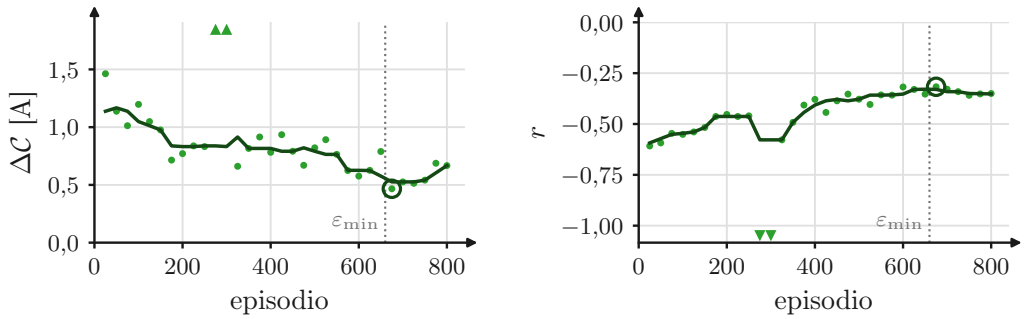
Las etapas intermedias, las rutas completas, los identificadores y las huellas de los archivos se detallan en el Apéndice C.



(a) entrenamiento previo con resistencia baja

(b) Agente 2

Figura 3.6: Obtención del Agente 2. (a) Entrenamiento desde cero con R de 0,5 a 3,5 Ω, *DC link* de 50 a 210 V y corriente hasta 5 A. (b) Parte de los pesos del episodio 1425 y amplía el *DC link* hasta 320 V y la corriente hasta 7,5 A; sus episodios se cuentan de nuevo desde cero. El costo resta la variación que produciría un seguimiento perfecto de las referencias. Los puntos corresponden a evaluaciones periódicas sin exploración y la línea representa la tendencia. En azul, el entrenamiento hasta la política seleccionada, marcada con un círculo; en gris, el resto. Los triángulos indican evaluaciones fuera de escala.



(a) costo sobre el seguimiento perfecto

(b) recompensa media

Figura 3.7: Evaluaciones *greedy* realizadas sobre ocho escenarios fijos. El costo resta la variación que produciría un seguimiento perfecto de las referencias. Los puntos corresponden a evaluaciones cada 25 episodios y la línea representa una mediana móvil de cinco evaluaciones. La línea vertical indica el fin de la reducción de ε, en el episodio 660. El círculo identifica la política seleccionada, en el episodio 675.

Entre las 32 evaluaciones completas, la política seleccionada obtuvo el menor costo sobre el seguimiento perfecto, $\Delta\mathcal{C} = 0{,}47$ A, de los cuales 0,41 A corresponden al sesgo y 0,06 A a la desviación estándar adicional. La misma política obtuvo la mayor recompensa media de evaluación; ambos criterios identificaron el episodio 675. Esta selección corresponde al criterio declarado y no constituye una demostración de optimalidad global.

3.6. Alcance y límites de transferencia

El entorno conserva la dinámica de la carga $R{-}L$ en coordenadas dq, con acoplamiento entre ejes. Durante el entrenamiento se varían R, L, la frecuencia angular del marco dq, la tensión del *DC link* y las referencias de corriente. También se representan el ruido de medida, la cuantización, una aproximación del tiempo muerto y un período de retardo en la actuación. El refinamiento mantiene esta misma planta y añade únicamente la perturbación de tensión en q. Los filtros y diferencias no aseguran que la observación sea un estado de Markov suficiente, porque la red no observa R, L ni p_q.

El entorno no incorpora las ecuaciones de estado del flujo de rotor ni de la dinámica mecánica, la saturación magnética, las caídas de tensión en los semiconductores o la conmutación explícita del inversor. Estas exclusiones delimitan lo aprendido por el agente: la dinámica de la máquina no se introdujo durante el entrenamiento. Por ello, la transferencia no se supone a partir del modelo, sino que se comprueba experimentalmente, primero sobre la carga $R{-}L$ y después en el accionamiento. En cada ensayo se utilizan pesos fijos y el tiempo de evaluación de la red se verifica por separado en la plataforma.

En esta memoria, sin entrenamiento adicional sobre la máquina significa que los pesos y sesgos de la red permanecen fijos durante la simulación y los ensayos. Los lazos externos del accionamiento sí se configuran y ajustan para generar las referencias de corriente, sin modificar la red neuronal. El entorno tampoco se construyó sin conocimiento de la aplicación: se utilizó información nominal de la máquina para dimensionar la perturbación p_q. Esto no incorpora a la planta de entrenamiento las ecuaciones del flujo de rotor ni la dinámica mecánica. El cálculo de su escala se presenta en el Apéndice A.3.

Capítulo 4

Verificación sobre la carga $R-L$

Este capítulo verifica experimentalmente la política aprendida por el Agente 2 sobre una carga trifásica $R-L$. Finalizado el entrenamiento, la red en línea seleccionada se implementó en el procesador de la plataforma BRAIn [25] y sus pesos permanecieron fijos durante los ensayos. La respuesta experimental se compara con la simulación bajo las mismas maniobras de referencia, antes de estudiar la transferencia de la política a la máquina de inducción.

4.1. Implementación y condiciones de ensayo

La Figura 4.1 muestra el montaje experimental; la configuración del banco y las conexiones con la plataforma se detallan en la Sección A.1. La salida del inversor alimenta una carga trifásica formada por una resistencia nominal de 1 Ω y una inductancia de 10 mH por fase. Estos valores corresponden a una condición considerada durante el entrenamiento del agente.

Las tensiones y corrientes adquiridas por BRAIn permiten además obtener una representación equivalente de la planta observada desde el lazo de corriente. Para cada maniobra se reconstruyó la tensión de fase a partir de la modulante y del V_{dc} medido, $v_a = m_a V_{dc}/2$, y se ajustaron por mínimos cuadrados las componentes fundamentales de v_a e i_a a 50 Hz sobre la ventana de estado estacionario. A partir de $Z = V_1/I_1 = R_{\mathrm{eq}} + j\omega L_{\mathrm{eq}}$ se obtuvieron ambos parámetros. Esta estimación no corresponde a una medición independiente de los componentes, sino a la impedancia equivalente obtenida a partir de las tensiones y corrientes medidas. La Tabla 4.1 compara ambos valores.

Tabla 4.1: Valores nominales de la carga y valores equivalentes obtenidos a partir de las mediciones de la plataforma BRAIn.

Parámetro	Nominal	Equivalente observado	Desviación
R [Ω]	1,00	1,16–1,52	+16 a +52 %
L [mH]	10,00	≈ 10,41	≈ +4,1 %

La inductancia equivalente permanece próxima al valor nominal. La resistencia estimada resulta mayor porque el cociente V_1/I_1 reúne la resistencia de la carga y las caídas de tensión que el cálculo no separa, como las del inversor, los conductores y sus uniones.

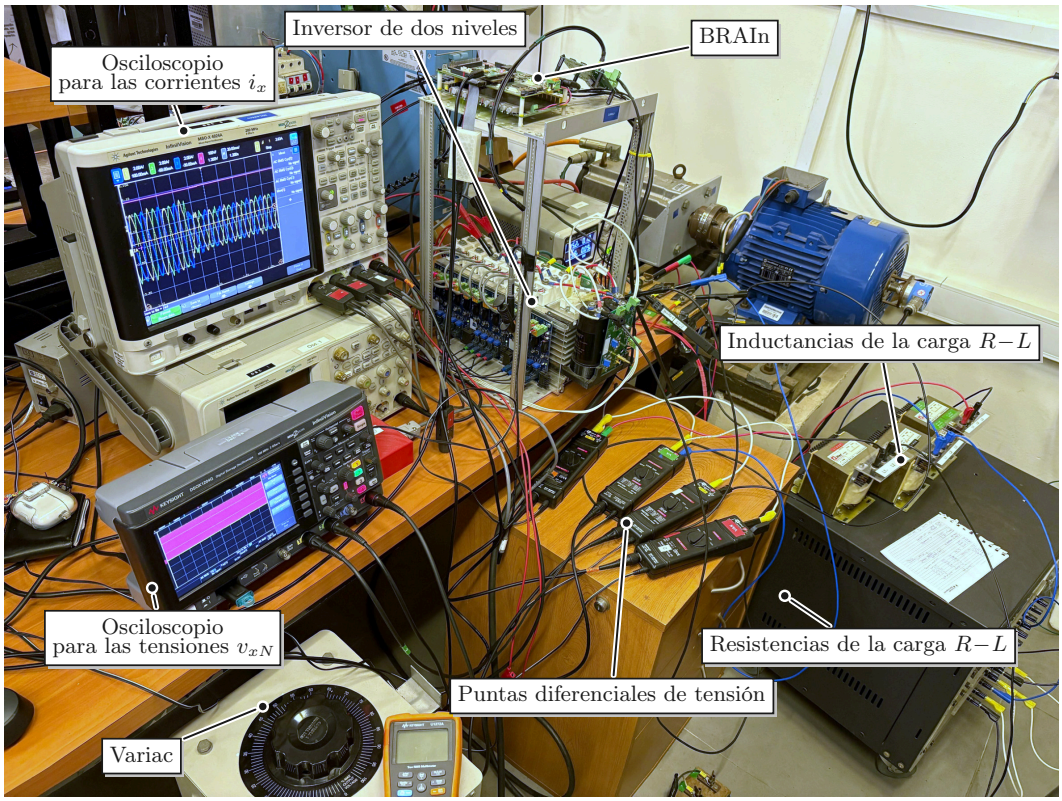


Figura 4.1: Plataforma BRAIn y etapa de potencia utilizadas en los ensayos sobre la carga trifásica $R-L$. La máquina visible al fondo no interviene en esta campaña.

Las condiciones principales de los ensayos se resumen en la Tabla 4.2.

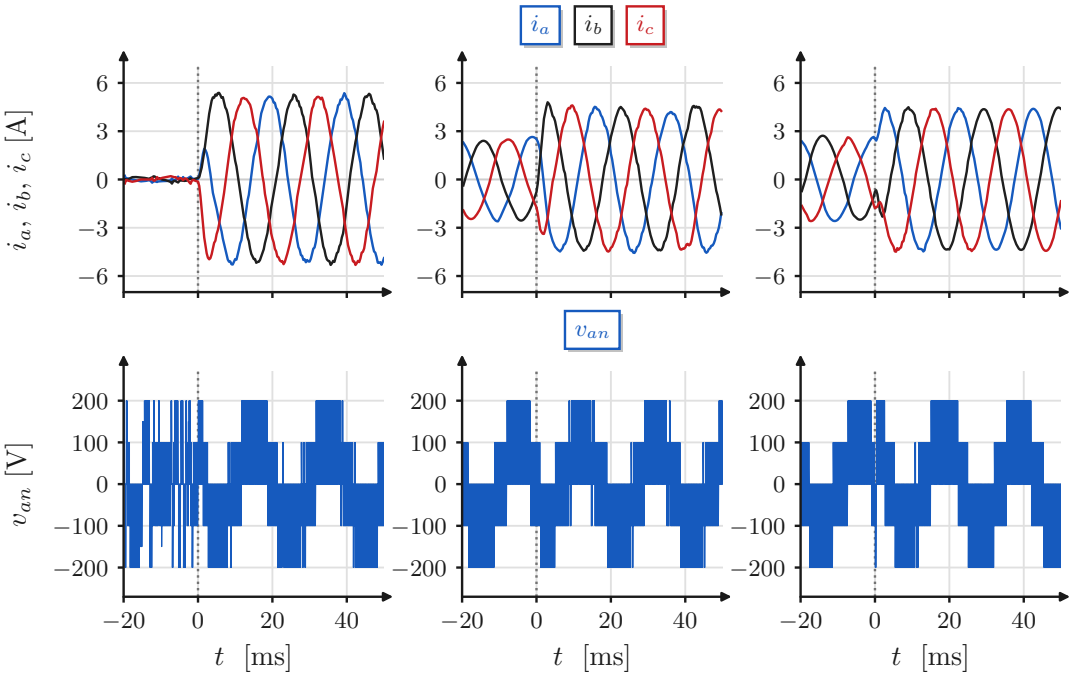
Tabla 4.2: Condiciones principales de los ensayos sobre la carga $R-L$.

Magnitud	Condición
Política	Agente 2, red en línea final con pesos fijos
Carga nominal	$R = 1\ \Omega$, $L = 10$ mH
Referencia del *DC link*	300 V
V_{dc} medido	289,8–293,7 V
Control y PWM	3 kHz
Marco dq	50 Hz

Se ensayan cuatro cambios de referencia: establecimiento de i_d de 0 a 5 A, reducción de 5 a 2,5 A y establecimiento de i_q desde cero hasta +3,5 A y hasta −3,5 A. En estos dos últimos casos, i_d^* se mantiene en 2,5 A.

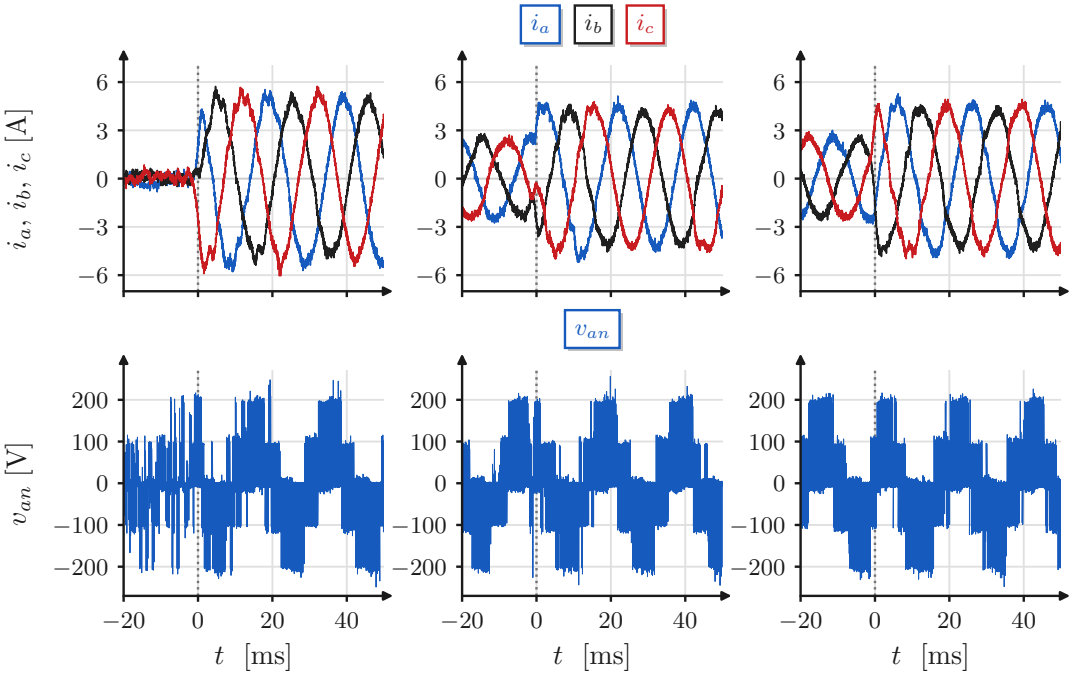
4.2. Verificación en bornes

Las Figuras 4.2 y 4.3 comparan las tensiones y corrientes en bornes obtenidas en el modelo implementado en PLECS [26] y en el ensayo. En cada captura, se definió $t = 0$ en el instante en que cambia la referencia. Las condiciones del banco son las de la Tabla 4.2; la simulación utiliza $V_{dc} = 300$ V y los cambios de corriente descritos en la Sección 4.1.



(a) escalón de i_d (b) escalón positivo de i_q (c) escalón negativo de i_q

Figura 4.2: Corrientes de fase y tensión fase–neutro obtenidas en simulación con $R = 1\ \Omega$, $L = 10$ mH, $V_{dc} = 300$ V y control a 3 kHz.



(a) escalón de i_d (b) escalón positivo de i_q (c) escalón negativo de i_q

Figura 4.3: Corrientes de fase medidas y tensión fase–neutro reconstruida a partir de v_{aN}, v_{bN} y v_{cN}, medidas entre cada salida del inversor y el negativo del *DC link*.

Las corrientes experimentales conservan la secuencia trifásica y presentan amplitudes del mismo orden que las obtenidas en simulación. La tensión fase–neutro reconstruida conserva la estructura de niveles esperada y presenta *peaks* de tensión durante las conmutaciones, con un mínimo de $-248{,}9$ V y un máximo de $+256{,}3$ V. Estas mediciones verifican desde los bornes la actuación física del inversor sobre la carga.

La cuantificación del error de control se realiza posteriormente en el marco dq, donde las corrientes pueden compararse directamente con sus referencias.

4.3. Seguimiento de corriente en el marco dq

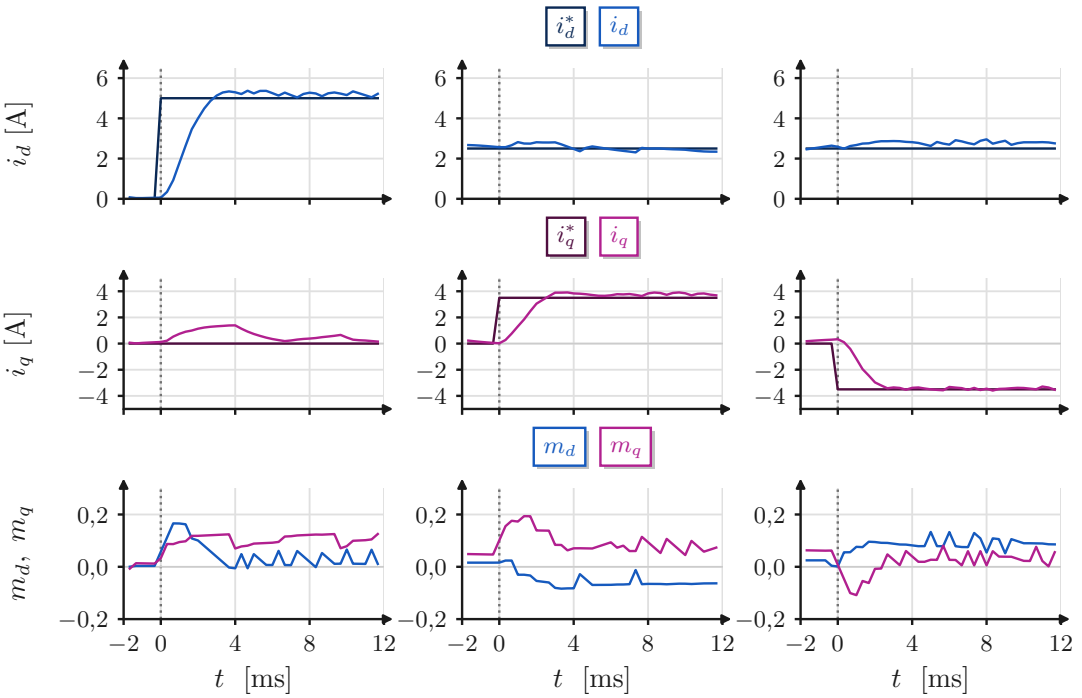
La evaluación directa del agente se realiza en el marco dq, donde i_d e i_q pueden compararse con sus respectivas referencias.

En estos ensayos el marco dq gira a $f = 50$ Hz. Como la carga no posee rotor ni *encoder*, la plataforma BRAIn genera el ángulo en cada período de control según

$$\theta_{k+1} = \mathrm{mod}(\theta_k + 2\pi f T_s,\, 2\pi)\,.$$

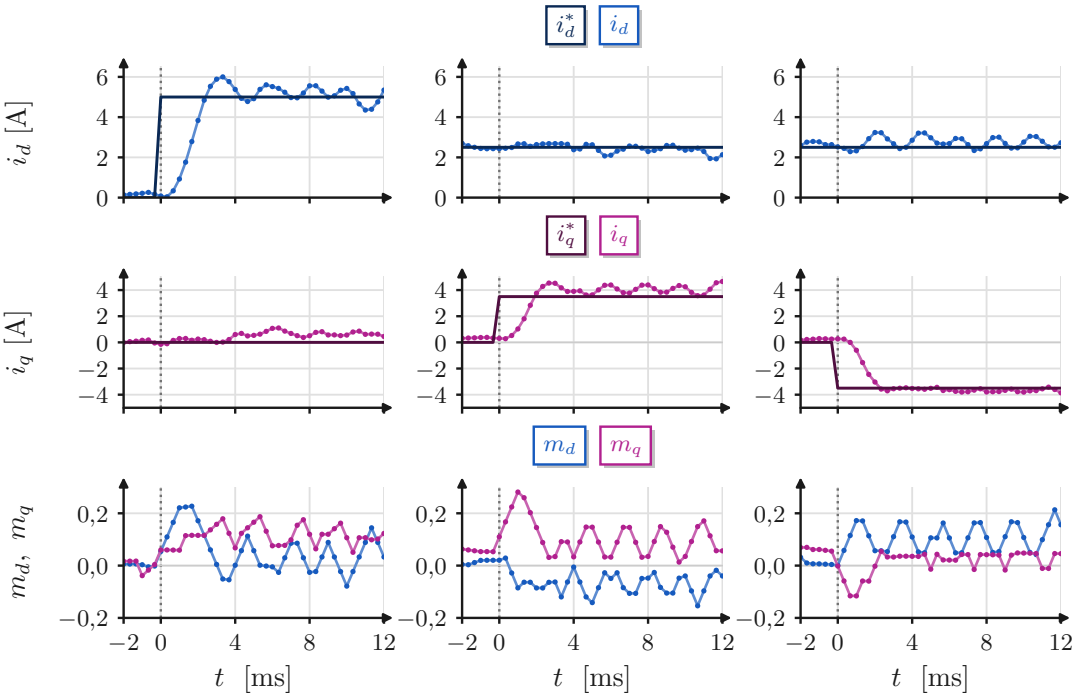
La transformación emplea $\gamma_k = \theta_k - \pi/2$, de acuerdo con la convención del Capítulo 2. Al expresar la carga R–L en este marco giratorio, los términos $\pm\omega L$ acoplan ambos ejes siempre que $\omega = 2\pi f \neq 0$; el agente debe compensar este acoplamiento durante el seguimiento.

Las Figuras 4.4 y 4.5 comparan las corrientes del marco dq, sus referencias y las modulantes obtenidas en simulación y en el ensayo.



(a) escalón de i_d (b) escalón positivo de i_q (c) escalón negativo de i_q

Figura 4.4: Corrientes, referencias y modulantes obtenidas en simulación.



(a) escalón de i_d (b) escalón positivo de i_q (c) escalón negativo de i_q

Figura 4.5: Corrientes, referencias y modulantes registradas por la plataforma BRAIn. La reducción de i_d^*, incluida en las métricas de este capítulo, se presenta en el Apéndice B.

El Agente 2 sigue los cambios del eje directo y establece corrientes de cuadratura de ambos signos, mientras mantiene el seguimiento de i_d. Las corrientes permanecen próximas a sus referencias; en simulación y en el ensayo se observa una desviación sostenida, y la respuesta experimental presenta además una mayor variabilidad.

El error se define como $e_x = i_x^* - i_x$, con $x \in \{d, q\}$. Para cuantificar el seguimiento se utiliza $\mathrm{RMSE}_x = \sqrt{N^{-1} \sum_{k=1}^{N} e_x^2[k]}$. En las ventanas analizadas se informan además el sesgo $\overline{e}_x = N^{-1} \sum_{k=1}^{N} e_x[k]$ y la desviación estándar del error $\sigma_{e_x} = \sqrt{N^{-1} \sum_{k=1}^{N} (e_x[k] - \overline{e}_x)^2}$. La Tabla 4.3 reúne estas métricas para cada maniobra.

El RMSE conjunto experimental en las ventanas analizadas alcanza 164,4 mA en el eje d y 264,7 mA en el eje q. La mayor desviación sostenida aparece cuando aumenta la referencia de i_q, con un sesgo de $-372{,}3$ mA. En simulación, el RMSE conjunto es de 146,7 y 263,5 mA, y el mayor sesgo aparece en la misma maniobra, con $-348{,}9$ mA. La simulación también presenta un sesgo de magnitud similar al experimental, por lo que este no se explica únicamente por la implementación en el banco. El ensayo presenta mayor desviación estándar del error en las tres maniobras de las Figuras 4.4 y 4.5, salvo en el eje q del escalón de i_d, y menor en la bajada de i_d.

Tabla 4.3: RMSE, sesgo y desviación estándar del error en las ventanas analizadas, en mA, simulados (Sim.) y experimentales (Exp.). La ventana comienza 30 ms después del cambio de referencia y termina al final del registro experimental, entre 310 y 347 ms después del cambio; la simulación usa la misma ventana. Las letras corresponden a los paneles de las Figuras 4.4 y 4.5. Para calcular el RMSE conjunto, se reúnen las muestras de las cuatro ventanas y se utiliza la ecuación (4.1).

		eje d			eje q		
Maniobra		RMSE	Sesgo	σ_e	RMSE	Sesgo	σ_e
(a) Escalón de i_d, 0 a 5 A	Sim.	164,6	−146,9	74,2	140,1	−19,8	138,7
	Exp.	176,6	−155,4	83,9	125,8	−27,7	122,7
(b) Escalón positivo de i_q, 0 a +3,5 A	Sim.	123,0	47,2	113,6	380,5	−348,9	151,8
	Exp.	177,8	18,8	176,8	427,2	−372,3	209,4
(c) Escalón negativo de i_q, 0 a −3,5 A	Sim.	150,9	−102,3	111,0	174,6	127,2	119,6
	Exp.	185,7	−82,6	166,3	185,1	110,4	148,5
Escalón de bajada de i_d, 5 a 2,5 A †	Sim.	143,5	−85,7	115,1	292,9	−255,1	143,9
	Exp.	109,8	−12,5	109,1	233,7	−204,8	112,6
Conjunto	Sim.	**146,7**	–	–	**263,5**	–	–
	Exp.	**164,4**	–	–	**264,7**	–	–

† No aparece en las Figuras 4.4 y 4.5; se presenta en el Apéndice B.

El RMSE conjunto se calcula con todas las muestras de las cuatro maniobras:

$$\mathrm{RMSE}_x = \sqrt{\frac{\sum_{j=1}^{4}\sum_{k=1}^{N_j} e_{x,j}^2[k]}{\sum_{j=1}^{4} N_j}}, \qquad x \in \{d, q\}. \tag{4.1}$$

Aquí, N_j es el número de muestras de la maniobra j. Por tanto, cada maniobra contribuye según $N_j / \sum_{\ell=1}^{4} N_\ell$. No se promedian los cuatro RMSE calculados por separado.

La ventana analizada comienza 30 ms después del cambio de referencia. En simulación, algunas maniobras presentan una evolución del error hasta aproximadamente 100–160 ms, y hasta 280 ms en el eje d del escalón positivo de i_q; posteriormente, el sesgo persiste hasta el siguiente cambio, a los 400 ms. El agente evaluado presenta, por tanto, un error sostenido de seguimiento, principalmente en el eje de cuadratura.

4.4. Respuesta transitoria

La respuesta transitoria se compara mediante el tiempo de subida del 10 al 90 % y el sobrepaso [27], con el mismo criterio para los resultados simulados y experimentales. El recorrido se define entre el promedio de los 20 ms anteriores al cambio de referencia y el valor medio alcanzado en la ventana de estado estacionario; el sobrepaso corresponde al extremo de la primera oscilación posterior al cambio, esto es, antes de que la respuesta vuelva a cruzar su valor final. Por el error en estado estacionario del agente, ese valor final no coincide con la referencia: en la simulación del escalón de i_d, por ejemplo, la corriente se establece en 5,15 A. Medido respecto de la referencia, el sobrepaso mezclaría el transitorio con el sesgo. La Tabla 4.4 resume ambos indicadores para cada maniobra.

Los tiempos obtenidos muestran una respuesta del orden de pocos milisegundos tanto en simulación como en el banco. No se espera una coincidencia exacta, porque el modelo y el banco difieren en la planta efectiva, la tensión disponible y la cadena de implementación.

Tabla 4.4: Tiempo de subida y sobrepaso, simulados (Sim.) y experimentales (Exp.). Cada maniobra cambia la referencia de un solo eje; el guion indica el eje que no recibe escalón. Las letras corresponden a los paneles de las Figuras 4.4 y 4.5.

	eje d				eje q			
	t_{10-90} [ms]		Sobrepaso [%]		t_{10-90} [ms]		Sobrepaso [%]	
Maniobra	Sim.	Exp.	Sim.	Exp.	Sim.	Exp.	Sim.	Exp.
(a) Escalón de i_d, 0 a 5 A	2,0	1,4	4,5	16,8	–	–	–	–
(b) Escalón positivo de i_q, 0 a +3,5 A	–	–	–	–	2,0	1,2	1,7	18,2
(c) Escalón negativo de i_q, 0 a −3,5 A	–	–	–	–	1,9	1,3	0,7	2,5
Escalón de bajada de i_d, 5 a 2,5 A †	2,0	2,5	7,9	0,5	–	–	–	–

† No aparece en las Figuras 4.4 y 4.5; se presenta en el Apéndice B.

El tiempo de asentamiento dentro de una banda de ±5 % no se informa, porque la variación de la corriente registrada provoca salidas de la banda durante prácticamente toda la ventana registrada. El sobrepaso se evalúa en la primera oscilación posterior al escalón. Así se evita incluir extremos de esa variación que aparecen después del transitorio y que pueden ser mayores que el sobrepaso inicial.

4.5. Resultado de la verificación

La red en línea final del Agente 2, ejecutada en el procesador de la plataforma BRAIn, controla experimentalmente las corrientes de la carga trifásica $R-L$. Las mediciones en bornes verifican la actuación física del inversor, mientras que las variables del marco dq permiten cuantificar directamente el seguimiento.

El RMSE conjunto en estado estacionario es de 164,4 mA en el eje d y 264,7 mA en el eje q, con un sesgo máximo de 372 mA cuando aumenta la referencia de i_q y una desviación estándar del error entre 84 y 209 mA. Aunque el agente presenta un error sostenido de seguimiento, completa las maniobras consideradas con los pesos de la red fijos. El Capítulo 6 estudia la transferencia del Agente 2 refinado, obtenido al continuar el aprendizaje en el entorno $R-L$ con la perturbación del eje q.

Capítulo 5

La referencia convencional

El control convencional proporciona la referencia con la que se evalúa el Agente 2 refinado. Las corrientes, el flujo y la velocidad se controlan mediante controladores proporcional–integrales (PI). Este capítulo presenta la planta de cada lazo, el criterio de ajuste y su implementación, manteniendo la orientación de campo, el estimador y la modulación del Capítulo 2.

5.1. El controlador y los criterios de ajuste

Los tres lazos emplean la misma estructura de controlador PI, con ganancias ajustadas a la dinámica de cada uno,

$$C(s) = k_p + \frac{k_i}{s} = k_p \, \frac{1 + s\,T_i}{s\,T_i}, \qquad T_i = \frac{k_p}{k_i}, \tag{5.1}$$

donde k_p y k_i son las ganancias proporcional e integral y T_i la constante del controlador, cuyo cero queda en $s = -1/T_i$ [1, 27].

El criterio de ajuste se elige a partir de la planta que ve cada controlador. Para una planta formada por un retardo lento y otro rápido, el óptimo del módulo (OM) cancela el polo lento con el cero del PI y ajusta la ganancia. Según la función de lazo cerrado de Fröhr y Orttenburger, el amortiguamiento resulta $\zeta = 1/\sqrt{2}$ y el sobrepaso ante un escalón es aproximadamente 4,3 % [27, pp. 268–269, ecs. (6.52)–(6.55)]. Barrios presenta este criterio y lo aplica al lazo de corriente de su banco de pruebas [1, pp. 57–58 y 63].

Si la constante mayor, T_1, supera cuatro veces la suma σ de las pequeñas, el retardo lento actúa, en primera aproximación, como un integrador. Fröhr y Orttenburger explican que esta aproximación permite emplear el óptimo simétrico [27, pp. 286–287], que Barrios recomienda para regular con mayor rapidez las perturbaciones [1, p. 59, ec. (4.19)]. Esta variante para una planta sin integrador se denomina aquí óptimo simétrico realista (OSR). Para el modelo con un integrador y retardos rápidos se emplea el óptimo simétrico (OS) [27, pp. 282–283, ecs. (6.74)–(6.77)]. Por ello, primero se identifica la planta y sus constantes de tiempo, se elige el criterio y después se calcula el controlador.

En el ajuste clásico del OS, con $T_i = 4\sigma$, la respuesta a un escalón presenta un sobrepaso de 43,4 %. Un alisamiento de primer orden en la referencia, de constante 4σ, lo reduce a 8,1 % [27, pp. 284–285 y 292–294, Figs. 6.30 y 6.35]. En una planta formada sólo por

retardos de primer orden, el sobrepaso es menor y se aproxima al del OM cuando $T_1/(4\sigma)$ se acerca a uno [27, pp. 287 y 294]. Estas cifras corresponden a los modelos lineales considerados y suponen que el controlador no satura; no describen una maniobra con saturación.

5.2. Ajuste del lazo de corriente

En el marco orientado al flujo de rotor, la dinámica de corriente de cada eje queda determinada por la inductancia transitoria y por la resistencia equivalente $R_\sigma = R_s + k_r^2 R_r'$, con $k_r = L_m/L_r'$ [2]. El acoplamiento entre ejes, $\mp\omega_s\,\sigma L_s\,i$, y la tensión inducida por el flujo de rotor se tratan como perturbaciones. La orden calculada se aplica en el período siguiente y se mantiene constante hasta la próxima actualización, como en un retenedor de orden cero (*Zero-Order Hold*, ZOH). Esta retención se explica en el apunte de tiempo discreto de Accionamientos [28, p. 11, Fig. 9]. Para ajustar el PI, el efecto del ZOH se aproxima mediante un retardo de $T_s/2$. Sumado al período T_s entre medir y aplicar la orden, se obtiene un retardo total de $1{,}5T_s$, siguiendo la aproximación empleada por Barrios [1, pp. 61–62]. Este retardo se representa mediante un bloque de primer orden de constante $\tau_c = 1{,}5T_s$; no es el modelo exacto del retenedor. Así, el controlador de cada eje ve la planta

$$G_i(s) = \frac{1/R_\sigma}{(1 + s\,\tau_\sigma)(1 + s\,\tau_c)}, \qquad \tau_\sigma = \frac{\sigma L_s}{R_\sigma}, \tag{5.2}$$

donde $\sigma L_s = 13{,}131$ mH, $R_\sigma = 1{,}781\ \Omega$, $\tau_\sigma = 7{,}3727$ ms [22] y $\tau_c = 0{,}50$ ms. La Figura 5.1 representa ese lazo.

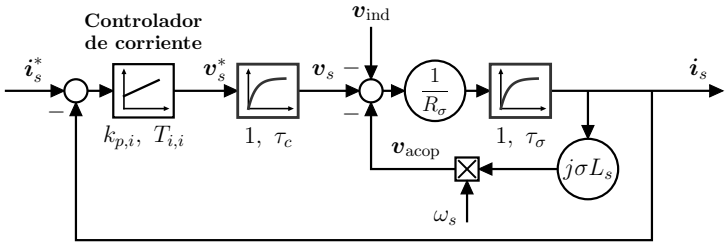


Figura 5.1: Lazo cerrado de corriente de estator sin compensación. Bajo el controlador se indican su ganancia $k_{p,i}$ y su tiempo integral $T_{i,i}$. El convertidor se representa por el primer orden de constante τ_c, y $1/R_\sigma$ y τ_σ son la ganancia y la constante de tiempo de (5.2). La tensión inducida por el flujo de rotor, $\boldsymbol{v}_{\text{ind}}$, y la de acoplamiento, $\boldsymbol{v}_{\text{acop}} = j\sigma L_s \omega_s \boldsymbol{i}_s$, entran como perturbaciones. El esquema omite la prealimentación y las saturaciones, activas en el equipo. Adaptada de [2].

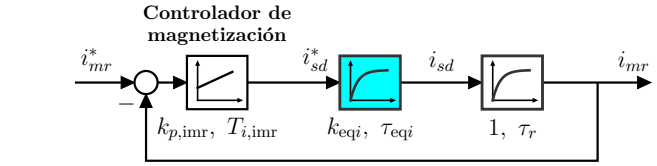
La planta tiene dos retardos de primer orden: uno lento, asociado a la máquina, y otro rápido, asociado al convertidor. La relación $\tau_\sigma/\tau_c \simeq 14{,}7$ supera cuatro, por lo que también sería admisible el óptimo simétrico realista (OSR). Se adopta el óptimo del módulo (OM), tomando como referencia el ajuste del lazo de corriente empleado por David Barrios [1, pp. 62–63, ecs. (4.23)–(4.24)]. Con este criterio, el cero del controlador cancela el polo del retardo lento y la ganancia se ajusta para obtener un amortiguamiento $\zeta = 1/\sqrt{2}$ en el modelo de diseño,

$$T_{i,i} = \tau_\sigma, \qquad k_{p,i} = \frac{\tau_\sigma R_\sigma}{2\,\tau_c} = \frac{\sigma L_s}{2\,\tau_c}, \qquad k_{i,i} = \frac{k_{p,i}}{T_{i,i}}, \tag{5.3}$$

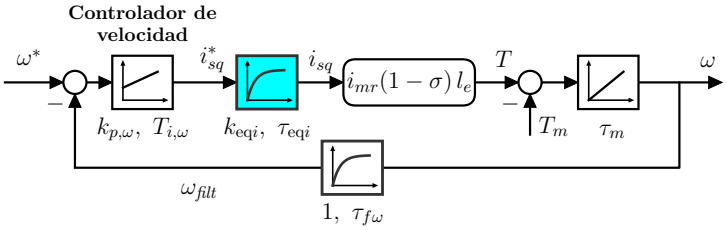
de donde $T_{i,i} = 7{,}3727$ ms, $k_{p,i} = 13{,}1315$ V/A y $k_{i,i} = 1781{,}1$ V/(A·s). Con ese ajuste la transferencia de lazo abierto se reduce a $1/[2\tau_c s\,(1+\tau_c s)]$ y el lazo cerrado es de segundo orden. Para la sintonía de los lazos externos se aproxima por un primer orden de constante $\tau_{eqi} = 2\,\tau_c = 1{,}0$ ms.

5.3. Ajuste de los lazos externos

La Figura 5.2 representa los dos lazos externos, que generan las referencias del interno y ven su planta en serie con el lazo cerrado de corriente. El de flujo se apoya en la dinámica de rotor de (2.4); el de velocidad, en la producción de torque de (2.2) y el integrador de la ecuación mecánica (2.3).



(a) flujo



(b) velocidad

Figura 5.2: En (a), la referencia de corriente de magnetización se compara con la que entrega el estimador de flujo; en (b), la velocidad mecánica medida, filtrada con constante $\tau_{f\omega}$, con su referencia. Con el flujo regulado en 1 pu el torque es proporcional a i_{sq}. Bajo cada controlador se indican su ganancia y su tiempo integral. Los dos esquemas omiten el limitador de pendiente y las saturaciones, activos en el equipo. Elaboración propia.

Los dos lazos externos ven plantas distintas y les corresponden criterios distintos. Con el parámetro a del óptimo simétrico, que fija a la vez el cero y la ganancia [1, 27],

$$T_i = a^2\,\tau_{eq}, \qquad k_{p,\mathrm{imr}} = \frac{\tau_r}{a\,\tau_{eq}}, \qquad k_{p,\omega} = \frac{1}{a\,k_\omega\,\tau_{eq}}, \tag{5.4}$$

donde τ_{eq} es la suma de las constantes rápidas que ve cada lazo y $k_\omega = k_r/(2H) = 4{,}946\ \mathrm{s}^{-1}$ la ganancia de la planta mecánica en por unidad. Se emplea $a = 2$, la forma canónica del criterio, con $T_i = 4\,\tau_{eq}$ [1, 27].

El lazo de flujo ve la constante de rotor, $\tau_r = 193{,}39$ ms, en serie con el lazo cerrado de corriente, de constante τ_{eqi}. Son dos retardos de primer orden, y en principio les correspondería el OM. Su cociente, $\tau_r/\tau_{eqi} \simeq 193$, supera con creces cuatro, y el retardo de rotor actúa como una componente de acción integral en la escala de tiempo del lazo. Por tanto, se adopta el OSR, que regula con mayor rapidez las perturbaciones del flujo [1, 27]. Con $\tau_{eq} = \tau_{eqi}$, (5.4) entrega $T_{i,\mathrm{imr}} = 4$ ms y $k_{p,\mathrm{imr}} = 96{,}6955$. La referencia de flujo cambia una sola vez, en la magnetización, y ese escalón exige una corriente del eje directo muy superior

a su límite de 13,54 A: la salida del controlador satura y el antienrollamiento determina la subida. La maniobra es de gran señal y no reproduce la respuesta lineal del criterio [1], de modo que su sobrepaso no se compara con el teórico. El control convencional no alisa esta referencia, y el flujo se establece con un sobrepaso inferior al 1 % (Tabla 7.2).

El lazo de velocidad ve el integrador de la ecuación mecánica en serie con el lazo cerrado de corriente y con el filtro de primer orden de la medida de velocidad, y le corresponde el OS [1, 27]. El filtro, de corte nominal en 80 Hz y constante $\tau_{f\omega} = 1{,}989$ ms, permanece dentro del lazo, y su constante se suma a la del lazo interno en el ajuste [1]; su frecuencia de corte se fijó por separación de bandas y un barrido del polo, según la Sección A.4.6. Con $\tau_{eq} = \tau_{eqi} + \tau_{f\omega} = 2{,}989$ ms, el ajuste teórico resulta $T_{i,\omega} = 11{,}96$ ms y $k_{p,\omega} = 33{,}82$.

Ensayado sobre el banco el 11 de septiembre de 2026, ese ajuste resulta demasiado agresivo para la planta experimental. La magnetización y el establecimiento del flujo se mantienen operativos; con la máquina en giro, en cambio, la referencia de corriente de cuadratura entra en un ciclo límite sostenido de 50 a 65 Hz que la corriente medida sigue, y lo mismo ocurre con el criterio aplicado sin el filtro. Por tanto, el lazo de velocidad se hace más lento y se implementa con una sintonía común a los dos accionamientos, conservadora y validada sobre el banco, la misma del agente: $k_{p,\omega} = 8{,}4243$ con $T_{i,\omega} = 48$ ms, la que entrega (5.4) con $\tau_{eq} = 12$ ms, adoptada a partir de una identificación en simulación realizada durante el desarrollo del agente, según el Apéndice A.4, sin sumar la del filtro. No es, por tanto, el óptimo simétrico del control convencional. Con el filtro dentro del lazo, esa sintonía deja un margen de fase de 58° con el control convencional y de 33° con el agente, frente a los 37° que el óptimo simétrico asigna con $a = 2$. La sintonía adoptada deja el lazo de velocidad idéntico en las dos campañas, de modo que no añade una diferencia entre ellas, con una respuesta más lenta que la del ajuste teórico. La Sección A.4.6 recoge los ensayos con sus cifras. El PI de flujo, en cambio, no es común: con el agente se utiliza otro ajuste. El agente sustituye sólo el control interno de corriente y no controla directamente el flujo. La Tabla A.10 reúne la configuración de cada caso.

5.4. Implementación de la referencia convencional

El PI se calcula primero en tiempo continuo, con los ajustes de las secciones anteriores. Para ejecutarlo en el procesador se discretiza mediante Euler hacia adelante: en cada período T_s se compara la referencia con la señal realimentada, se calcula la orden de control y se actualiza la integral acumulando el error [28, pp. 11–12, ecs. (22)–(25)]. La parte proporcional responde al error actual; la integral corrige el error que se mantiene con el tiempo. Euler hacia adelante permite calcular esa integral en el programa; el ZOH describe cómo se mantiene la orden aplicada entre actualizaciones.

La salida de cada PI se limita al rango permitido por su bloque. Si la orden calculada supera ese rango, el controlador corrige su integral para evitar que siga acumulando una orden que no puede aplicar. Esta protección se denomina antienrollamiento y se describe en el apunte de tiempo discreto [28, pp. 17–18, Fig. 10]. La corrección considera el límite del propio PI; los límites posteriores de corriente y modulación actúan por separado.

En el equipo se añaden compensaciones a las salidas de los PI de corriente. Dos términos compensan el acoplamiento entre los ejes d y q; otro aporta la tensión asociada al flujo del rotor en el eje q. Se calculan con las corrientes medidas y las señales del estimador. Su finalidad es reducir la corrección que deben aportar los PI. Las ganancias se conservan

como fueron calculadas para el modelo de diseño, aunque la respuesta del equipo incluye estas acciones adicionales y no corresponde a una cancelación exacta de ese modelo.

También hay una diferencia en la conversión de la orden de tensión a modulante: la normalización y el modulador emplean escalas distintas, de modo que la tensión media aplicada queda multiplicada por $\sqrt{3}/2$ respecto de la orden. Este factor se mantuvo en el programa ensayado y modifica la ganancia efectiva del lazo de corriente.

En esta memoria, el control convencional es este PI implementado, con su normalización de tensión, el factor $\sqrt{3}/2$ y sus compensaciones. Las curvas del control convencional del Capítulo 6 corresponden al modelo de esa implementación, y las del Capítulo 7, al programa ensayado.

El modelo del PI del Capítulo 6 reproduce las ganancias, las compensaciones y los límites del lazo de corriente del programa ensayado, incluido el factor $\sqrt{3}/2$ en la tensión aplicada. Conserva las ganancias de los lazos externos de este capítulo: el PI de flujo convencional y el PI de velocidad común, e incluye un filtro de primer orden de 80 Hz nominales en la realimentación del PI de velocidad, como el modelo del agente. Utiliza un *DC link* ideal de 300 V y una medida de velocidad sin cuantización. El detalle de la implementación y las diferencias con el equipo se presentan en el Apéndice A.4.2.

5.5. Alcance de la referencia convencional

La referencia convencional queda definida por estos criterios y coeficientes, con las diferencias entre modelo y equipo declaradas en la sección anterior. En el modelo empleado para ajustar los PI, la acción integral permite eliminar el error en estado estacionario ante perturbaciones constantes, siempre que el lazo sea estable y no permanezca saturado. La respuesta del lazo de corriente no se verificó mediante un ensayo aislado de pequeña señal: durante el arranque y la inversión su referencia la genera el lazo de velocidad, mientras que durante la magnetización aumenta la corriente del eje directo y actúan los límites del controlador. Esta diferencia debe considerarse al interpretar los transitorios del Capítulo 7.

Capítulo 6

Transferencia a la máquina de inducción

La transferencia se evalúa primero sobre el modelo de la máquina de inducción de 5,5 kW. El Agente 2 refinado mantiene los pesos seleccionados en el entorno $R-L$ perturbado y controla el lazo interno de corriente. Se compara su respuesta con la del control convencional y se evalúa la reflexión del eje de cuadratura durante la inversión de marcha.

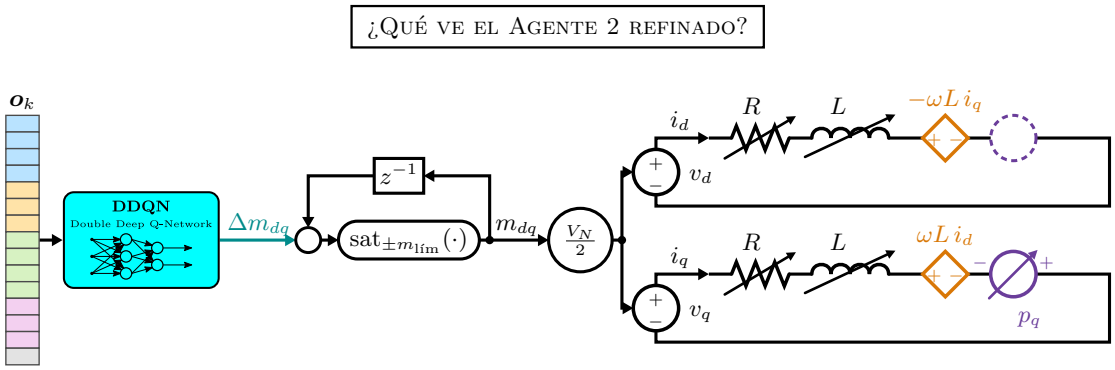
Entre los dos modelos también cambia el ajuste del lazo de flujo. La comparación describe las implementaciones completas y no permite atribuir todas sus diferencias al controlador de corriente.

6.1. Integración del Agente 2 refinado en el IFOC

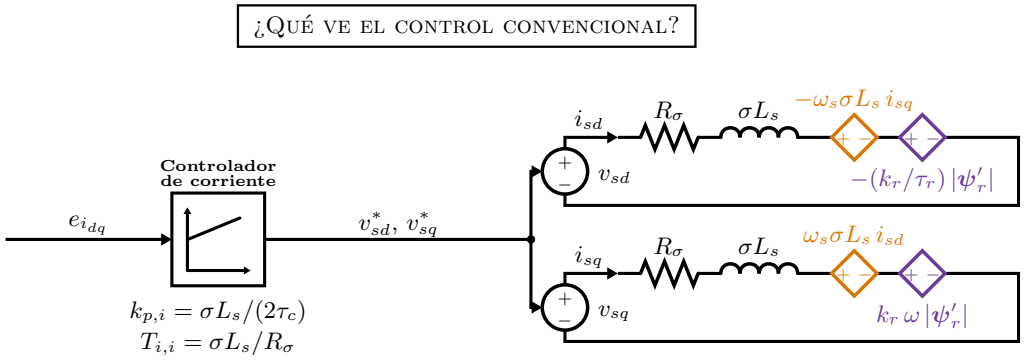
La transferencia comprueba si el seguimiento aprendido sobre la carga $R-L$ se conserva al controlar una máquina, sin entregar a la red sus ecuaciones de flujo de rotor ni su dinámica mecánica. Para preparar al agente frente a una mayor exigencia de tensión, el refinamiento añadió una perturbación en q sin cambiar las matrices de la carga (Sección 3.2). Durante el entrenamiento la red recibe las señales del entorno $R-L$; en el accionamiento recibe las variables correspondientes, expresadas con la misma observación. Los pesos no se actualizan al controlar la máquina. Por ello, la dificultad evaluada no es volver a aprender una planta conocida, sino conservar la política al conectarla a una dinámica que no formó parte del entrenamiento.

La Figura 6.1 compara el entorno de entrenamiento del agente con el modelo utilizado para ajustar el control convencional. Ambos presentan términos resistivos, inductivos y de acoplamiento entre ejes, pero la máquina incorpora además las tensiones asociadas al rotor. Esta semejanza permite explicar el fundamento de la transferencia; su funcionamiento se verifica mediante simulación y ensayos, manteniendo los pesos de la red fijos.

Durante la transferencia se conservan la orientación de campo, el estimador de flujo, los lazos externos de flujo y velocidad y la modulación descritos en el Capítulo 2, y el agente controla el seguimiento de corriente del lazo interno. La Tabla 6.1 identifica los bloques adicionales y su función.



(a)



(b)

Figura 6.1: Qué ve cada controlador en el lazo de corriente, como circuito equivalente con el eje d arriba y el q abajo. (a) El Agente 2 refinado recibe $\boldsymbol{o}_k$, con sus cinco grupos de la Sección 3.4 en colores, y su acción se acumula, se limita y se expresa en tensión con $V_N/2$, como en la Figura 3.4. La planta es la carga $R-L$ de (3.1); las flechas diagonales marcan lo que se sortea en cada episodio, con p_q de (A.6), y el círculo punteado, que el eje d no tiene tensión adicional. (b) Al control convencional entra el error de corriente y sale la tensión de referencia. La planta es la máquina de la Sección 2.1.1 en el marco orientado al flujo de rotor, con las tensiones del rotor de (3.8) y las ganancias de (5.3). En naranja, el acoplamiento entre ejes; en morado, la tensión del rotor y su sustituto p_q. Elaboración propia.

Tabla 6.1: Elementos añadidos al pasar de la carga $R-L$ a la máquina, y la razón de cada uno. Ninguno forma parte del Agente 2 refinado.

elemento	por qué
Lazos externos de flujo y velocidad	Generan las referencias de corriente a partir del flujo y la velocidad. Se sintonizan sobre la constante de tiempo equivalente del lazo interno del agente, de 12 ms identificados en simulación, y no sobre la del controlador PI (Capítulo 5). El de velocidad es común a ambos controles; el de flujo es distinto (Sección A.4.7).
Rampa en la referencia de velocidad	Sin ella el máximo de arranque alcanza 18,3 A. Limita la exigencia, no el error.
Filtro de la medida de velocidad	Suaviza las variaciones rápidas de la velocidad medida antes de calcular el error del PI de velocidad. Se usa un filtro de primer orden de 80 Hz nominales, elegido mediante pruebas con distintas frecuencias de corte (Sección A.4.6).
Límites de la modulante	Evitan pedir una tensión mayor que la permitida. El agente conserva primero la componente del eje directo y limita la del eje de cuadratura según el margen disponible. En el PI se limita el módulo de la orden total a uno. Después se aplica Min–Max en ambos casos.
Reflexión del eje de cuadratura	Explota una simetría que las ecuaciones del modelo cumplen de forma exacta, para la inversión de marcha (Sección 6.3).

Conviven dos normalizaciones de la modulante. El agente decide en la escala del entorno, referida a la base fija V_N del Capítulo 3, y el accionamiento expresa la orden respecto de la tensión de enlace medida. Sus incrementos se convierten con el cociente entre ambas bases, de modo que cada paso aplica la misma tensión con que el agente entrenó. En el modelo, la orden del agente se limita mediante $|m_d| \leq 2/\sqrt{3}$ y, con prioridad del eje directo, $m_d^2 + m_q^2 \leq 4/3$, el círculo inscrito en el hexágono del inversor [6] que describe la Sección 2.2.2. La orden total del PI se limita en módulo a uno antes de aplicar Min–Max, como en el *firmware* ensayado. Este recorte es distinto del límite lineal de la modulación; Min–Max sigue activo en ambos modelos. En el programa del banco, el estado del agente conserva el límite $\pm m_{\text{lím}}$ por eje del entrenamiento, expresado en la base $V_N/2$ y equivalente a $\pm V_{dc}/2$. Después, ya en la base $V_{dc}/2$, la orden se limita en módulo a uno. Tras Min–Max, cada ciclo de trabajo se recorta entre el 12,5 y el 87,5 %. Ese recorte, común a los dos controles, limita finalmente la orden realizable a $0{,}433\,V_{dc}$, que son 122 V con 282 V. La Sección A.6 ordena estas operaciones en la Tabla A.12 y da un paso numérico. La Figura 6.2 muestra dónde se incorpora el agente y qué operaciones permanecen fuera de la red.

Durante el desarrollo del modelo se comprobó cómo cambia la respuesta del agente ante una diferencia pequeña en la simulación. La prueba se hizo el 12 de agosto de 2026, con un agente anterior del linaje y escalones de carga hasta el torque nominal. Se modificó el antienrollamiento del PI de flujo, de modo que su salida difería en una muestra en cada escalón de i^*_{mr}. Las corrientes de las dos simulaciones empezaron a diferir en $t = 0{,}2737$ s, con 5,4 mA, y la diferencia alcanzó 1 A en 89 ms. Los errores medios de los dos ejes cambiaron un 4,5 %. Ese resultado corresponde al caso evaluado y no es un umbral general de incertidumbre. Muestra que diferencias de ese orden en los indicadores no bastan por sí solas para afirmar que una respuesta es mejor.

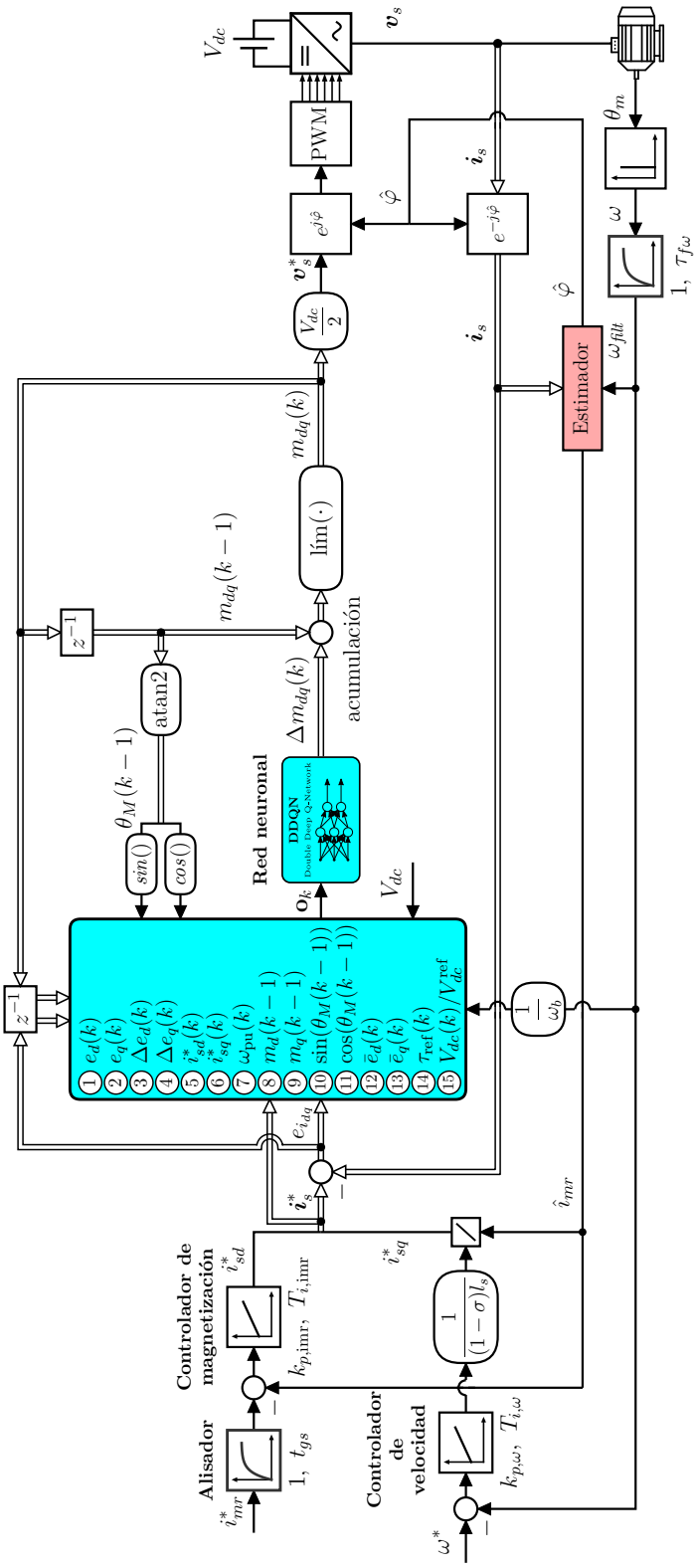


Figura 6.2: Esquema funcional del accionamiento. Los lazos PI de flujo y velocidad generan las referencias de corriente; el Agente 2 refinado recibe quince componentes y selecciona un incremento de la modulante acumulada. El alisador, de constante $t_{gs} = 60$ ms, actúa sólo con el agente sobre la referencia i^*_{mr}. Bajo cada controlador se indican su ganancia y su tiempo integral. El nodo de acumulación suma el incremento $\Delta m_{dq}(k)$ con la modulante anterior $m_{dq}(k-1)$. Después del límite, el resultado $m_{dq}(k)$ se conserva para el período siguiente, según (3.14). Los límites y cambios de base se detallan en la Tabla A.12. Elaboración propia.

6.2. Comparación en simulación con el control convencional

Las Figuras 6.3 y 6.4 presentan la magnetización, el arranque y la inversión de marcha del control convencional y del Agente 2 refinado, respectivamente.

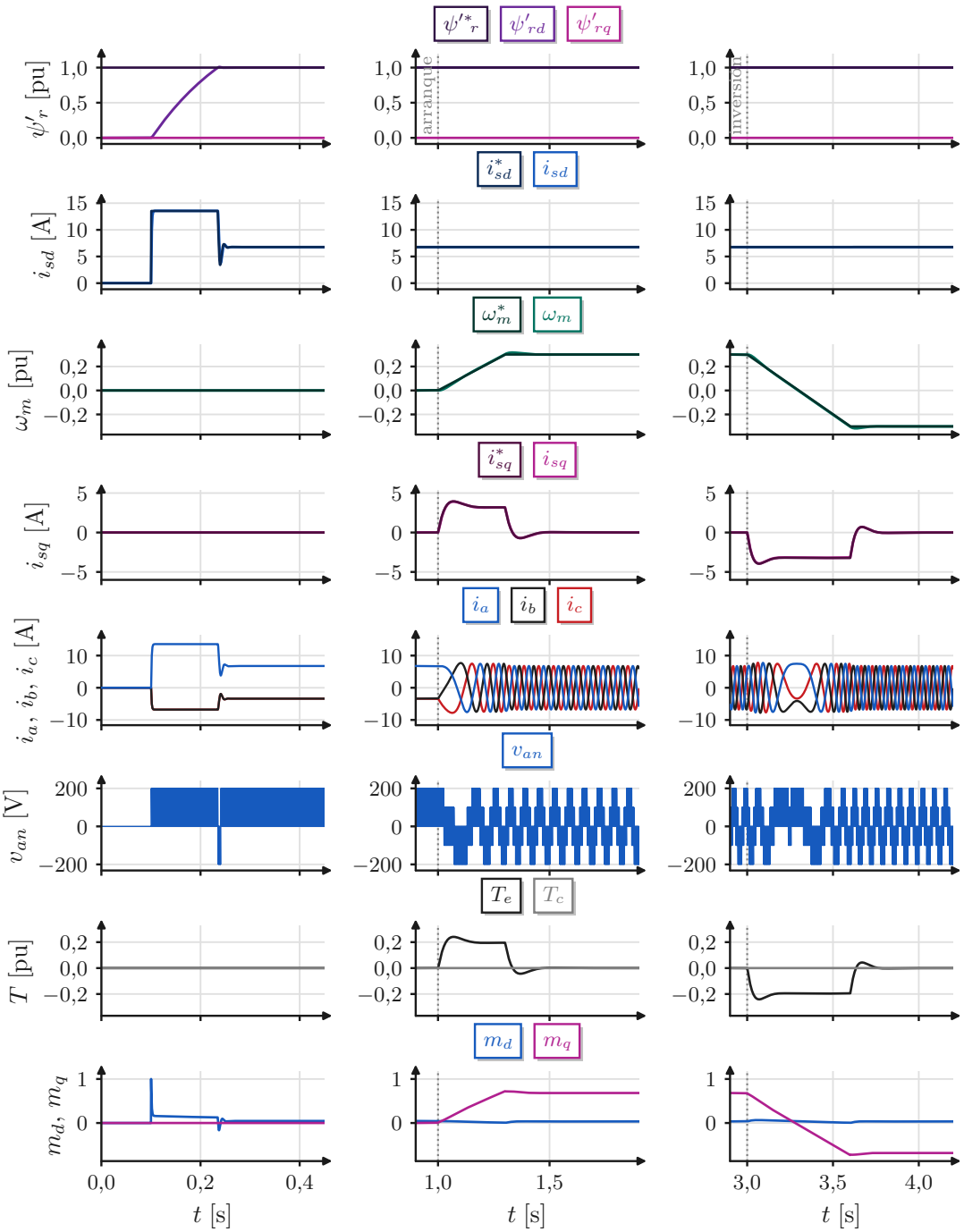


Figura 6.3: Maniobra simulada en PLECS con el control convencional: magnetización, arranque e inversión de marcha. Sin carga mecánica, *DC link* de 300 V y referencia de ±0,30 pu con rampa de 1 pu/s. Filas y escalas comunes con la Figura 6.4.

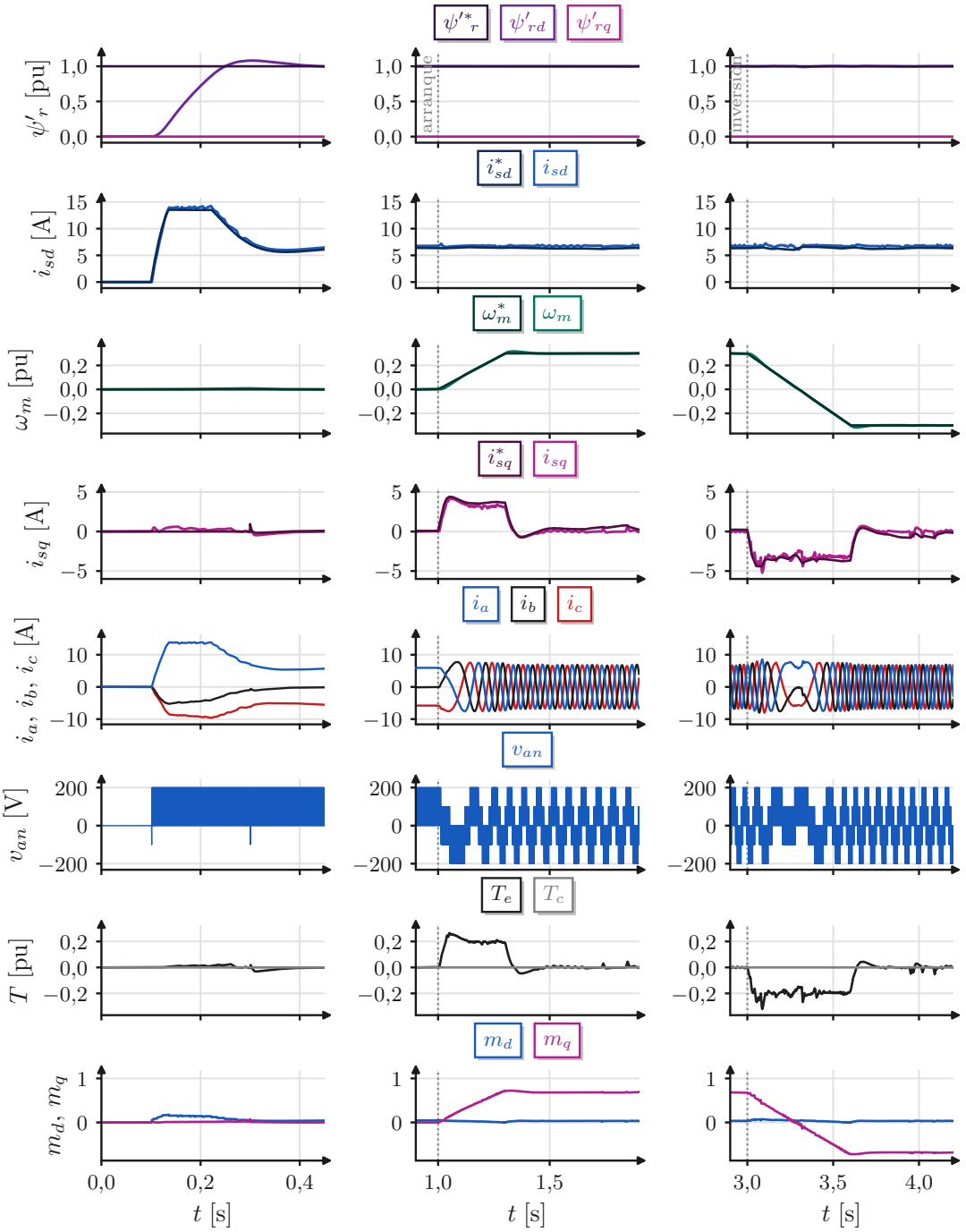


Figura 6.4: La misma maniobra de la Figura 6.3, con el Agente 2 refinado en el lazo interno y la reflexión del eje de cuadratura activa tras la inversión de marcha. Las condiciones, las filas y las escalas son las mismas que en la figura del control convencional.

Ambos modelos comparten el PI de velocidad, el filtro de su medida de 80 Hz nominales y la rampa de referencia de 1 pu/s. Esta rampa evita la demanda inicial de 2,53 pu de corriente que produciría un escalón de 0,30 pu. El PI de flujo usa los parámetros del Capítulo 5 en el control convencional y los del Apéndice A.4 con el agente, cuya reflexión del eje de cuadratura está activa con velocidad negativa. Con el agente, ese PI usa la mitad de las ganancias de su diseño, $k_{p,\mathrm{imr}} = 4{,}0290$ con $T_{i,\mathrm{imr}} = 48$ ms, y su referencia se alisa con 60 ms. El agente sustituye el control interno de corriente y no controla directamente el flujo. La Tabla A.10 reúne la configuración de ambos modelos.

La comparación abarca 4,5 s; las figuras amplían las tres etapas hasta 4,20 s. El control opera a 3 kHz y las tensiones se representan por período PWM. Las corrientes de fase se reconstruyen a partir de las corrientes del marco sincrónico utilizadas por el control.

La Tabla 6.2 compara los transitorios de corriente, flujo y velocidad durante la magnetización y el arranque. El tiempo de subida se mide entre los primeros cruces del 10 y el 90 % del recorrido; el asentamiento exige permanecer en la banda de ±5 % hasta el final de la ventana evaluada.

Tabla 6.2: Transitorios de la magnetización y del arranque. Asentamiento en ±5 %. Las ventanas, los valores finales y el procesamiento se detallan en la Sección B.1.

Indicador	Unidad	Control convencional	Agente 2 refinado
Corriente del eje directo: magnetización			
Tiempo de subida (10–90 %)	ms	2,19	24,49
Tiempo de asentamiento (±5 %)	ms	3,33	121,67
Sobrepaso	%	0,26	5,28
Corriente del eje de cuadratura: arranque			
Tiempo de subida (10–90 %)	ms	25,65	23,03
Tiempo de asentamiento (±5 %)	ms	142,33	267,33
Sobrepaso	%	23,54	32,33
Flujo de rotor estimado: magnetización			
Tiempo de subida (10–90 %)	ms	105,32	102,46
Tiempo de asentamiento (±5 %)	ms	125,67	257,33
Sobrepaso	%	0,93	8,18
Velocidad mecánica: arranque			
Tiempo de subida (10–90 %)	ms	224,14	222,63
Tiempo de asentamiento (±5 %)	ms	349,33	359,67
Sobrepaso	%	5,46	5,68

La respuesta de corriente del eje directo incluye el alisamiento aplicado con el agente. En el eje de cuadratura, la referencia depende del lazo de velocidad: estos indicadores describen el transitorio de la maniobra y no un ensayo aislado del control de corriente. La Figura 6.5 amplía esas señales en las ventanas de la Tabla 6.2.

El agente presenta mayor sobrepaso y tiempo de asentamiento del flujo, con un tiempo de subida similar. Estos resultados incluyen las diferencias de ajuste del lazo de flujo y el alisamiento aplicado con el agente. Las respuestas de velocidad son similares. Sus tiempos dependen principalmente del limitador de pendiente de 1 pu/s aplicado a la referencia de velocidad y no permiten comparar por sí solos la rapidez del control de corriente.

La Tabla 6.3 resume el seguimiento. El sesgo es el promedio de referencia menos corriente medida, el recorrido *peak to peak* es la diferencia entre los valores máximo y mínimo de la corriente registrada, y la desviación estándar de la corriente lo acompaña. Los tres

se calculan en la ventana de estado estacionario, los últimos 200 ms de cada tramo de velocidad constante, con el mismo criterio en los dos controles. La Figura 6.6 señala las dos ventanas sobre la velocidad y muestra en ellas el error de corriente de los dos controles.

Tabla 6.3: Resultados de la simulación hasta 4,5 s. Los indicadores de corriente se evalúan durante los últimos 200 ms de cada tramo de velocidad constante: 2,80–3,00 s a +0,30 pu y 4,30–4,50 s a −0,30 pu. Se utiliza el mismo criterio que en el Capítulo 7.

Indicador	Unidad	Control convencional	Agente 2 refinado
Arranque: tiempo hasta +0,30 pu	ms	298	298
Inversión de marcha: tiempo hasta −0,30 pu	ms	598	598
Primera entrada del flujo en ±0,2 % del nominal	s	0,235	0,249
Velocidad máxima absoluta en la magnetización	pu	0	0,0087
Velocidad de +0,30 pu			
Sesgo de i_{sq}	A	0,000	+0,233
Valor absoluto del sesgo de i_{sd}	A	0,000	0,436
Recorrido *peak to peak* de i_{sq}	A	0,000	0,209
Desviación estándar de i_{sq}	A	0,000	0,065
Velocidad de −0,30 pu			
Sesgo de i_{sq}	A	0,000	−0,235
Valor absoluto del sesgo de i_{sd}	A	0,000	0,442
Recorrido *peak to peak* de i_{sq}	A	0,001	0,212
Desviación estándar de i_{sq}	A	0,000	0,064

Ambos controles completan la maniobra. Los tiempos de velocidad dependen principalmente de la rampa y no permiten comparar por sí solos la rapidez del control de corriente. El control convencional presenta sesgos próximos a cero; el agente presenta mayor sesgo y mayor recorrido de la corriente, de magnitud similar en los dos sentidos de giro. Los PI de flujo y velocidad compensan ese sesgo ajustando las referencias, sin eliminar el error del lazo interno.

Antes de las ventanas evaluadas, la corriente del eje de cuadratura del agente presenta *peaks* de +0,723 y −0,861 A durante el ajuste posterior al arranque y a la inversión de marcha. Estos *peaks* forman parte del transitorio y se analizan por separado del recorrido de la corriente en estado estacionario. En ambas figuras, las ampliaciones llegan hasta 1,90 s en el arranque y 4,20 s en la inversión de marcha.

Se opera a ±0,30 pu de velocidad para mantener margen de tensión. Después del breve transitorio inicial de magnetización del control convencional, el módulo de la modulante no supera aproximadamente 0,73 en ambos modelos, por debajo del límite lineal de 1 correspondiente a la modulación sin inyección Min–Max. Min–Max permanece activo en ambos casos. El torque mostrado es el estimado por el control.

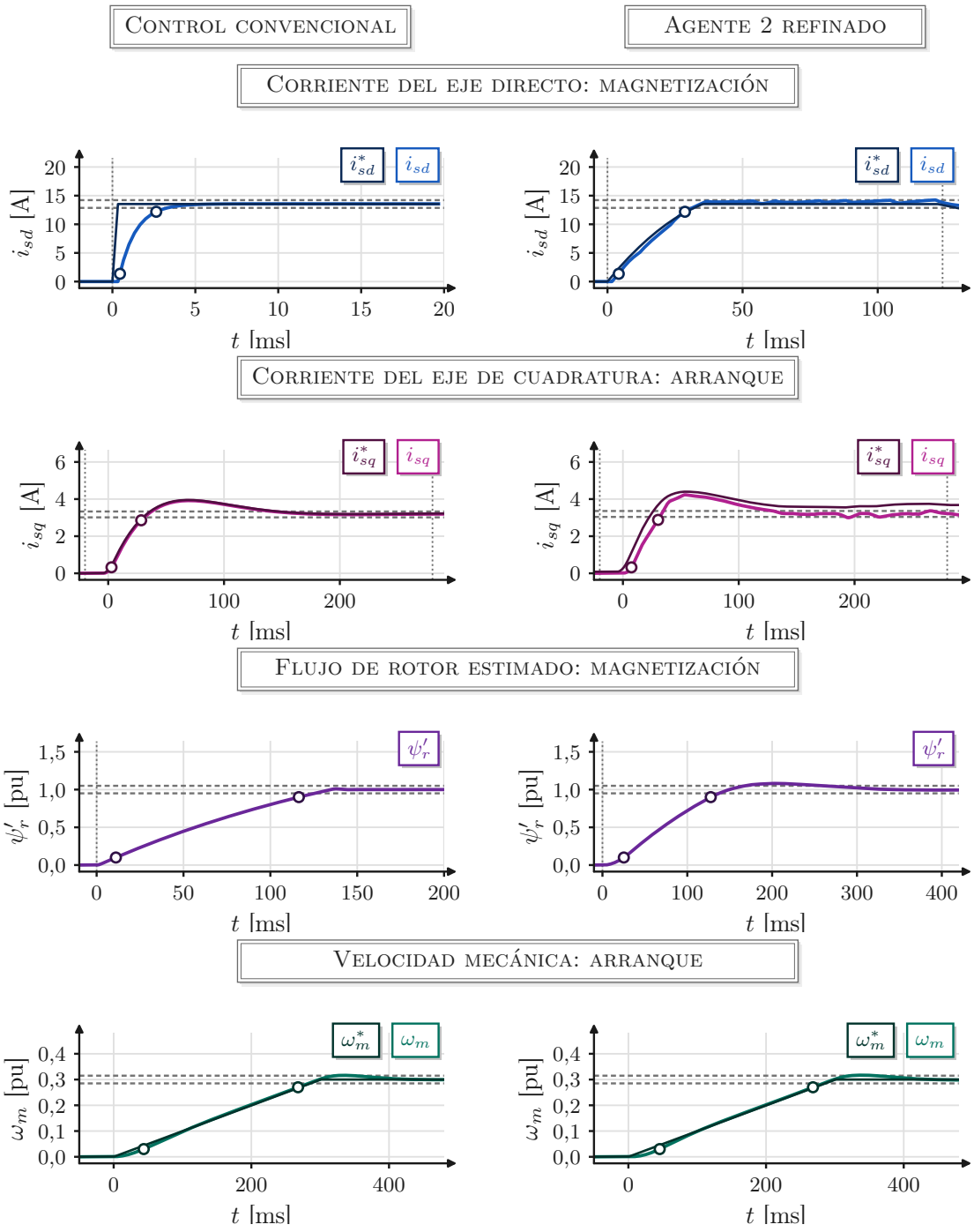


Figura 6.5: Transitorios simulados de la Tabla 6.2. Cada panel muestra la señal evaluada y, en tono oscuro, su referencia. El eje del tiempo parte del evento de cada fila y se ajusta a cada transitorio: es más corto para el control convencional en i_{sd} y el flujo. Las escalas verticales son comunes por fila. Los círculos marcan los cruces del 10 y el 90 % del recorrido, y las horizontales a trazos, la banda de ±5 % del valor final. La señal se asienta al entrar en esa banda para no salir; si el sobrepaso es menor que el 5 %, como en i_{sd} y el flujo del control convencional, eso ocurre antes del máximo. Las verticales punteadas marcan los límites visibles de la ventana, detallada en la Sección B.1. Se muestran i_{sq} y la velocidad promediadas en 10 y 20 ms. Condiciones: las de la Figura 6.3.

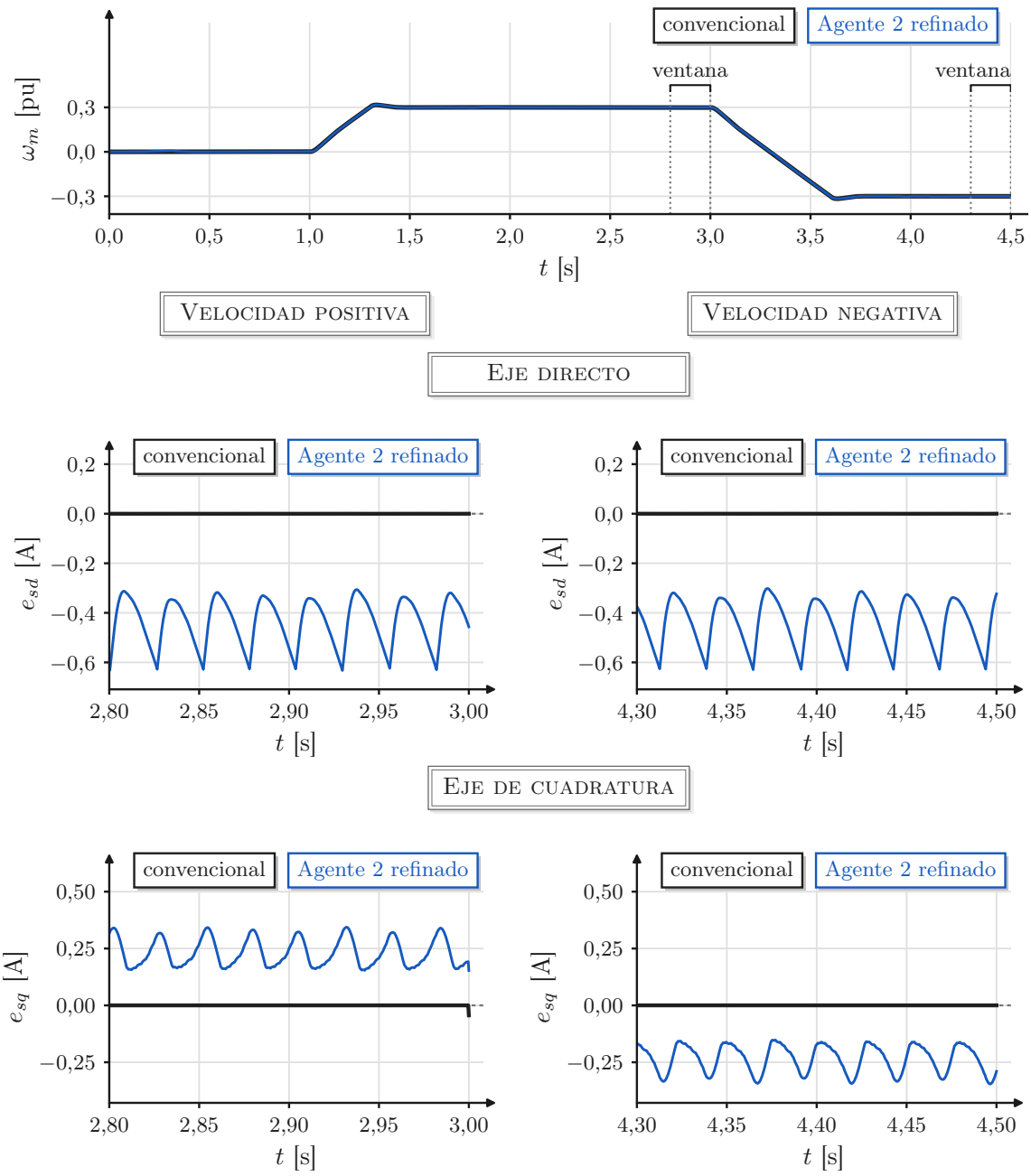


Figura 6.6: Errores de corriente simulados en las ventanas de la Tabla 6.3. Arriba, la velocidad de la maniobra completa; las de los dos controles casi coinciden. Las ventanas son los últimos 200 ms de cada tramo de velocidad constante: de 2,80 a 3,00 s, a +0,30 pu, y de 4,30 a 4,50 s, a −0,30 pu. Abajo, $e_{sd} = i^*_{sd} - i_{sd}$ y $e_{sq} = i^*_{sq} - i_{sq}$ en esas ventanas, con escala común por fila. La media de e_{sq} es el sesgo de i_{sq} de la Tabla 6.3, y el valor absoluto de la media de e_{sd}, el del sesgo de i_{sd}. Condiciones: las de la Figura 6.3.

6.3. Reflexión del eje de cuadratura

La Figura 6.7 compara la inversión de marcha con y sin reflexión del eje de cuadratura, en las mismas condiciones y con velocidad negativa mantenida hasta los 8 s.

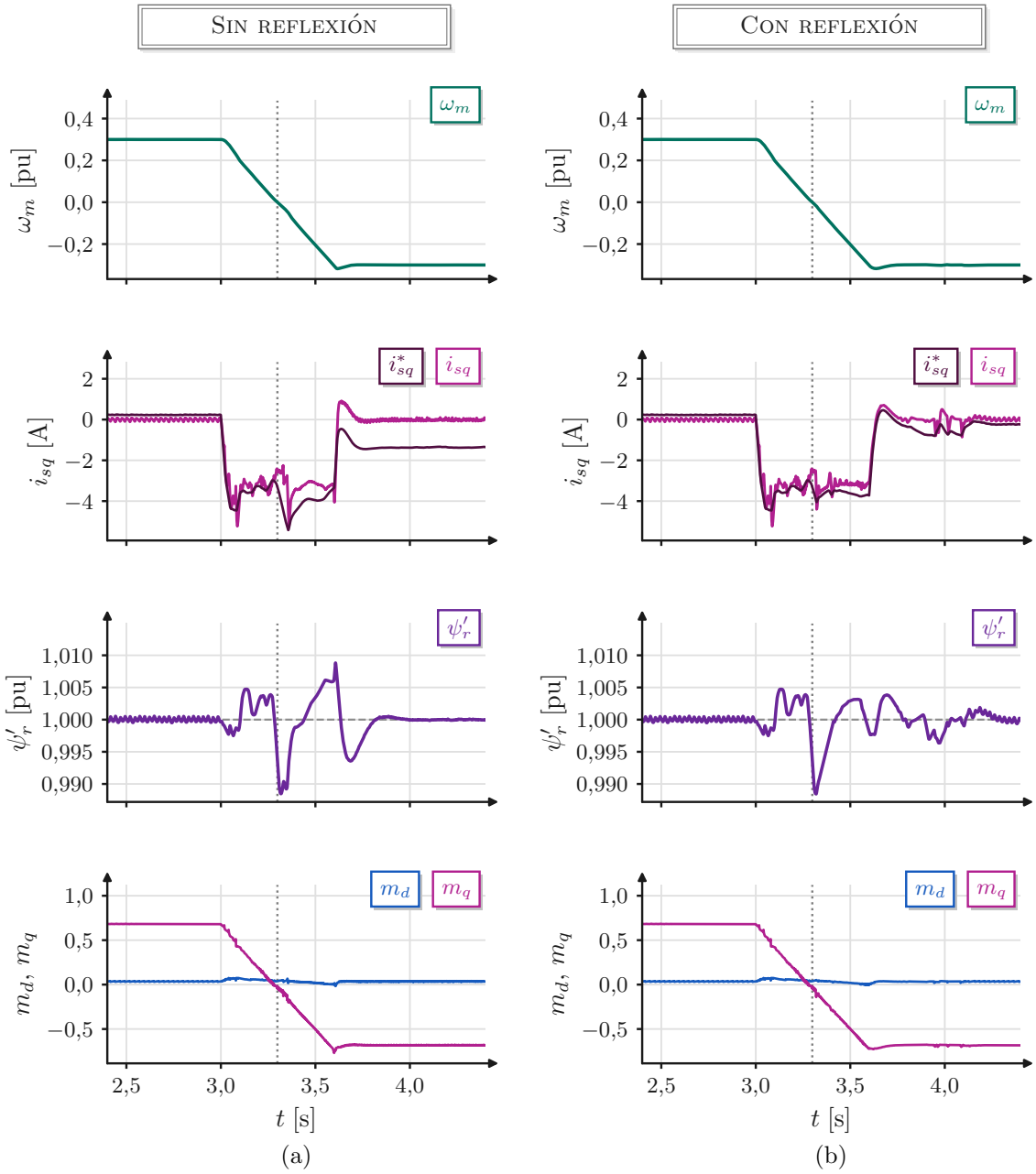


Figura 6.7: Inversión de marcha en PLECS: (a) sin reflexión; (b) con reflexión, con escalas comunes entre columnas. Condiciones: las de la Figura 6.3. La vertical marca el paso por velocidad cero. En (b), la reflexión se activa 17 ms después, cuando la entrada ω_{pu} del agente baja de $-0{,}02$ pu.

Sin la reflexión, el agente también completa la maniobra: sigue la rampa con una pendiente de 1,011 pu/s y mantiene el flujo de rotor en el 99,3 % del nominal al cruzar por cero, sin bajar del 98,8 % durante toda la inversión de marcha. Pero el error de seguimiento

con velocidad negativa es casi cinco veces el de velocidad positiva, 1,354 frente a 0,281 A. Una posible explicación sería una pérdida de orientación del marco *dq*. En cambio, los registros internos no muestran una pérdida del flujo estimado, que permanece en 1,000 pu durante todo el tramo, ni del seguimiento del eje directo, que sigue a su referencia como con velocidad positiva; sin una medida independiente del flujo, la asimetría no se atribuye a una pérdida de orientación. La diferencia aparece en el eje de cuadratura, en la Figura 6.7(a). Sin carga y a velocidad constante se espera una corriente media de cuadratura próxima a cero. En la simulación, i_{sq} vale 0,00 A. Su referencia permanece en $-1{,}35$ A durante todo el tramo, porque el controlador de velocidad se desplaza esa cantidad para compensar un sesgo del lazo interno. El Agente 2 refinado no sigue las referencias negativas del eje de cuadratura con la misma precisión que las positivas, y esa asimetría no se atribuye a una ausencia de velocidades negativas en el entrenamiento, porque el dominio final sí las contiene según la Sección 3.6.

Cuando la velocidad es negativa, se cambia el signo de las componentes correspondientes de la observación antes de evaluar al agente. Luego se cambia el signo de la componente de cuadratura de su acción. Los cambios de signo se representan mediante

$$(i_{sd},\, i_{sq},\, v_{sd},\, v_{sq},\, \omega) \;\longrightarrow\; (i_{sd},\, -i_{sq},\, v_{sd},\, -v_{sq},\, -\omega), \tag{6.1}$$

donde i_{sd} e i_{sq} son las corrientes del marco sincrónico, v_{sd} y v_{sq} las tensiones correspondientes y ω la velocidad. Esta operación no modifica los pesos del agente. La reflexión está activa mientras la entrada ω_{pu}, sin reflejar, es menor que $-0{,}02$ pu, sin histéresis. Esa entrada es la frecuencia del marco sincrónico, $(\omega + \omega_{sl})/\omega_b$, y no la velocidad del rotor. En la simulación cruza el umbral una sola vez, en la inversión. Los errores previos y la modulante acumulada se guardan en el marco *dq* sin reflejar, de modo que al activarse la reflexión Δe_q sólo cambia de signo. La red recibe la observación reflejada, y su acción vuelve al marco real antes de acumularse. La Tabla 6.4 detalla qué entradas cambian de signo. Con estas reglas, la red evaluada sobre las observaciones registradas en la simulación repite la acción del modelo en las 14 052 muestras con la reflexión activa. La Tabla 6.5 cuantifica el efecto sobre el error de seguimiento.

Tabla 6.4: Entradas y acción bajo la reflexión del eje de cuadratura, con la numeración de la Figura 6.2. j_d y j_q son los índices de cada eje en la grilla, de 0 a 14.

No cambian	1 e_d, 3 Δe_d, 5 i_d^*, 8 m_d, 11 $\cos\theta_M$, 12 $\bar{e}_d$, 13 $\bar{e}_q$, 14 τ_{ref}, 15 $V_{dc}/V_{dc,\mathrm{ref}}$
Cambian de signo	2 e_q, 4 Δe_q, 6 i_q^*, 7 ω_{pu}, 9 m_q, 10 $\sin\theta_M$
Acción	el índice $15j_q + j_d$ pasa a $15(14 - j_q) + j_d$: Δm_d se conserva y Δm_q cambia de signo

Tabla 6.5: Error de seguimiento con y sin la reflexión, en valor medio absoluto y en amperes. Condiciones: las de la Figura 6.7. Ventanas: velocidad positiva entre 1,60 y 3,00 s, la meseta a +0,30 pu, velocidad negativa entre 4,5 y 8 s; el error global recorre los dos ejes desde que el flujo se asienta. $|i|_{\mathrm{máx}}$ es el máximo del módulo de la corriente de estator.

	$\lvert e_{sq}\rvert$, $\omega > 0$	$\lvert e_{sq}\rvert$, $\omega < 0$	$\lvert e_{sd}\rvert$, $\omega < 0$	global	$\lvert i\rvert_{\mathrm{máx}}$
Sin reflexión	0,281	1,354	0,324	0,647	8,84 A
Con reflexión	0,281	**0,232**	0,439	**0,348**	8,57 A

Con la reflexión, el error del eje de cuadratura con velocidad negativa se reduce 5,8 veces y queda por debajo del de velocidad positiva, y el error global disminuye un 46 %. En

cambio, el error del eje directo con velocidad negativa empeora de 0,324 a 0,439 A, un 35 % más. Por tanto, el error global mejora, aunque el del eje directo aumenta. Esta prueba no permite determinar la causa de ese aumento. La numeración de las acciones identifica pares de incrementos dq; su promedio no representa por sí solo una tensión ni permite explicar la mejora del seguimiento.

Esta maniobra se realizó sin carga mecánica: el módulo de la modulante no supera 0,770, un 67 % del límite lineal de $2/\sqrt{3}$, y con la reflexión su máximo es aún menor, de modo que el límite del módulo de la modulante no actúa en ninguna muestra. La reflexión está implementada también en el programa del banco, y la campaña del Capítulo 7 la lleva activa: su tramo de giro con velocidad negativa es la versión experimental de la Figura 6.7(b).

6.4. Resultado de la transferencia en simulación

El Agente 2 refinado completa la maniobra sobre el modelo de la máquina. Los lazos externos ajustan las referencias de corriente para mantener el flujo y la velocidad. La reflexión del eje de cuadratura reduce el error global en la inversión de marcha, aunque el error del eje directo aumenta; el resultado no representa una mejora uniforme en todas las variables.

La acumulación de (3.14) se parece a un integrador, pero no integra el error de corriente. Cada incremento es el que la red elige entre sus 225 pares de incrementos, y no un valor proporcional al error. En un integrador ideal, la salida sólo deja de cambiar cuando su entrada, el error, es cero. El agente deja quieta la modulante cada vez que elige la acción nula, y puede elegirla con un error distinto de cero si sus valores Q la prefieren. Por eso la acumulación no garantiza eliminar el error en estado estacionario. Ese error puede deberse a la penalización del movimiento después de τ_Δ, al tamaño de los incrementos, a las variables que la red no observa o a las diferencias entre el entorno y la máquina. Estas causas no se aislaron mediante pruebas y se mantienen como hipótesis.

Los ensayos se realizaron sin carga mecánica y no incluyeron escalones de carga. Además, la diferencia en el ajuste del lazo de flujo también influye en la respuesta. Por ello, esta comparación evalúa los accionamientos completos y no únicamente sus controles de corriente. El capítulo siguiente verifica la transferencia sobre el banco. Su campaña utiliza la reflexión, pero no incluye una repetición con ella deshabilitada que permita separar su efecto experimental.

Capítulo 7

Ensayos experimentales en la máquina de inducción

La validación experimental compara el control convencional y el Agente 2 refinado sobre la máquina de inducción del banco. El agente utiliza la política aprendida en el entorno $R-L$ perturbado, con pesos fijos. Se evalúan magnetización, arranque e inversión de marcha mediante el seguimiento de corriente, el flujo estimado y la velocidad. Las mediciones externas contrastan las escalas de tensión y corriente.

7.1. El banco, las maniobras y la procedencia de los registros

Las campañas se realizan sobre la máquina de inducción del banco experimental, utilizando la plataforma de control BRAIn [25]. Se evalúan la magnetización con el eje detenido, el arranque y la inversión de marcha. La unidad de control y dos osciloscopios registran simultáneamente cada maniobra.

La Figura 7.1 muestra el montaje. La máquina de carga permanece acoplada al eje, sin aplicar carga mecánica externa. El *encoder* entrega la posición y la velocidad se calcula a partir de sus incrementos. El ángulo de orientación se obtiene al integrar la velocidad eléctrica y el deslizamiento estimado según (2.4) [2]. La plataforma BRAIn y las interfaces del banco se detallan en la Sección A.1.

7.1.1. Condiciones comunes y diferencias entre las campañas

Ambas campañas utilizan la misma máquina, los mismos instrumentos y el mismo inversor. También mantienen el mismo ajuste del lazo de velocidad, el limitador de pendiente de su referencia y la modulación PWM. El filtro de la medida de velocidad es el mismo. La Tabla 7.1 reúne las condiciones comunes; los ajustes particulares de cada controlador se detallan en la Sección A.4.7. Los registros proceden de sesiones distintas.

La corrección de escala del *encoder* sólo se aplica al ángulo de orientación. El lazo de velocidad, las protecciones, el agente y las figuras usan la medida sin corregir, como detalla la Sección B.5. Sólo se comparan registros posteriores a esta calibración. En cada captura, el tiempo se mide desde el inicio de la maniobra. La procedencia de los registros y los criterios de selección se detallan en los Apéndices B y C.

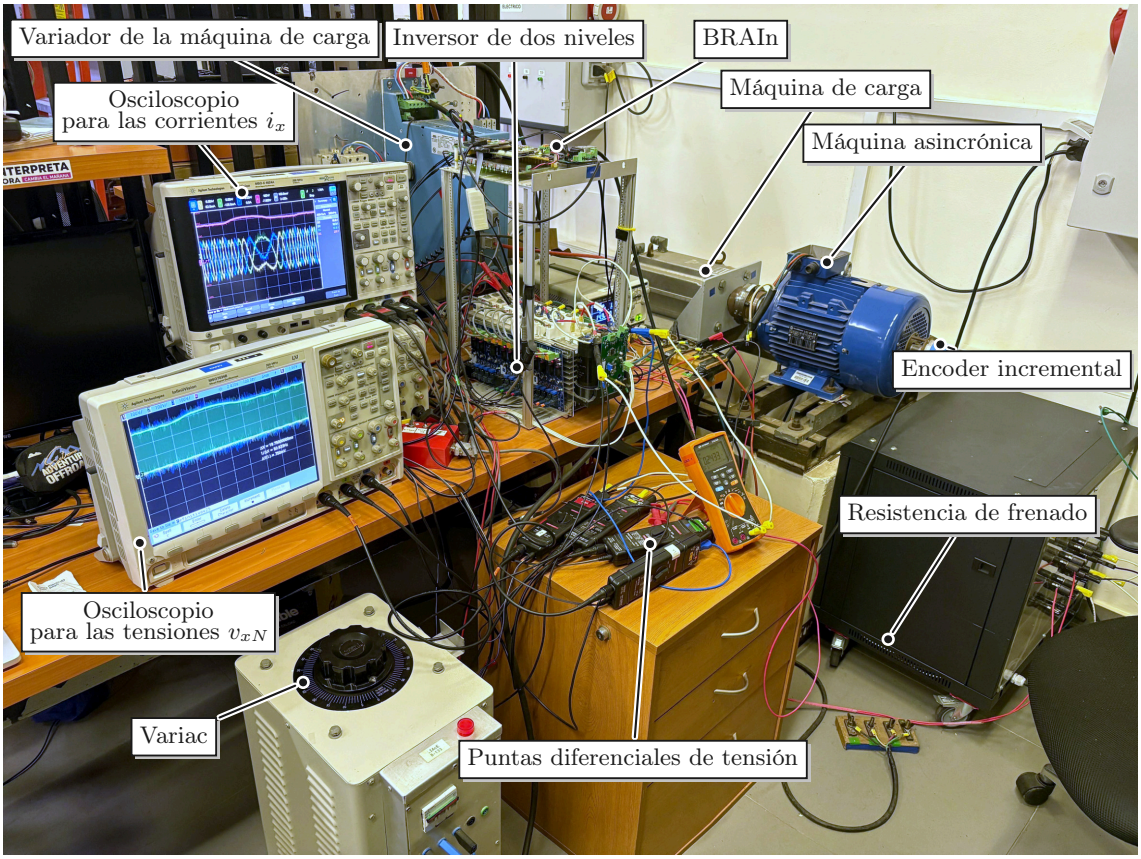


Figura 7.1: Montaje del banco experimental con la plataforma de control BRAIn [25]. Un osciloscopio mide con pinzas las tres corrientes de fase i_x y el otro, con puntas diferenciales, las tres tensiones v_{xN}, medidas entre cada salida $x \in \{a, b, c\}$ del inversor y N, el negativo del enlace. Condiciones: las de la Tabla 7.1.

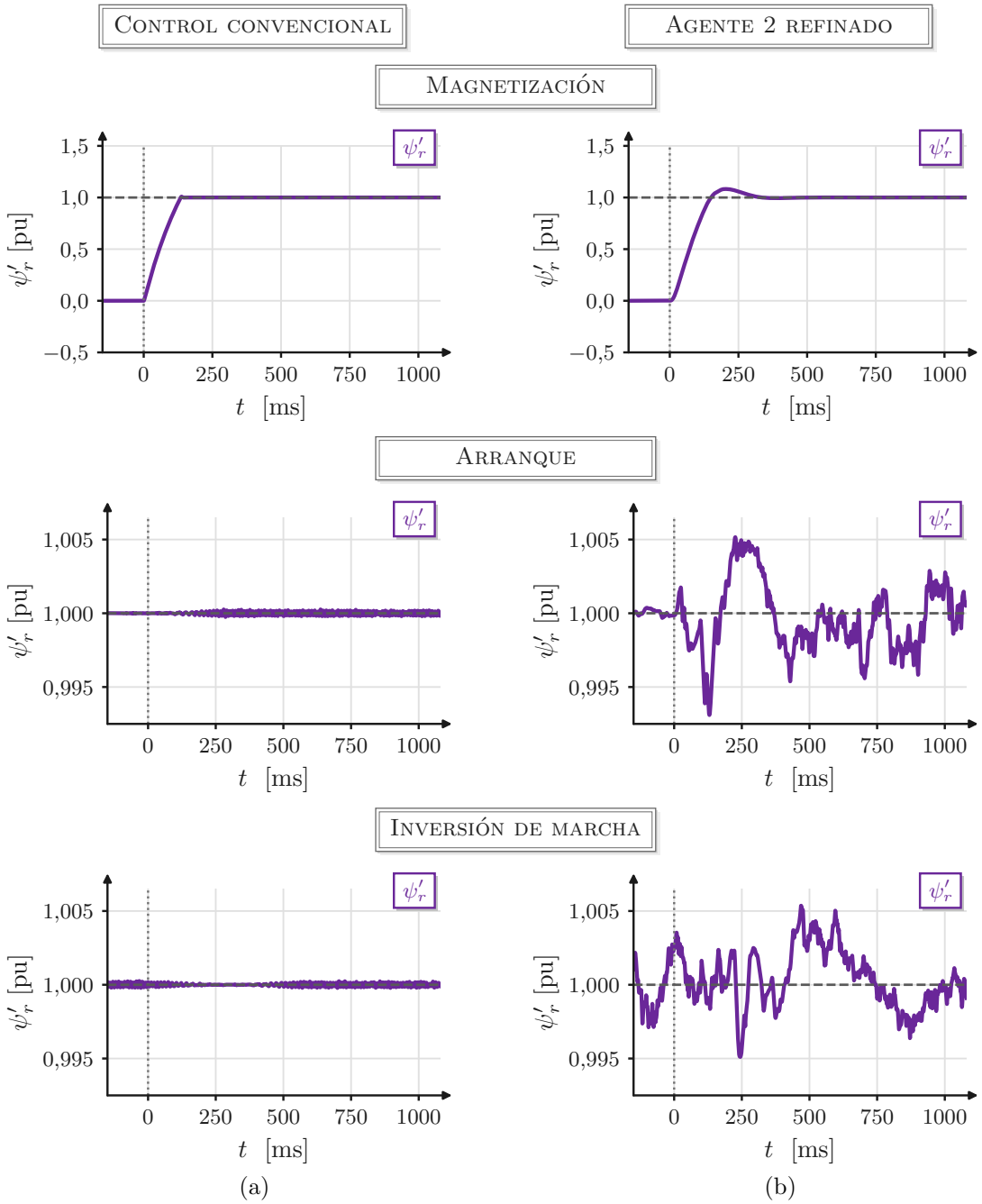
Tabla 7.1: Condiciones de los ensayos experimentales.

Condición	Valor
Potencia nominal de la máquina	5,5 kW
Referencia final de velocidad	±0,30 pu internos, ±450 rpm
Velocidad física estimada	±0,30699 pu, ±460,5 rpm
Pendiente de la referencia	1 pu/s
Carga mecánica aplicada	Sin carga
Frecuencia de control y PWM	3 kHz
Frecuencia de registro de la unidad	1500,15 Hz
Frecuencia del filtro de velocidad	80 Hz
Corrección de escala del *encoder*	1,0233, sólo en el ángulo
Enlace en estado estacionario, control convencional	282–285 V
Enlace en estado estacionario, agente	281–284 V

El Agente 2 refinado controla el lazo interno de corriente. El PI de velocidad, con $k_{p,\omega} = 8{,}4243$ y $T_{i,\omega} = 48$ ms, se ajustó para la constante equivalente de 12 ms del lazo de corriente con el Agente 2 refinado, y el control convencional usa ese mismo ajuste. El PI de flujo, en cambio, tiene un ajuste distinto en cada accionamiento, como muestra la Tabla A.10. Los ajustes y la comprobación de ejecución se detallan en las Secciones A.4.7 y A.5.1.

7.2. Resultados experimentales

Las Figuras 7.2 a 7.9 comparan las tres maniobras con escalas y ventanas comunes. El tiempo se mide desde el inicio del cambio de referencia.



(a) (b)

Figura 7.2: Flujo de rotor estimado. (a) control convencional; (b) Agente 2 refinado. La línea discontinua indica la referencia de 1 pu. Condiciones: las de la Tabla 7.1.

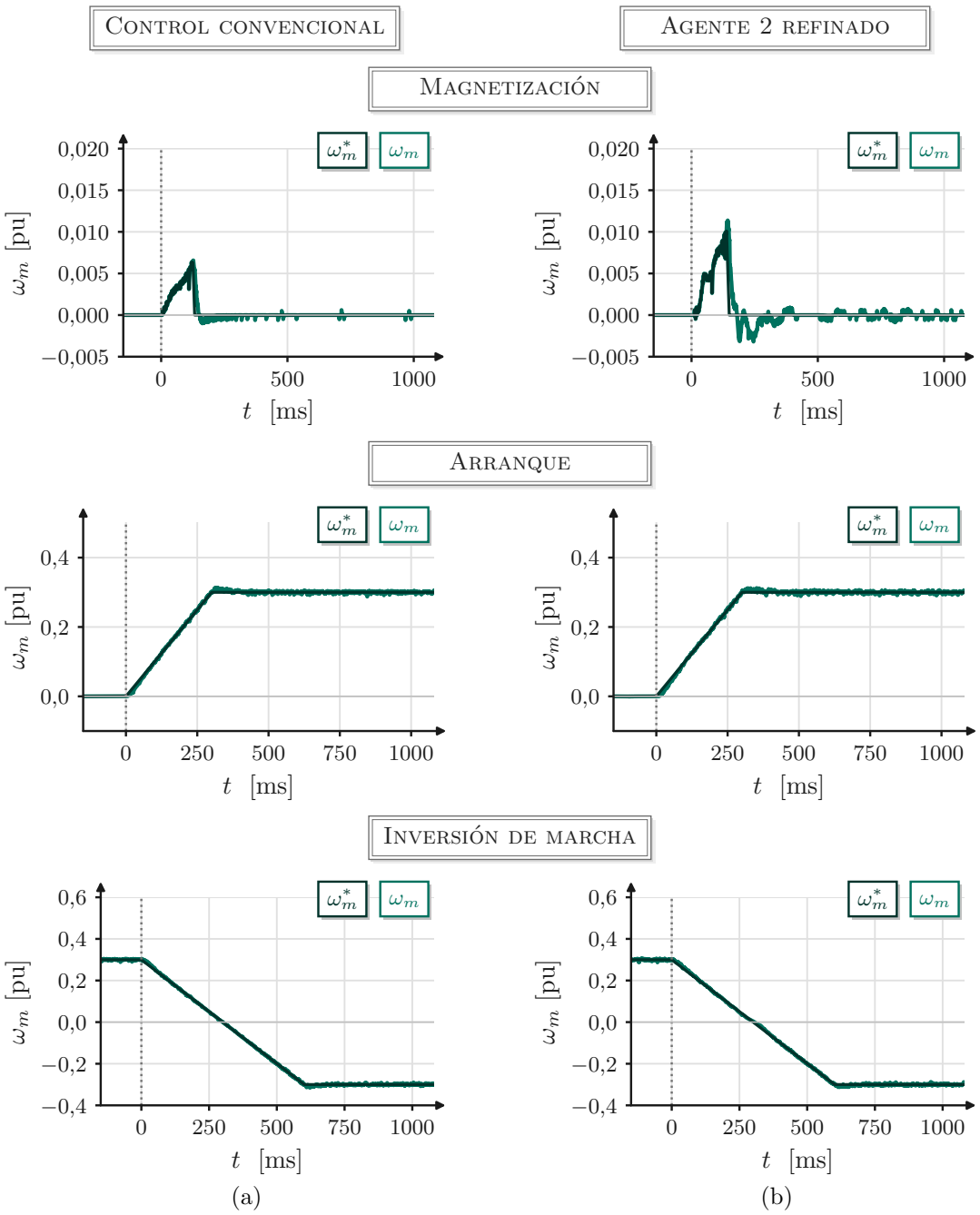


Figura 7.3: Velocidad mecánica medida y su referencia. Las curvas muestran la medida interna sin corregir, en la que 0,30 pu son 450 rpm. Con el factor 1,0233, la velocidad física estimada es 0,30699 pu, unos 460,5 rpm. (a) control convencional; (b) Agente 2 refinado. Cada maniobra utiliza una escala vertical común a ambos controladores. El sobrepaso no se aprecia a esta escala; lo cuantifica la Tabla 7.2. Condiciones: las de la Figura 7.2.

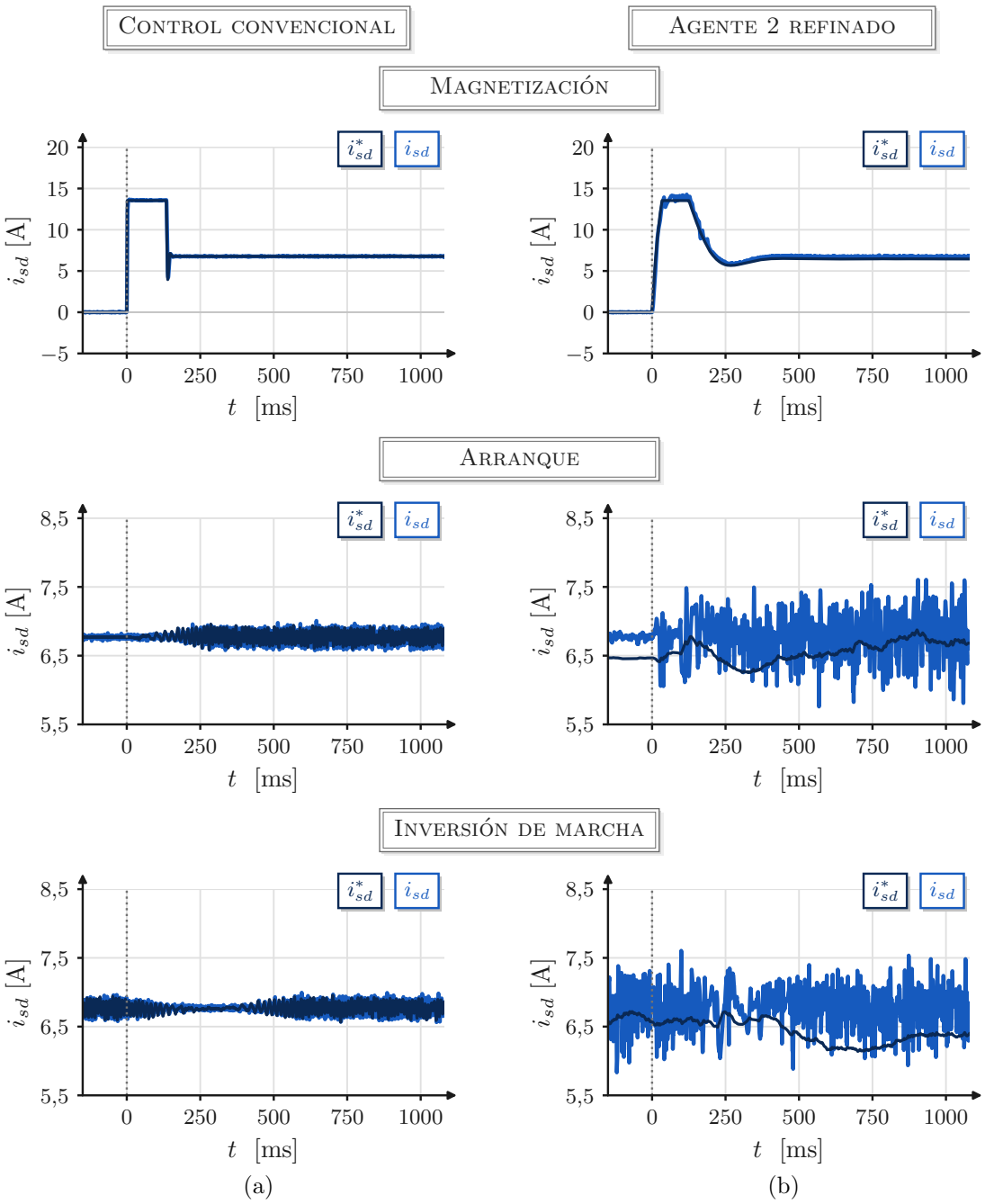


Figura 7.4: Corriente del eje directo con su referencia superpuesta en el tono oscuro del mismo color. (a) control convencional; (b) Agente 2 refinado. El arranque y la inversión de marcha utilizan la misma escala vertical, ampliada alrededor del valor de corriente para mostrar la variación de la corriente registrada; la magnetización utiliza una escala distinta. Condiciones: las de la Figura 7.2.

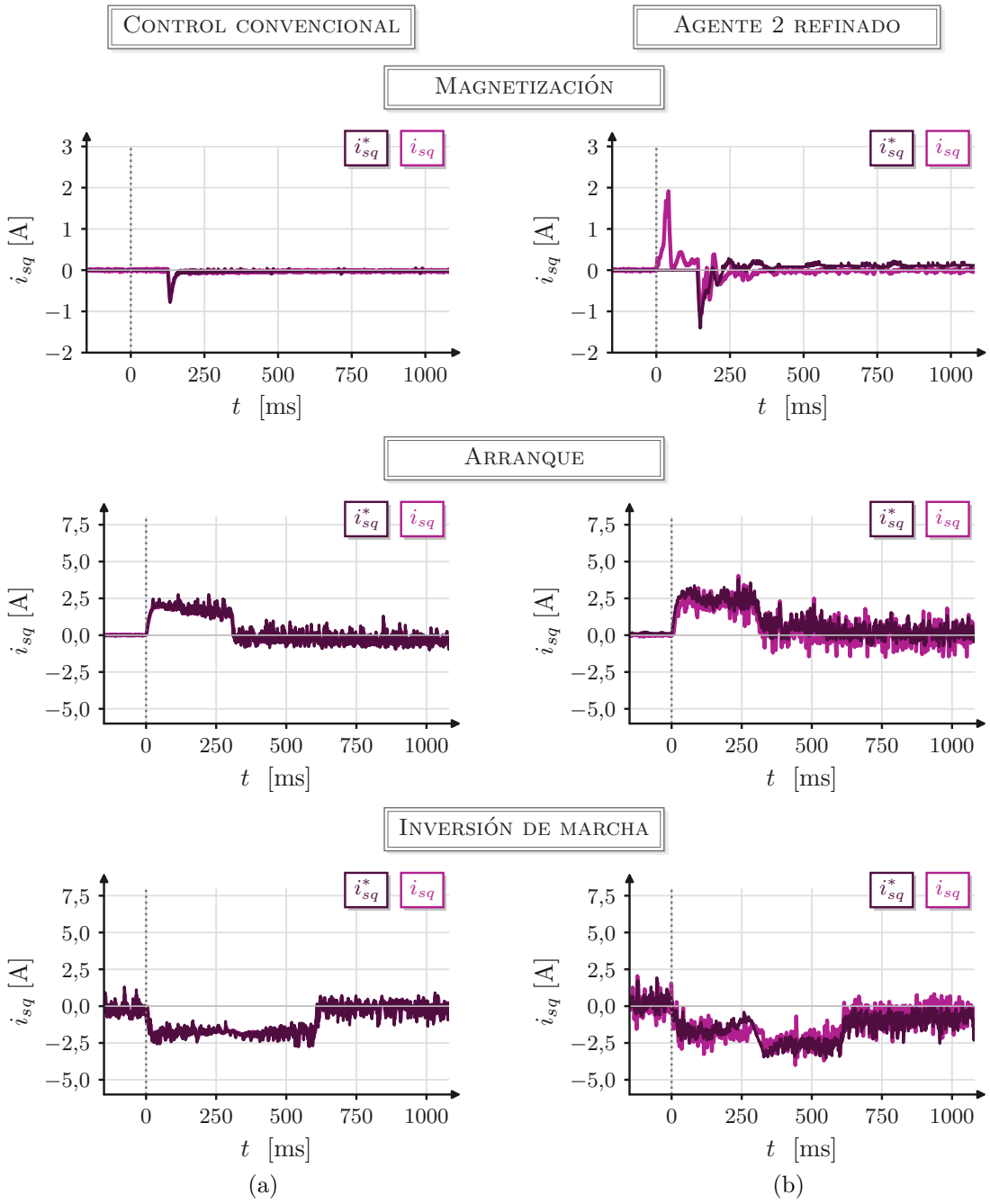


Figura 7.5: Corriente del eje de cuadratura, proporcional al torque estimado una vez establecido el flujo en 1 pu, con su referencia superpuesta. (a) control convencional; (b) Agente 2 refinado. El arranque y la inversión de marcha utilizan la misma escala vertical para ambos controladores. Condiciones: las de la Figura 7.2.

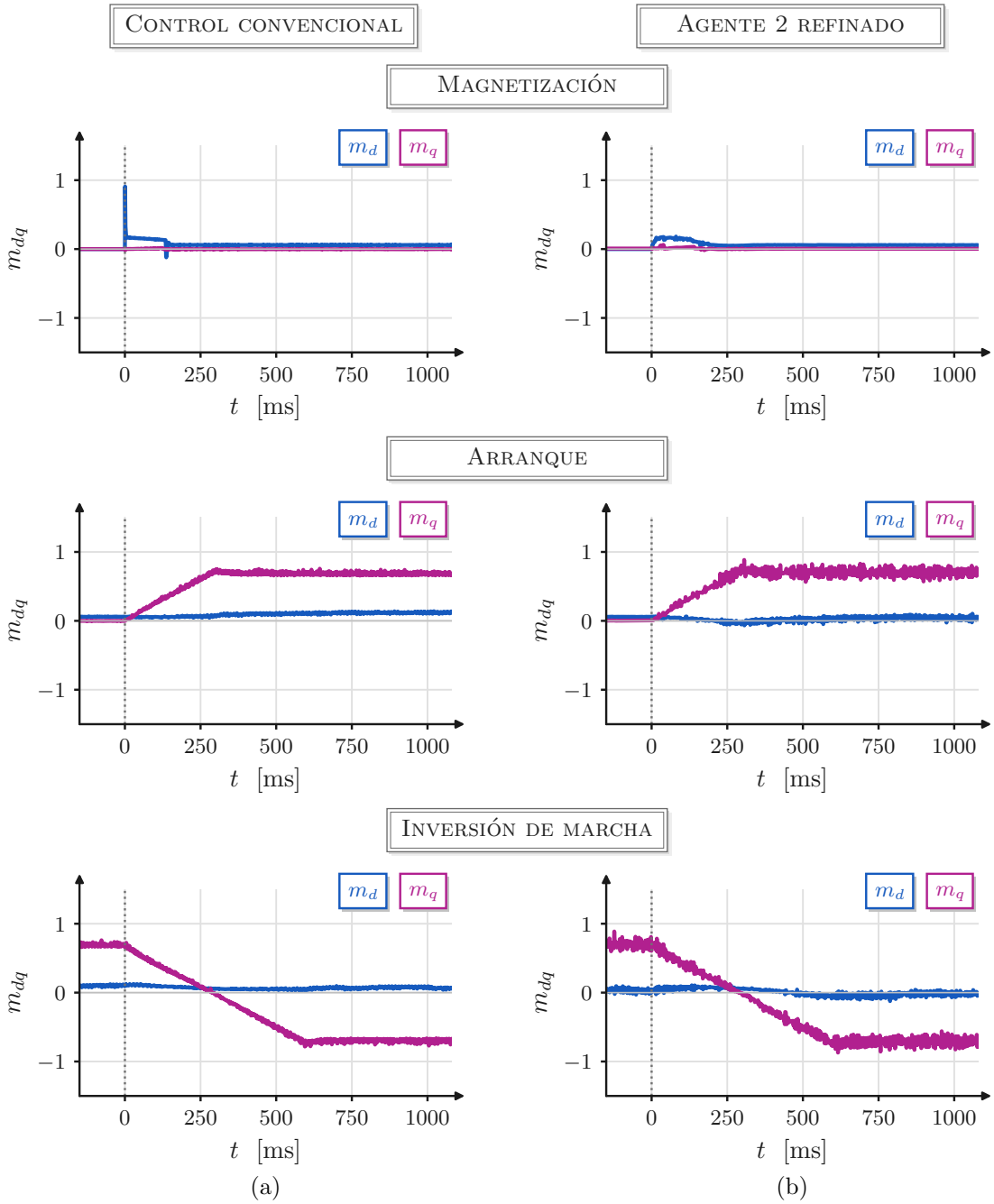


Figura 7.6: Modulantes m_d y m_q registradas antes del modulador. (a) control convencional; (b) Agente 2 refinado. Las seis gráficas comparten escala vertical. Las modulantes registradas corresponden a la orden previa a los límites propios del modulador. En los dos controles, el modulador las aplica con la escala $V_{dc}/2$. En el agente ya incluyen su límite por eje. Ninguna incluye el límite de módulo, Min–Max ni el recorte de los ciclos de trabajo. Condiciones: las de la Figura 7.2.

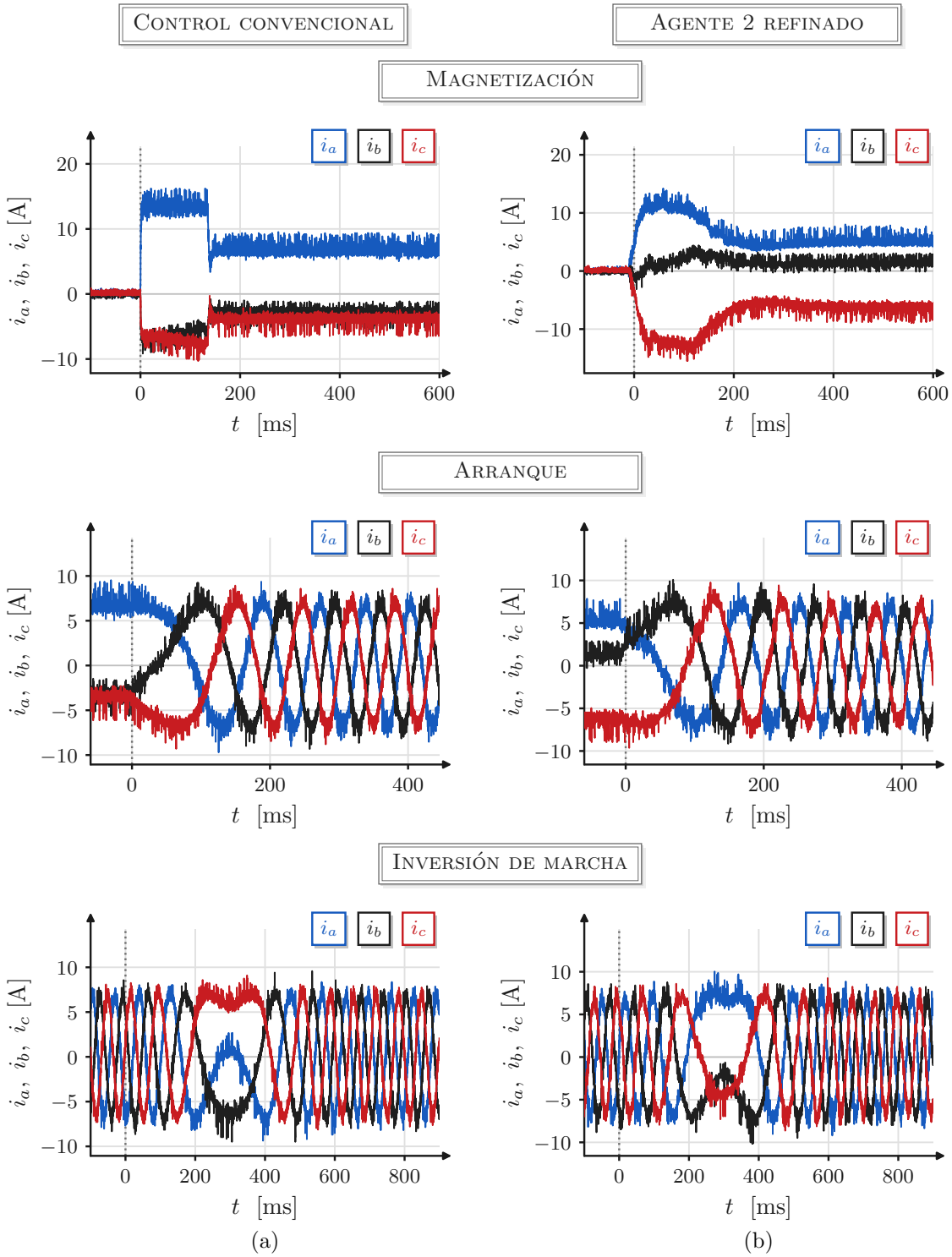


Figura 7.7: Corrientes de fase medidas en bornes. Se muestran los valores mínimos y máximos en pequeños intervalos de tiempo. (a) control convencional; (b) Agente 2 refinado. Las tres maniobras utilizan la misma escala vertical; los dos controladores se muestran con la misma ventana de tiempo en cada maniobra. Condiciones: las de la Tabla 7.1; el tiempo se mide desde el inicio de la maniobra.

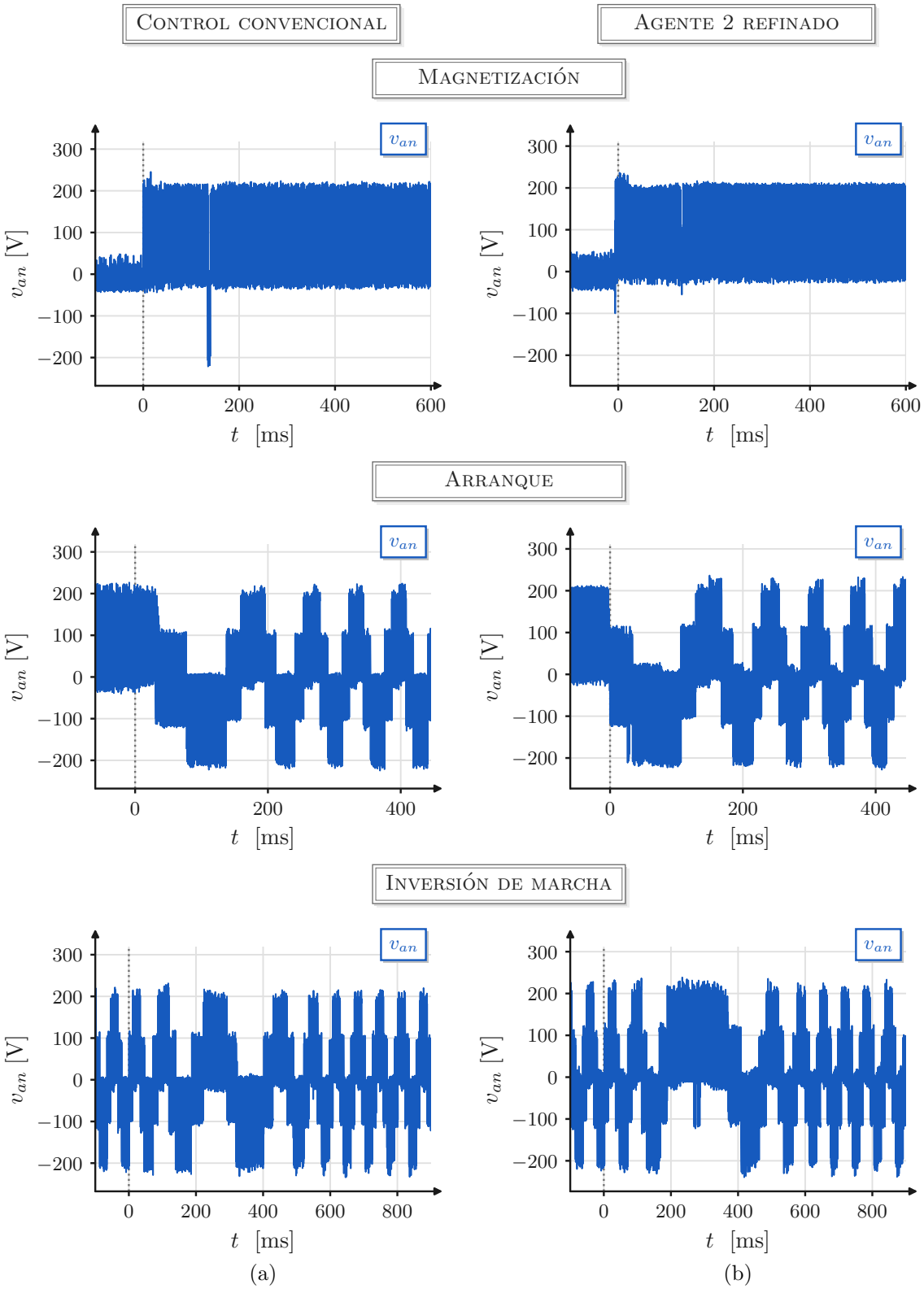


Figura 7.8: Tensión fase–neutro reconstruida a partir de v_{xN}, medida entre cada salida $x \in \{a, b, c\}$ del inversor y N, el negativo del enlace. Se muestran mínimos y máximos en pequeños intervalos de tiempo, con escalas y ventanas iguales para ambos controles. (a) control convencional; (b) Agente 2 refinado. Condiciones: las de la Figura 7.7.

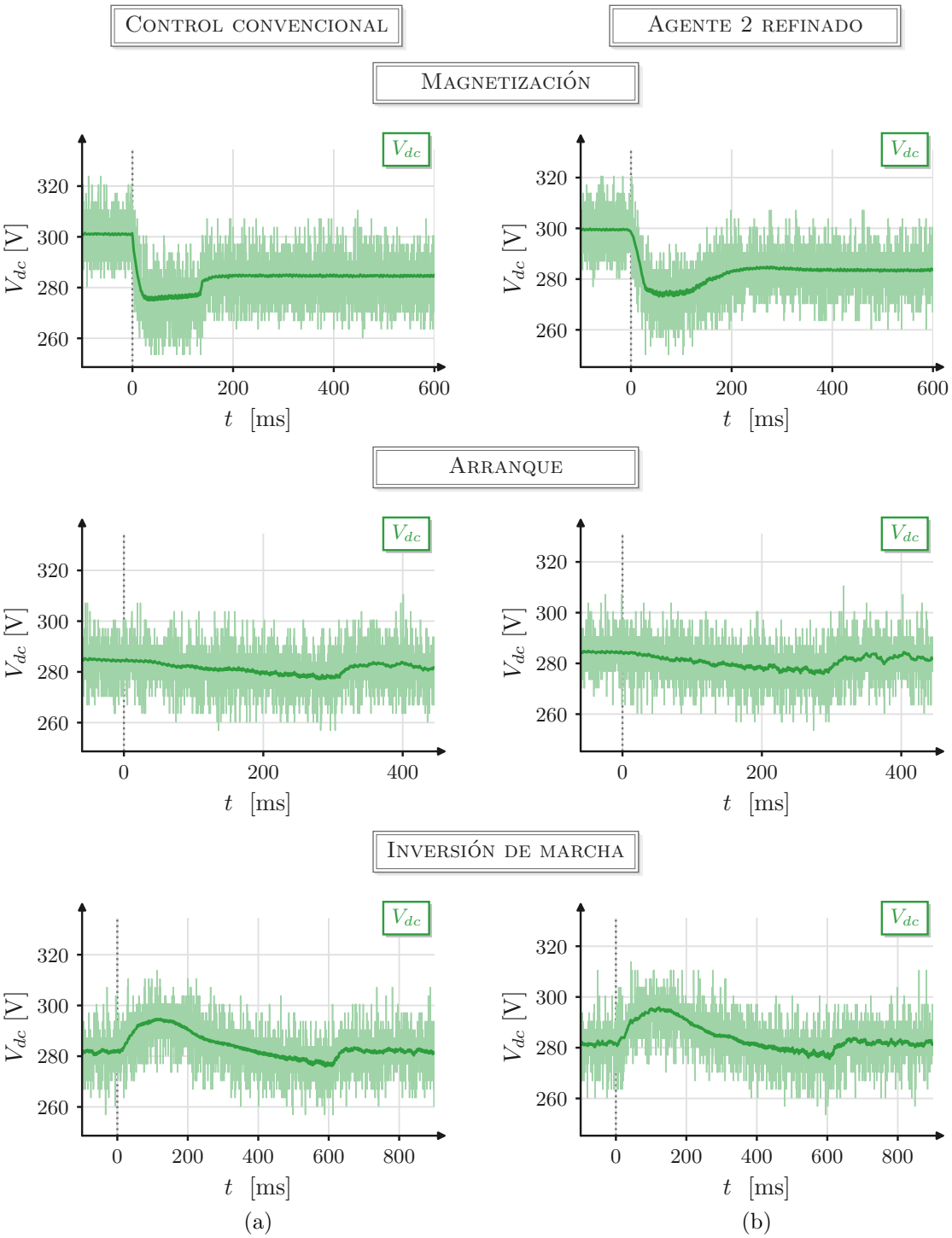


Figura 7.9: Tensión del *DC link*. En tono claro se muestran los valores mínimos y máximos de las muestras, y en tono oscuro, la media móvil de 1 ms. Las tres maniobras utilizan la misma escala vertical. (a) control convencional; (b) Agente 2 refinado. Condiciones: las de la Figura 7.7.

Las referencias de corriente no cambian como escalones en todas las maniobras. En la magnetización, la referencia i^*_{sd} del agente incorpora un alisamiento de 60 ms. Durante el arranque y el frenado, la referencia i^*_{sq} la genera el lazo de velocidad mientras su referencia pasa por un limitador de pendiente; por eso cambia junto con el error de velocidad. La Figura 7.10 compara el seguimiento de cada corriente respecto de su propia referencia, no la respuesta de ambos controles ante un mismo escalón de corriente.

La Figura 7.10 amplía el inicio de los transitorios mostrados en las figuras anteriores. La Tabla 7.2 reúne los indicadores de flujo y velocidad y el seguimiento transitorio de corriente; la Tabla 7.3 resume el RMSE, el sesgo y la desviación estándar del error de corriente en los últimos 200 ms de cada registro, sin suavizar. Las definiciones y ventanas se detallan en la Sección B.1.

Tabla 7.2: Indicadores transitorios de flujo, velocidad y corriente. En i_{sq}, los tiempos y el error $e = i^* - i$ se calculan después de promediar, en cada instante, los valores de los 10 ms anteriores; el RMSE y el sesgo corresponden a los primeros 100 ms. Criterios: Sección B.1.

Indicador	Unidad	Control convencional	Agente 2 refinado
Magnetización: flujo estimado			
Tiempo de subida	ms	105,2	104,5
Sobrepaso	%	0,87	8,24
Tiempo de asentamiento (±5 %)	ms	125,3	260,6
Magnetización: i_{sd}			
Tiempo de subida de la referencia	ms	escalón	26,5
Tiempo de subida de la corriente	ms	2,2	26,7
Arranque: velocidad			
Tiempo de subida	ms	231,3	228,6
Sobrepaso	%	3,52	3,05
Tiempo de asentamiento (±5 %)	ms	286,0	286,6
Arranque: i_{sq}			
Tiempo de subida de la referencia	ms	17,8	19,1
Tiempo de subida de la corriente	ms	17,6	23,3
RMSE transitorio del error	A	0,05	0,56
Sesgo transitorio del error	A	0,03	0,53
Inversión de marcha: velocidad			
Tiempo de subida (10–90 %)	ms	478,6	476,0
Sobrepaso	%	1,77	1,48
Tiempo de asentamiento (±5 %)	ms	571,3	571,3
Inversión de marcha: i_{sq} al iniciar el frenado			
Tiempo de subida de la corriente (10–90 %)	ms	19,7	24,6
RMSE transitorio del error	A	0,05	0,20
Sesgo transitorio del error	A	−0,01	0,01

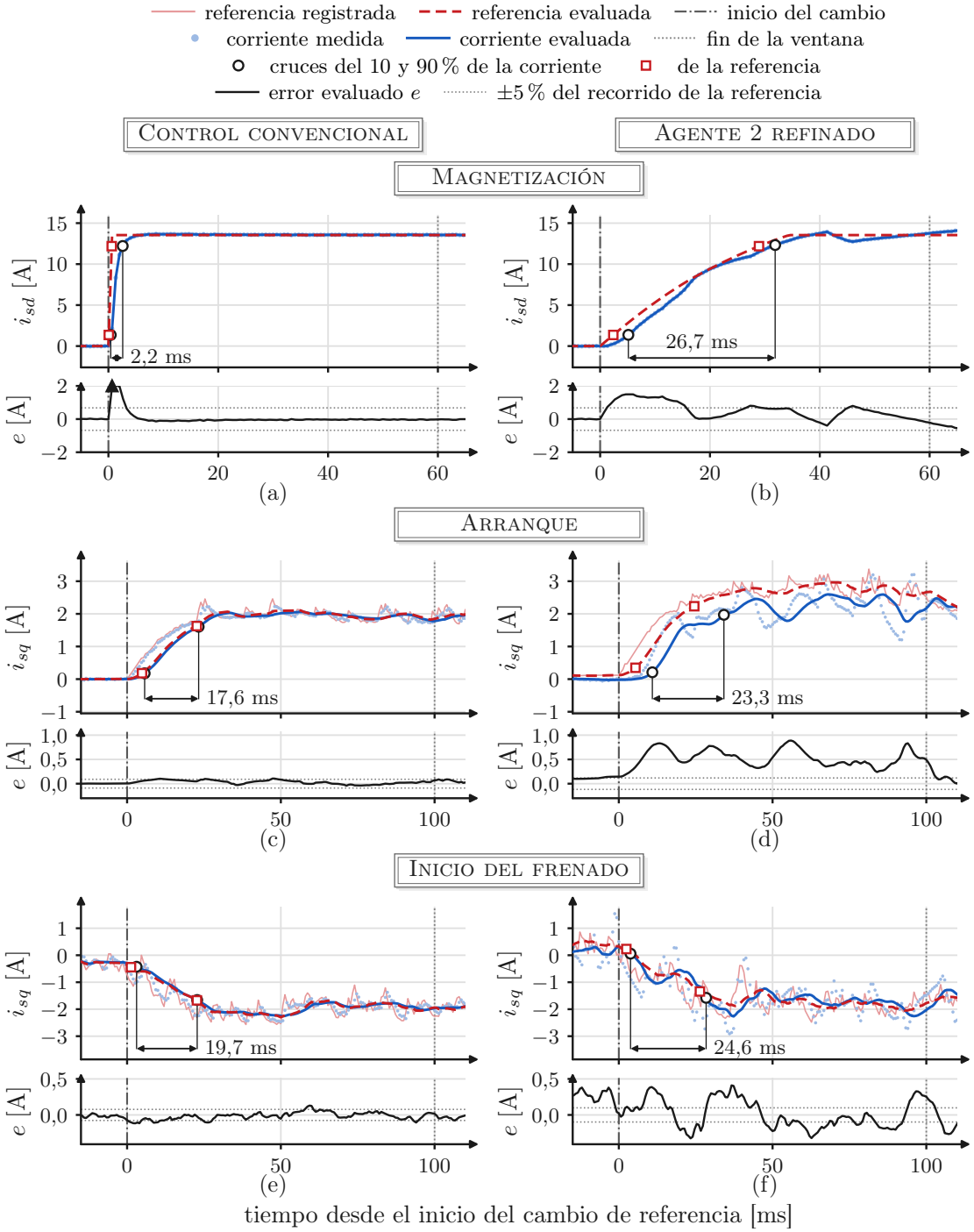


Figura 7.10: Seguimiento de la referencia de corriente durante la magnetización, el arranque y el inicio del frenado. (a), (c) y (e) control convencional; (b), (d) y (f) Agente 2 refinado. Bajo cada gráfica se muestra el error $e = i^* - i$. En i_{sq}, la referencia y la corriente se promedian, en cada instante, con los valores de los 10 ms anteriores. La cota indica el tiempo entre los cruces del 10 y el 90 % de la corriente. Condiciones: las de la Tabla 7.1.

Tabla 7.3: Error de corriente en los últimos 200 ms de cada registro, sin suavizar: RMSE, sesgo y desviación estándar del error $e = i^* - i$.

Maniobra	Control	Eje	RMSE [mA]	Sesgo [mA]	σ_e [mA]
Magnetización	Control convencional	d	18,3	−0,5	18,3
	Agente 2 refinado	d	310,8	−308,3	39,1
	Control convencional	q	16,6	+0,7	16,6
	Agente 2 refinado	q	126,7	+122,0	34,4
Arranque	Control convencional	d	106,3	+0,7	106,3
	Agente 2 refinado	d	362,9	−92,8	350,8
	Control convencional	q	244,0	+11,3	243,8
	Agente 2 refinado	q	536,2	+109,5	524,9
Inversión de marcha	Control convencional	d	106,0	−3,7	105,9
	Agente 2 refinado	d	502,5	−422,4	272,2
	Control convencional	q	297,5	−8,4	297,3
	Agente 2 refinado	q	627,3	−275,5	563,6

7.2.1. Magnetización

La magnetización establece el flujo con el eje detenido. Ambos accionamientos alcanzan la referencia de flujo; con el agente, el flujo presenta mayor sobrepaso y tiempo de asentamiento. El agente no recibe el flujo estimado ni su error en la observación y fue entrenado con una carga $R-L$, sin la dinámica del rotor. Por ello, el flujo continúa regulado por un PI externo. Su ajuste partió de la constante equivalente de 12 ms identificada en PLECS para la dinámica del lazo interno; esta constante no corresponde al tiempo de cálculo del agente. La sintonía resultante se probó en PLECS y se redujeron a la mitad las ganancias de ese diseño, de $k_{p,\mathrm{imr}} = 8{,}0580$ a $4{,}0290$, manteniendo $T_i = 48$ ms, porque las variaciones lentas de i_{sd} hacían variar el flujo y el torque. El lazo de velocidad respondía modificando la referencia i^*_{sq}, que terminaba oscilando. Como esa reducción aumentó el sobrepaso durante la magnetización, se aplicó un alisamiento de 60 ms a la referencia i^*_{mr} del lazo de flujo, que genera la referencia de corriente i^*_{sd}. Este valor se eligió mediante un barrido en simulación dentro del intervalo que Fröhr propone para reducir el sobrepaso de la consigna, a costa de hacer más lenta la subida [27, pp. 291–294]. Los ajustes se detallan en la Sección A.4.7.

La referencia i^*_{sd} del control convencional cambia prácticamente como un escalón y su corriente sube en 2,2 ms. Con el agente, la referencia sube en 26,5 ms por el alisamiento y la corriente tarda 26,7 ms. Por ello, esta diferencia describe principalmente la forma de las referencias y no permite atribuir por sí sola una menor rapidez al control interno. La corriente del agente vuelve a entrar en la banda de ±5 % apenas 6,7 ms antes de que la referencia comience a disminuir. Ese intervalo no basta para comprobar su tiempo de asentamiento. La evidencia se presenta en la Figura B.1.

7.2.2. Arranque

Durante el arranque, ambos accionamientos siguen la referencia de velocidad generada por el limitador de pendiente. Sus tiempos de velocidad describen el seguimiento de esa rampa y no la rapidez aislada del control de corriente.

En los primeros 100 ms, la corriente de cuadratura del control convencional reproduce el tiempo de subida de su referencia: 17,6 frente a 17,8 ms. Con el agente, la referencia sube en 19,1 ms y la corriente en 23,3 ms. Sobre las señales evaluadas, el RMSE transitorio es 0,05 A con el control convencional y 0,56 A con el agente; los sesgos son 0,03 y 0,53 A, respectivamente. Estas cifras comparan el seguimiento observado dentro del accionamiento completo.

7.2.3. Inversión de marcha

La inversión comprende el frenado, el paso por velocidad cero y el giro en sentido contrario. Ambos accionamientos completan la maniobra y mantienen la velocidad próxima a su referencia final.

Para calcular el tiempo de subida, el recorrido de velocidad entre +0,30 y −0,30 pu se normaliza de 0 a 1. El intervalo entre el 10 y el 90 % tarda 478,6 ms con el control convencional y 476,0 ms con el agente, y corresponde aproximadamente al tramo entre +0,24 y −0,24 pu. No representa la duración completa de la inversión ni el tiempo hasta el paso por velocidad cero.

Al iniciar el frenado, el tiempo de subida normalizado de i_{sq} es de 19,7 ms con el control convencional y 24,6 ms con el agente. En los primeros 100 ms, el RMSE transitorio es 0,05 y 0,20 A, mientras que el sesgo permanece próximo a cero en ambos casos. Al pasar por velocidad cero, la referencia del agente vuelve a cambiar y la corriente adopta otro valor medio; ese tramo no contiene un cambio común y aislado que permita definir un segundo tiempo de subida para ambos controles.

En el estado estacionario con velocidad negativa, el sesgo de cuadratura del agente cambia de signo respecto del arranque. La Tabla 7.3 conserva ambas condiciones por separado para evitar que sus errores medios se cancelen. La tensión del enlace aumenta durante el frenado sin activar el circuito de frenado (Figura 7.9); su comprobación se detalla en la Sección B.6.

7.3. Alcance de la validación experimental

El Agente 2 refinado controla el lazo interno de corriente de la máquina de inducción y completa la magnetización, el arranque y la inversión de marcha sin aprendizaje *online*. Los lazos de flujo y velocidad permanecen activos y ajustan las referencias de corriente. El resultado demuestra la transferencia experimental al accionamiento en las condiciones ensayadas, aunque el agente mantiene mayor RMSE y sesgo de corriente que el control convencional.

La comparación corresponde a los accionamientos implementados, incluidos sus ajustes del lazo de flujo. Los ensayos se realizaron sin carga mecánica y no incluyeron escalones de carga. La reflexión permaneció activa con el agente; sin repetir los ensayos con ella desactivada, no se aisló su efecto experimental.

7.4. Contraste con la simulación

La Tabla 7.4 selecciona los indicadores necesarios para contrastar las campañas con las simulaciones de la Sección 6.2. Se aplican los mismos criterios de flujo y velocidad, y el error de corriente se evalúa en los últimos 200 ms de cada registro. El procedimiento se detalla en la Sección B.1.

Tabla 7.4: Contraste de indicadores seleccionados entre simulación y banco. El RMSE de corriente corresponde a los últimos 200 ms de cada registro.

		Simulación		Banco	
Maniobra	Indicador	Convencional	Agente	Convencional	Agente
Magnetización	Sobrepaso del flujo [%]	0,93	8,18	0,87	8,24
Magnetización	Asentamiento del flujo [ms]	125,7	257,3	125,3	260,6
Arranque	Sobrepaso de velocidad [%]	5,46	5,68	3,52	3,05
Inversión de marcha	Sobrepaso de velocidad [%]	2,73	2,83	1,77	1,48
Magnetización	RMSE de e_{sd} [mA]	0,0	372,0	18,3	310,8
	RMSE de e_{sq} [mA]	0,0	84,4	16,6	126,7
Arranque	RMSE de e_{sd} [mA]	0,2	446,2	106,3	362,9
	RMSE de e_{sq} [mA]	0,0	228,9	244,0	536,2
Inversión de marcha	RMSE de e_{sd} [mA]	0,4	379,8	106,0	502,5
	RMSE de e_{sq} [mA]	0,2	651,1	297,5	627,3

La magnetización presenta resultados próximos entre simulación y banco. En velocidad, el sobrepaso simulado es mayor durante el arranque y la inversión. El RMSE de corriente del control convencional aumenta en el banco; con el agente, las diferencias dependen del eje y de la maniobra, sin una mejora o empeoramiento uniforme.

Las condiciones no son idénticas. En la simulación, el inversor se alimenta desde un *DC link* ideal de 300 V; por tanto, no se representan el variac, el rectificador ni los dos condensadores del *DC link* del banco. En el ensayo, la tensión del *DC link* varía durante las maniobras. La inercia utilizada en la simulación se heredó del modelo validado de Gálvez [22] y no se volvió a identificar experimentalmente en esta memoria; tampoco se representan el roce ni la cuantización del *encoder*. Estas diferencias, junto con las distintas sesiones de ensayo, limitan la atribución de los resultados a una causa única.

Capítulo 8

Mejora del transitorio del agente

Este capítulo describe cómo se mejoró el transitorio del Agente 2 refinado sin degradar su estado estacionario y sin cambiar los pesos de la red. Compara el accionamiento con el agente contra el control convencional, el PI de corriente implementado que define la Sección 5.4, en el mismo banco del Capítulo 7, y complementa así el sexto objetivo específico. El control convencional entrega una referencia experimental. El objetivo no es superarlo, sino comprobar si una política aprendida con una carga $R-L$ y una perturbación de tensión, con sus pesos fijos, puede controlar el lazo interno de la máquina real. La red no recibe el flujo de rotor ni los parámetros de la máquina. Su entrada de velocidad, ω_{pu}, fue en el entrenamiento la frecuencia del marco de la carga $R-L$. En el banco es la velocidad medida por el *encoder* más el deslizamiento estimado, según la Tabla A.3. Todos los resultados del capítulo se midieron en el banco. Las simulaciones sólo se usaron para elegir valores de ajuste y, junto con las comprobaciones complementarias, se reúnen en el Apéndice D. El punto de partida es el Agente 2 refinado con la configuración del banco del Capítulo 7.

8.1. El problema inicial

Con la configuración del banco del Capítulo 7, el Agente 2 refinado siguió sus referencias en estado estacionario, pero respondió tarde al arrancar. La corriente de cuadratura reaccionó 6,0 ms después de su referencia, frente a 0,7 ms con el control convencional. En los primeros 50 ms, la corriente media faltante fue 0,59 y 0,07 A, respectivamente.

El atraso no viene del cálculo de la red. En la lectura tomada al terminar la campaña del Capítulo 7, el ciclo del agente dentro de la interrupción duró 239,2 μs, dentro del período de control $T_s = 333{,}3$ μs, y el contador de ciclos tardíos valía cero, según la Sección A.5. En la rutina original, el agente corre antes que el IFOC, de modo que usa la referencia i_{sq}^{*} y el ángulo del marco calculados en el período anterior. La Figura D.3, en el Apéndice D, separa el período de control, el muestreo y la conmutación.

La acción también era poco refinada. Cada decisión suma a la orden de tensión uno de 15 niveles por eje, el cero y siete por signo, desde 0,025 hasta 8,33 V de fase, en la base $V_N/2 = 325$ V de la acción que describe la Sección A.6. En el arranque, la política sumó sólo los tres niveles menores hasta que el error de cuadratura llegó a unos 0,9 A, y entonces saltó a un paso de 1,21 V o más. En la magnetización, la corriente del eje directo

tardó 119,3 ms en asentarse en la banda de ±5 %, aunque su referencia alisada subió en 26,5 ms.

Al subir la velocidad apareció un tercer problema, esta vez en estado estacionario: un sesgo de la corriente de cuadratura que crecía con la velocidad, de +211 mA a 459 rpm hasta +739 mA a 1029 rpm.

8.2. La configuración final del agente

La configuración final reúne tres correcciones que actúan fuera de la red: el ángulo al día, la compensación del retardo y la ganancia adaptativa. Los pesos, la grilla de acciones y la forma de la observación no cambian. La Tabla 8.1 resume qué problema corrige cada una, con qué valor se usó y qué medición la respalda. Las tres se usaron juntas en los dos ensayos de la configuración final, a ±0,30 pu y a 1050 rpm, con valores distintos.

Tabla 8.1: Las tres correcciones de la configuración final. Los pesos de la red no cambian. Las evidencias se desarrollan en la Sección 8.3.

Corrección	Problema corregido	Valor empleado	Evidencia
Ángulo al día	El agente medía y armaba su orden con el ángulo de la muestra anterior; sesgo de i_{sq} creciente con la velocidad	Medición: $+1{,}00\,\omega_s T_s$; orden: $+1{,}40\,\omega_s T_s$, igual en los dos ensayos	Sesgo de i_{sq} de +739 a −108 mA cerca de 1050 rpm
Compensación del retardo	La referencia llegaba un período tarde y la política esperaba un error de unos 0,9 A antes del primer paso grande	$t_{d,\mathrm{ag}} = 4$ ms a ±0,30 pu y 2 ms a 1050 rpm, en los primeros 50 ms del arranque desde el reposo	Corriente media faltante de 0,59 a 0,23 A con la compensación sola, de 7 ms
Ganancia adaptativa	Paso de tensión grueso cerca de la referencia; la red no distinguía errores pequeños	±0,30 pu: $g = 2$, $k = 0{,}5$ en los dos ejes, $P = 1$, $k_{\mathrm{adapt}} = 2$. 1050 rpm: $P = (1{,}5;\ 2)$, $k_{\mathrm{adapt}} = (4;\ 2)$ en $(d;\ q)$, error observado recortado a ±0,30 pu	Asentamiento de i_{sd} en la magnetización de 119,3 a 33,3 ms; mayor salto de m_q de 0,119 a 0,059

8.3. Formalización de las tres correcciones

8.3.1. Ajuste experimental del ángulo

El ángulo usado por el agente tenía un desfase que cambiaba con la velocidad. Se varió la corrección angular y se conservó la que redujo el error de cuadratura. El signo se eligió por el camino más corto, porque 0° y 360° representan la misma orientación. Este ajuste experimental incluye los retardos de medida y de cálculo del banco, de modo que no permite atribuir el desfase sólo al *encoder*.

La Figura 8.1 muestra el resultado del ajuste. Cada punto es el sesgo $e_{sq} = i^*_{sq} - i_{sq}$ antes o después de corregir el ángulo. En las cuatro velocidades comparadas, el sesgo disminuyó y quedó cerca de cero.

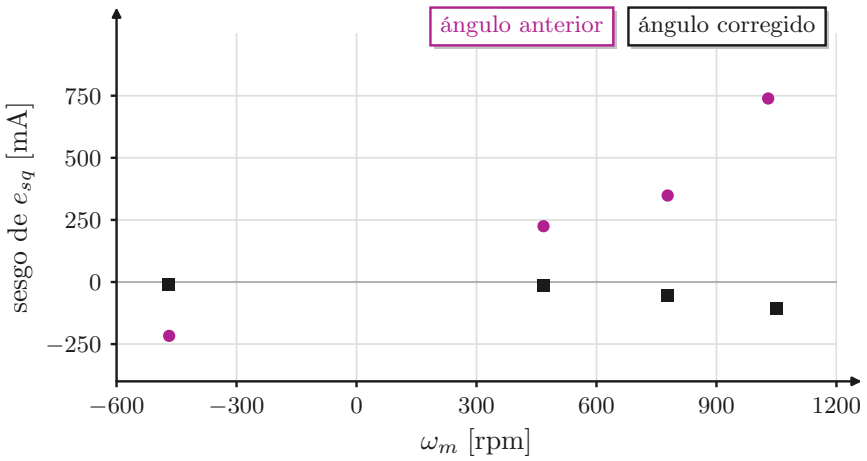


Figura 8.1: Sesgo de la corriente de cuadratura en estado estacionario antes y después de aplicar la corrección angular. Cada punto es la media de $e_{sq} = i^*_{sq} - i_{sq}$ en los últimos 200 ms. La velocidad se obtuvo desde la frecuencia de las corrientes. Condiciones: Agente 2 refinado, ganancias del error de 12 y 4, máquina sin carga, *DC link* entre 459 y 466 V y período de control de 333,3 μs.

La implementación aplica dos giros obtenidos en ese barrido:

$$\theta_{\text{med}}[k] = \theta_s[k-1] + 1{,}00\,\omega_s[k-1]\,T_s, \tag{8.1}$$

$$\theta_{\text{ord}}[k] = \theta_s[k-1] + 1{,}40\,\omega_s[k-1]\,T_s, \tag{8.2}$$

donde θ_{med} transforma las corrientes y θ_{ord} forma la orden de tensión. Los factores 1,00 y 1,40 son calibraciones de la secuencia de cálculo. Con la máquina detenida, $\omega_s = 0$ y no se corrige el ángulo.

Cerca de 1050 rpm, el sesgo de cuadratura cambió de +739 a −108 mA y su RMSE bajó de 953 a 495 mA. Esta comparación verifica el efecto conjunto de la corrección angular. No permite atribuir todo el error restante a una sola causa.

8.3.2. La compensación del retardo

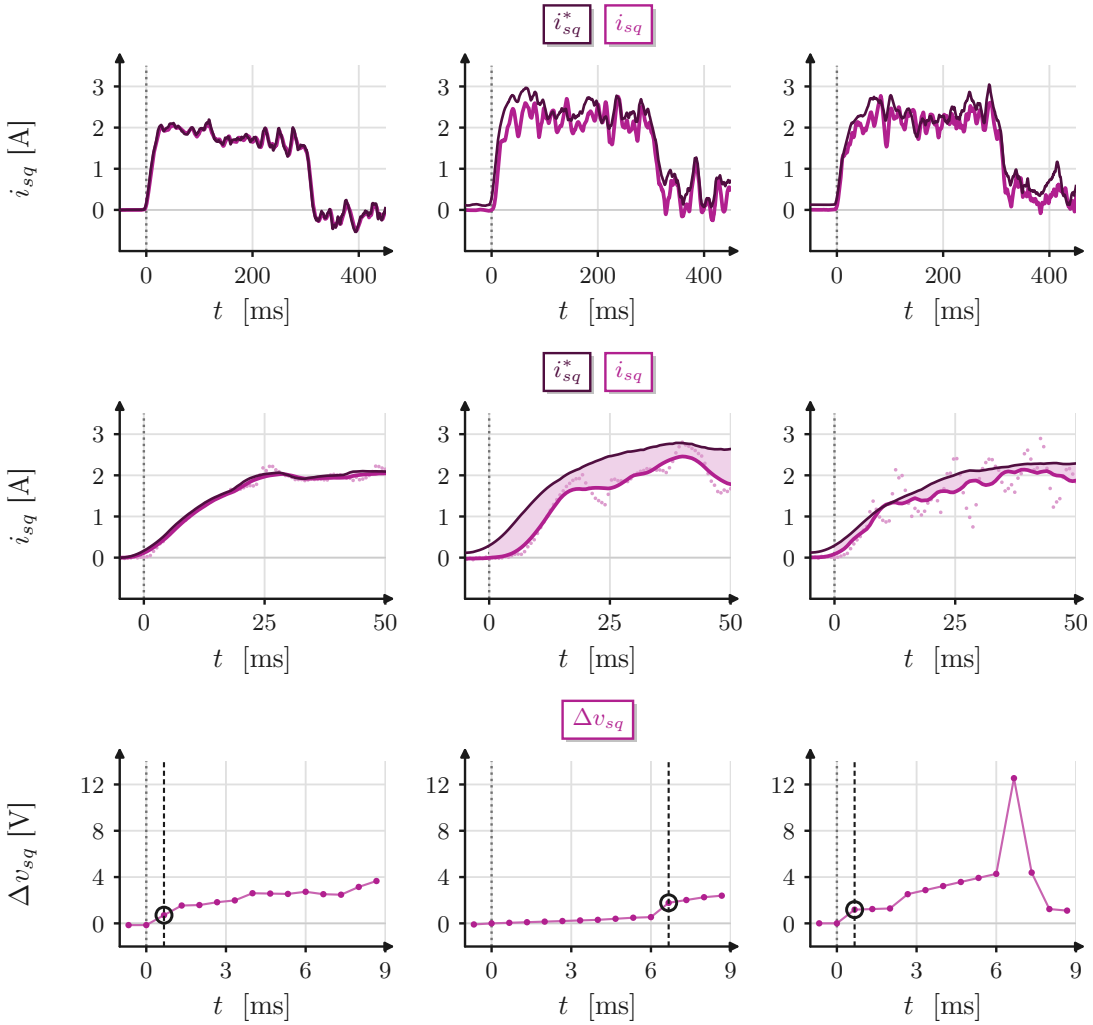
El agente recibe la referencia de cuadratura un período después que el control convencional, porque corre antes que el lazo de velocidad. Además, la política no da un paso grande mientras el error es pequeño. Al arrancar, la referencia de velocidad sube en rampa de 1 pu/s y la de cuadratura sube con ella. Durante los primeros 50 ms del arranque desde el reposo, la red recibe entonces la referencia prolongada

$$i^*_{sq,\text{ag}} = i^*_{sq} + t_{d,\text{ag}}\,\frac{\mathrm{d}i^*_{sq}}{\mathrm{d}t}, \tag{8.3}$$

donde $t_{d,\text{ag}}$ es el retardo que se compensa. La pendiente sale de la diferencia de i^*_{sq} entre dos muestras, con un filtro de primer orden de 80 Hz. La referencia que usan el lazo de velocidad y el IFOC no cambia.

Sin compensación, la política tarda en elegir un paso grande. Con la pendiente medida, la espera calculada es 6,1 ms y la medida, 6,67 ms. El Apéndice D muestra esta cuenta y distingue ese paso grande de las correcciones finas anteriores.

La Figura 8.2 aísla esta corrección con un valor de 7 ms. La corriente media faltante durante los primeros 50 ms bajó de 0,59 a 0,23 A. En la configuración final, el error duplicado redujo la espera de la política y bastó compensar 4 ms.



(a) control convencional

(b) configuración del Capítulo 7

(c) con la compensación del retardo

Figura 8.2: Comienzo del arranque medido por la unidad de control: (a) control convencional; (b) Agente 2 refinado con la configuración del banco del Capítulo 7; (c) el mismo agente con la compensación de un retardo de 7 ms y sin las otras dos correcciones. Arriba, el arranque completo; al medio, los primeros 50 ms, con la corriente que falta sombreada; abajo, el cambio de la orden de tensión de cuadratura, Δv_{sq}, con el primer paso grande marcado con un círculo. En i_{sq}, la referencia y la corriente llevan una media móvil de 10 ms sin desfase; al medio se agregan las muestras de la corriente. El área sombreada se divide por los 50 ms para expresar la corriente media faltante: (a) 0,07, (b) 0,59 y (c) 0,23 A. El primer paso grande llegó a los (a) 0,67, (b) 6,67 y (c) 0,67 ms. El registro guarda una muestra cada 0,67 ms. Condiciones: arranque a +0,30 pu sin carga y *DC link* entre 275 y 286 V; el origen de tiempos es el cambio de la referencia de velocidad.

La Figura 8.2 no incluye la ganancia adaptativa ni la corrección angular. La Figura 8.3 reúne las tres correcciones.

La compensación tiene tres límites. Actúa sólo en los primeros 50 ms de un arranque desde el reposo, no al comenzar el frenado. Su valor se eligió en simulación. Y no obliga al agente a reproducir la forma de la orden del control convencional, que aumenta su tensión en forma continua, mientras el agente lo hace con los pasos de su grilla.

8.3.3. La ganancia adaptativa

La DDQN se trató como una caja negra. Sus entradas de error, la 1 a la 4 de la Tabla A.3, se multiplicaron por g_x, y el paso de su acción, por k_x. Las medias de $|e_x|$, entradas 12 y 13, no se multiplicaron. El escalamiento total es

$$P_x = g_x k_x, \qquad k_{\text{adapt},x} = 1/k_x. \tag{8.4}$$

A ±0,30 pu se usó $g = 2$ y $k = 0{,}5$ en ambos ejes. La red observa el doble del error y aplica la mitad del paso, de modo que $P = 1$. La resolución se duplica sin cambiar este producto. Los pasos mínimo y máximo bajan de 0,025 y 8,33 V a 0,013 y 4,17 V.

El banco confirmó la reducción del paso: el mayor salto de la modulante de cuadratura entre dos muestras pasó de 0,119 a 0,059, una razón de 0,495 frente a 0,5 calculado. La magnetización aísla esta corrección, porque con la máquina detenida el ángulo al día vale cero y la compensación del retardo actúa sólo en el eje de cuadratura. En ella, el asentamiento de i_{sd} bajó de 119,3 a 33,3 ms y el RMSE del eje directo en estado estacionario, de 311 a 148 mA.

La corrección amplifica el ruido observado y reduce la rapidez máxima de la acción. Por eso $k_{\text{adapt}} = 2$ es un ajuste experimental de este agente, no un óptimo general. El Apéndice D presenta su barrido en simulación.

8.4. Resultado conjunto a ±0,30 pu

La configuración final se ensayó a ±0,30 pu con las tres correcciones a la vez, en las mismas maniobras y con el mismo lazo de velocidad del Capítulo 7. El lazo de flujo del agente conserva el alisamiento de 60 ms de ese capítulo. Como en ese capítulo, ±0,30 pu es el valor interno de la referencia de velocidad. El factor 1,0233 del *encoder* sólo corrige el ángulo, y con él la velocidad física estimada es de unos ±460,5 rpm, según la Tabla B.4. La Tabla 8.2 compara el lazo interno con el control convencional y con el Agente 2 refinado, en la configuración del Capítulo 7 y en la final, y la Figura 8.3 muestra el seguimiento de la configuración final junto al del control convencional.

La configuración final comenzó a corregir con pasos finos desde la primera o la segunda muestra registrada. La corriente reaccionó entre 0 y 1,3 ms después de su referencia en tres repeticiones. La corriente faltante durante los primeros 50 ms fue de 5,8 a 10,6 A ms, frente a 29,4 A ms con la configuración del Capítulo 7.

En la magnetización, la corriente de la configuración final se asentó a los 33,3 ms, frente a 119,3 ms con la configuración del Capítulo 7. La referencia alisada se asentó a los 32,0 ms.

Tabla 8.2: Agente 2 refinado con la configuración final y con la configuración del banco del Capítulo 7, frente al control convencional a ±0,30 pu, medidos por la unidad de control. El asentamiento y el RMSE de estado estacionario se calculan sin suavizar; el RMSE usa los últimos 200 ms de cada registro. Condiciones: sin carga, *DC link* entre 281 y 286 V.

Indicador	Unidad	Control convencional	Configuración del Capítulo 7	Configuración final
Indicadores transitorios				
Asentamiento de i_{sd} en la magnetización (±5 %)	ms	3,3	119,3	33,3
Estado estacionario, RMSE de $e = i^* - i$				
Eje directo, después de la magnetización	mA	18	311	148
Eje de cuadratura, después de la magnetización	mA	17	127	93
Eje directo, después del arranque	mA	106	363	351
Eje de cuadratura, después del arranque	mA	244	536	493
Eje directo, después de la inversión	mA	106	503	330
Eje de cuadratura, después de la inversión	mA	297	627	475

El estado estacionario no se degradó. En las seis ventanas de la Tabla 8.2, el RMSE de la configuración final fue menor que el de la configuración del Capítulo 7. El control convencional conservó el menor RMSE en estado estacionario, entre 17 y 297 mA.

Las Figuras 8.4 a 8.8 muestran la configuración final a ±0,30 pu en las tres maniobras: el flujo y la velocidad, las corrientes del marco, y las modulantes, medidos por la unidad de control, y el *DC link*, las corrientes de fase y la tensión fase–neutro, medidos con el osciloscopio. En la magnetización, el flujo alcanzó 1,081 pu. Durante el arranque y la inversión de marcha, se mantuvo entre 0,996 y 1,004 pu. La velocidad llegó a 0,314 pu en el arranque y a −0,313 pu en la inversión. La media de 1 ms del *DC link* se mantuvo entre 271 y 298 V.

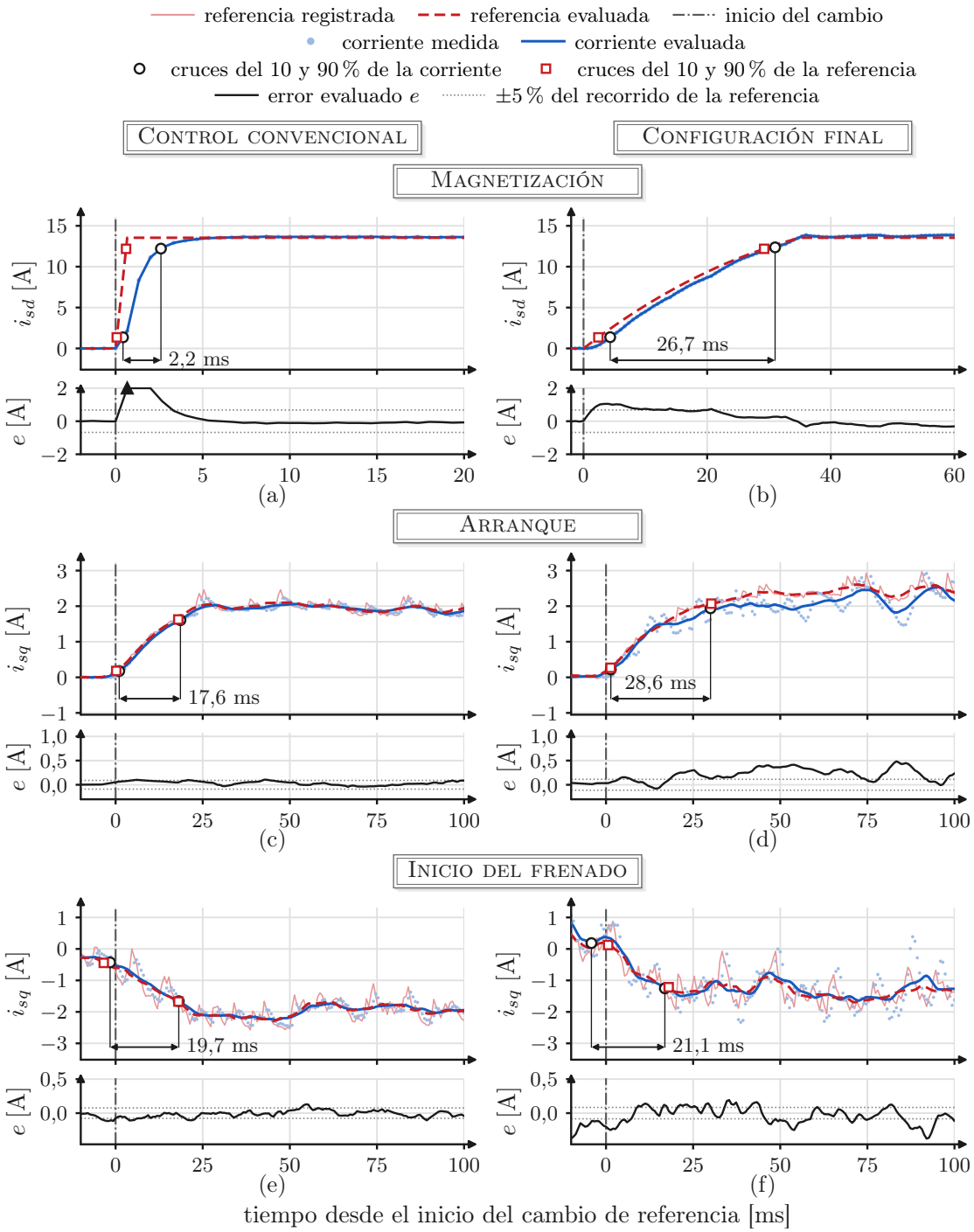


Figura 8.3: Seguimiento de la referencia de corriente medido por la unidad de control durante la magnetización, el arranque y el comienzo del frenado. (a), (c) y (e) control convencional; (b), (d) y (f) configuración final, con las tres correcciones. Bajo cada gráfica, el error $e = i^* - i$. En i_{sq}, la referencia y la corriente llevan una media móvil de 10 ms sin desfase; la cota marca los cruces del 10 y el 90 % de la corriente. Sólo en la magnetización cambia la escala de tiempo, 20 ms para el control convencional y 60 ms para la configuración final, para ver en ambos el comienzo del estado estacionario. Condiciones: ±0,30 pu sin carga y *DC link* entre 281 y 286 V.

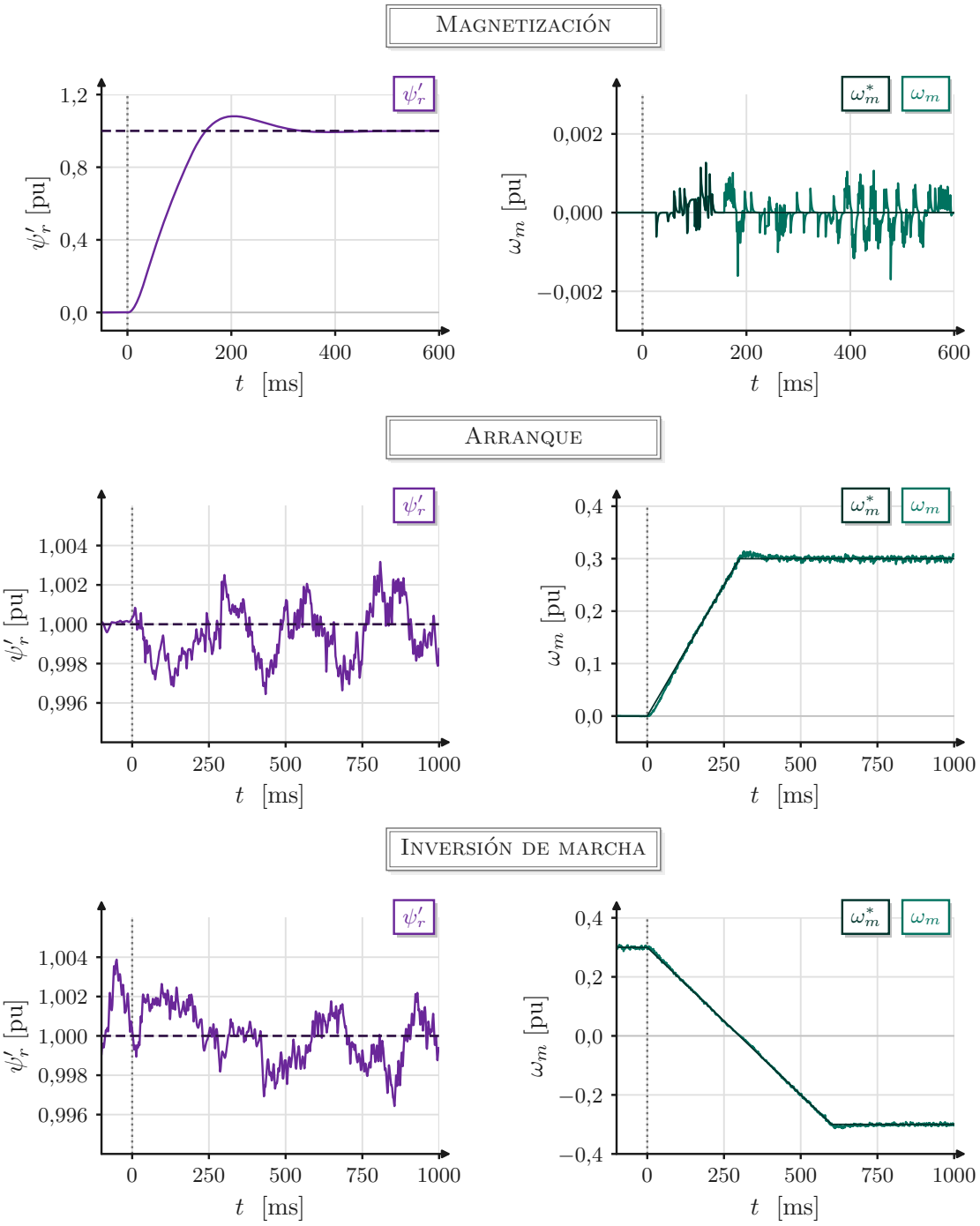


Figura 8.4: Configuración final a ±0,30 pu, medida por la unidad de control. Una fila por maniobra: a la izquierda, flujo de rotor estimado; la línea discontinua indica su referencia de 1 pu; a la derecha, velocidad con su referencia. La velocidad es la medida interna sin corregir, en la que 0,30 pu son 450 rpm. En el arranque y la inversión se amplía la escala del flujo alrededor de 1 pu para mostrar su variación. Condiciones: las de la Tabla 8.2; el origen de tiempos es el comienzo de la referencia de i_{sd} en la magnetización y el de la rampa de velocidad en las otras maniobras.

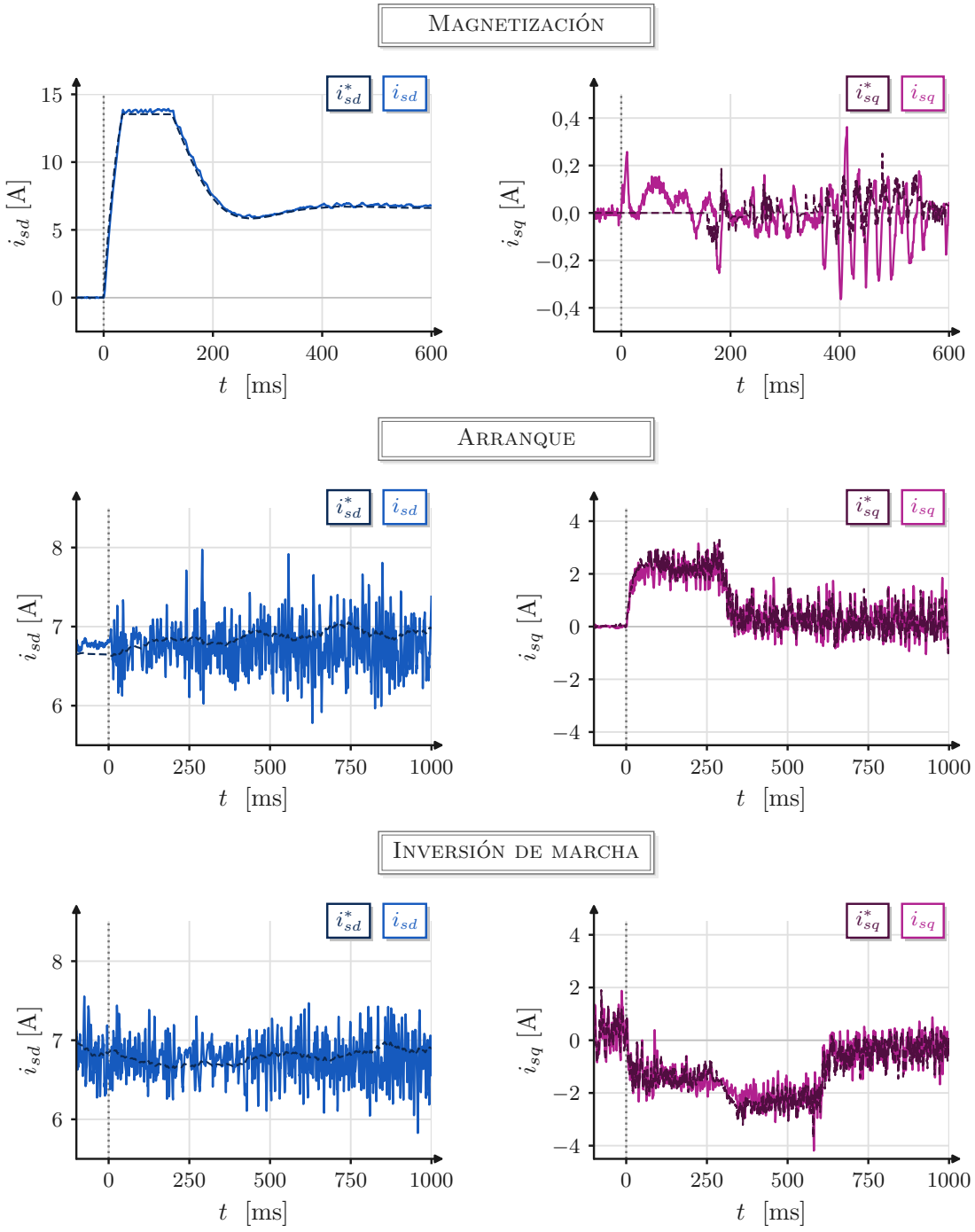


Figura 8.5: Corrientes del marco de la configuración final a ±0,30 pu, medidas por la unidad de control, con sus referencias a trazos. Una fila por maniobra: a la izquierda, i_{sd}; a la derecha, i_{sq}. El arranque y la inversión comparten la escala de cada corriente. Condiciones: las de la Tabla 8.2; el origen de tiempos es el de la Figura 8.4.

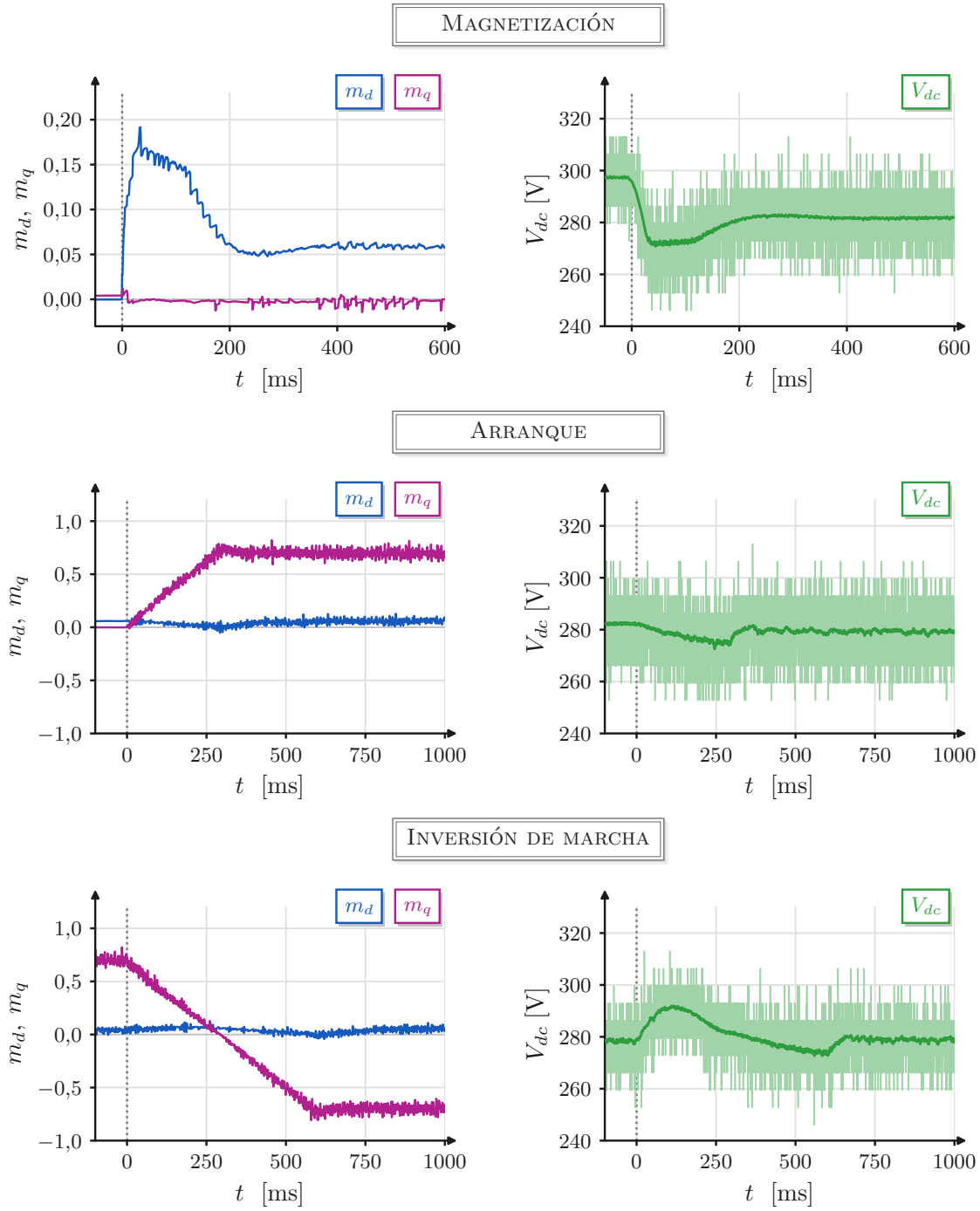


Figura 8.6: Modulantes y *DC link* de la configuración final a ±0,30 pu. Una fila por maniobra: a la izquierda, m_d y m_q registradas por la unidad de control antes del modulador; a la derecha, el *DC link* medido con el osciloscopio en las capturas de la Figura 8.7, en tono claro la señal medida y en tono oscuro su media de 1 ms. El arranque y la inversión comparten la escala de cada magnitud. Condiciones: las de la Tabla 8.2 y de la Figura 8.7; el origen de tiempos es el de la Figura 8.4 a la izquierda y el de la Figura 8.7 a la derecha.

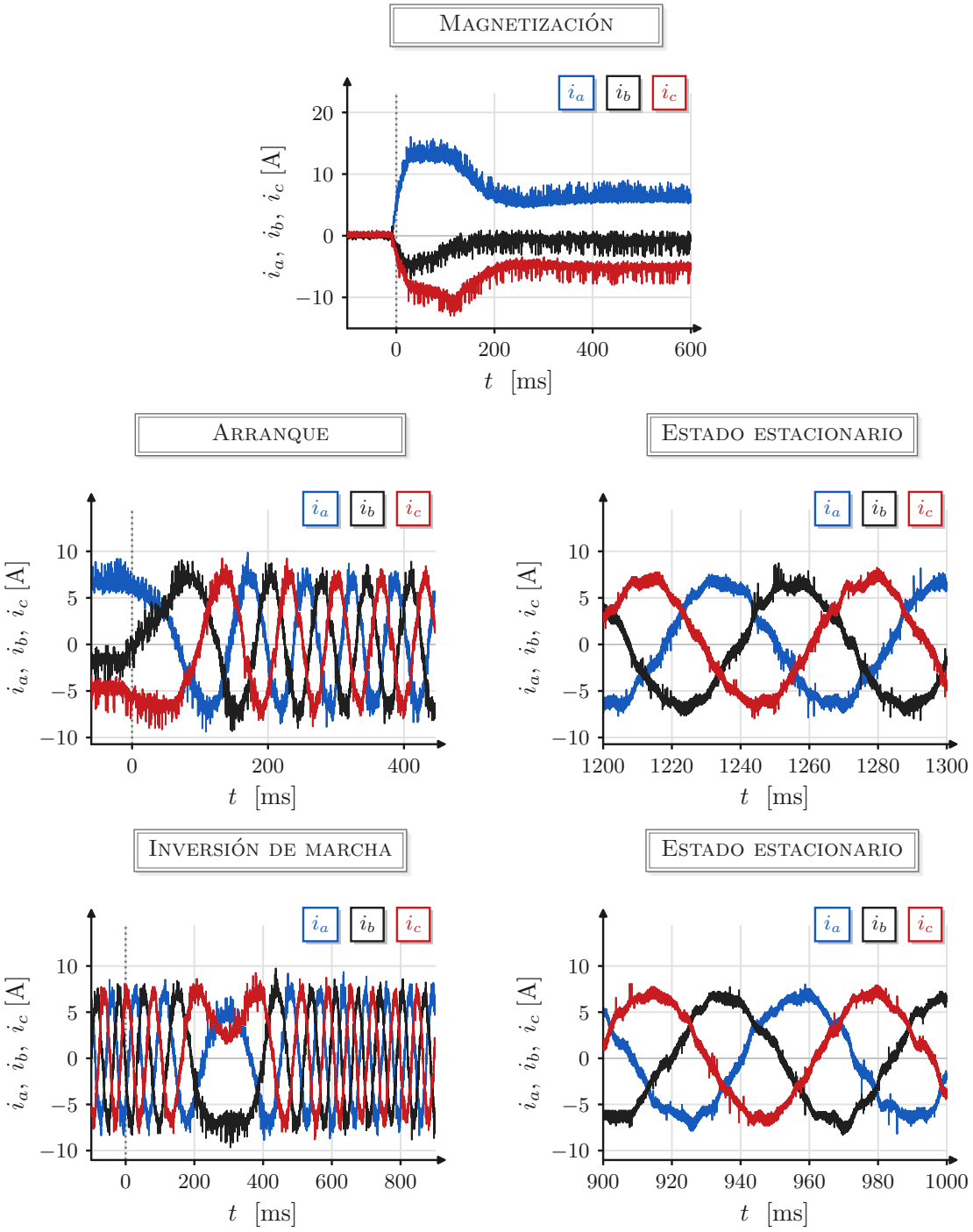


Figura 8.7: Corrientes de fase de la configuración final a ±0,30 pu, medidas en bornes con el osciloscopio. Arriba, la magnetización; al medio y abajo, el arranque y la inversión de marcha, con 100 ms de estado estacionario a la derecha. Condiciones: ±0,30 pu sin carga, *DC link* entre 271 y 298 V; el origen de tiempos es la aparición de la corriente, el despegue de su ángulo y el comienzo de la rampa.

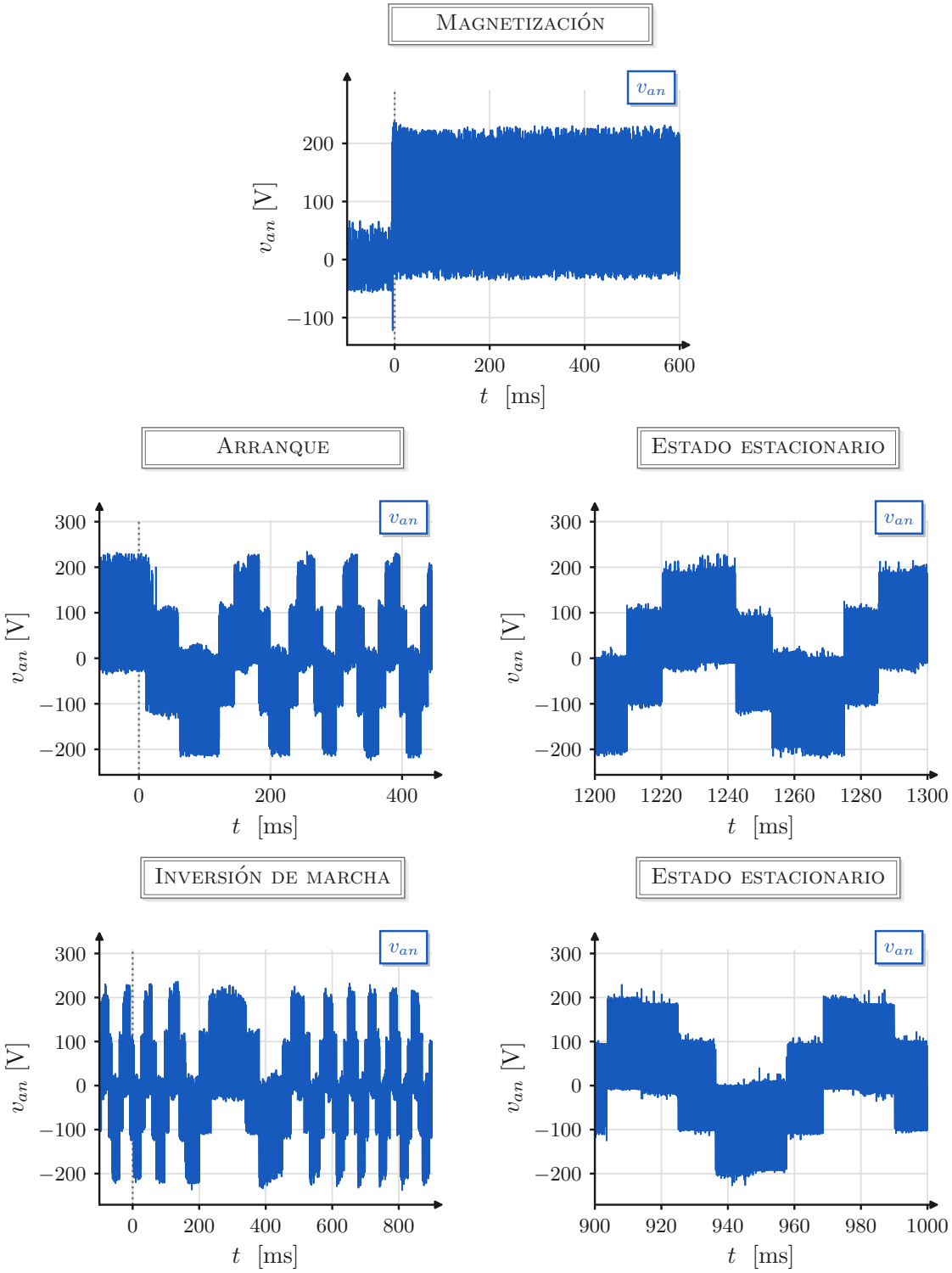


Figura 8.8: Tensión fase–neutro v_{an} de la configuración final a ±0,30 pu, medida con el osciloscopio en las mismas maniobras y ventanas de la Figura 8.7. Condiciones: las de la Figura 8.7.

8.5. Funcionamiento a 1050 rpm

La velocidad más alta ensayada fue 1050 rpm, con el *DC link* en 478 V y sin carga. La referencia interna fue ±0,674 pu. Como en el Capítulo 7, el factor del *encoder* sólo corrige el ángulo. Valió 1,0389 hacia adelante y 1,0421 hacia atrás, y con él la referencia corresponde a unos ±1050 rpm. La frecuencia de las corrientes medida con el osciloscopio, 35,013 Hz, corresponde a unos 1050 rpm. No se usó un instrumento independiente de velocidad. Esta sección presenta sólo la configuración final, que conservó las tres correcciones con los valores de la Tabla 8.1. A esta velocidad la ganancia adaptativa refina más el eje directo, con $k_{\text{adapt},d} = 4$, y ya no conserva $P = 1$.

8.5.1. El ruido y el filtro de la velocidad

Al subir la velocidad crece el ruido. La velocidad medida por el *encoder* trae más ruido, el lazo de velocidad lo convierte en ruido de la referencia i_{sq}^{*} y la ganancia adaptativa lo amplifica en la observación del agente. La Tabla 8.3 muestra que, entre 459 y 928 rpm, la desviación estándar de la velocidad medida se duplicó con creces y la de i_{sd} se multiplicó por cuatro. También creció el armónico de seis veces la frecuencia de las corrientes asociado al tiempo muerto del inversor.

Tabla 8.3: Ruido al subir la velocidad y efecto del filtro de la velocidad del lazo de velocidad, medidos por la unidad de control con el Agente 2 refinado antes del ángulo al día, con ganancias del error de 12 y 4 y la referencia de cuadratura del agente filtrada a 100 Hz. Arriba, últimos 400 ms del arranque, con el filtro de 40 Hz a 459 rpm y de 25 Hz a 928 rpm; f_e es la frecuencia de las corrientes. Abajo, últimos 300 ms del arranque y de la inversión de marcha. Condiciones: sin carga; arriba, *DC link* de 282 y 429 V, bancos del 25 y 26 de septiembre; abajo, ±930 rpm, *DC link* entre 421 y 485 V, banco del 26 de septiembre.

Indicador	Unidad	459 rpm	928 rpm
Desviación estándar de la velocidad medida	10^{-3} pu	2,8	6,1
Desviación estándar de i_{sd}	A	0,10	0,44
Armónico de $6f_e$ en i_{sd}	mA	81	283
Filtro de la velocidad del lazo de velocidad		40 Hz	25 Hz
RMSE de i_{sq}, después del arranque / de la inversión	mA	936 / 996	799 / 806
Desviación estándar de i_{sq}^{*}, arranque / inversión	A	0,58 / 0,50	0,54 / 0,36
Desviación estándar de i_{sd}, arranque / inversión	A	0,44 / 0,45	0,44 / 0,42

Para limitar ese ruido, el controlador PI de velocidad usa la velocidad medida con un filtro de primer orden,

$$\omega_f[k] = \omega_f[k-1] + \alpha\big(\omega[k] - \omega_f[k-1]\big), \qquad \alpha = 2\pi f_c T_s = 0{,}0524, \tag{8.5}$$

con frecuencia de corte $f_c = 25$ Hz. El ángulo de orientación, la velocidad que observa el agente y las protecciones conservan el filtro de 80 Hz del Capítulo 7. A 1050 rpm el lazo de velocidad usó este filtro. A ±0,30 pu el lazo de velocidad usó la velocidad de 80 Hz, como en el Capítulo 7.

El filtro se evaluó en el banco cerca de 930 rpm, comparando 40 y 25 Hz en la misma sesión. Con 25 Hz, el RMSE de cuadratura bajó de 936 a 799 mA después del arranque y de 996 a 806 mA después de la inversión de marcha, y la desviación estándar de i_{sq}^{*}

también bajó. El eje directo casi no cambió. Esa comparación se hizo antes del ángulo al día y con la referencia de cuadratura del agente filtrada a 100 Hz, un filtro que las dos configuraciones finales ya no usan.

8.5.2. La magnetización, el arranque y la inversión de marcha

La Figura 8.9 muestra la configuración final en las tres maniobras. En la magnetización, el flujo alcanzó 1,080 pu y el eje permaneció entre −8,5 y 2,6 rpm. Durante el arranque y la inversión de marcha, el flujo se mantuvo entre 0,995 y 1,005 pu. La velocidad llegó a 1085 rpm en el arranque y a −1082 rpm en la inversión.

La Tabla 8.4 resume los indicadores de corriente de la configuración final. La Figura 8.10 muestra sus corrientes con sus referencias, y la Figura 8.11, sus modulantes junto al *DC link*. La configuración final actuó en la primera muestra del registro, a 1,00 ms. En la magnetización, la corriente se asentó a los 34,0 ms.

Tabla 8.4: Configuración final a 1050 rpm, medida por la unidad de control, con las cuentas de la Tabla 8.2. Condiciones: sin carga, *DC link* de 478 V al comenzar, filtro de 25 Hz en la velocidad del lazo de velocidad.

Indicador	Unidad	Configuración final
Indicadores transitorios		
Asentamiento de i_{sd} en la magnetización (±5 %)	ms	34,0
Estado estacionario, RMSE de $e = i^* - i$		
Eje directo, después de la magnetización	mA	75
Eje de cuadratura, después de la magnetización	mA	62
Eje directo, después del arranque	mA	389
Eje de cuadratura, después del arranque	mA	423
Eje directo, después de la inversión	mA	573
Eje de cuadratura, después de la inversión	mA	526

La Figura 8.12 muestra las corrientes de fase medidas con el osciloscopio en cinco etapas de la maniobra, y la Figura 8.13, la tensión fase–neutro en las mismas capturas. En la magnetización, la corriente de la fase a se mantuvo en 13,4 A durante el refuerzo, con 13,54 A de referencia del eje directo, y se asentó en 6,80 A, mientras el *DC link* bajó de 484 a 470 V. En estado estacionario, la frecuencia de las corrientes fue 35,013 Hz, frente a 35,016 Hz del marco calculado, y su amplitud fue 6,63 A en el osciloscopio y 6,79 A en la unidad de control. En la inversión de marcha, la frecuencia de las corrientes pasó de +35,085 a −35,245 Hz sin una interrupción visible. El flujo estimado no bajó de 0,9952 pu y el *DC link* llegó a 534 V, bajo los 570 V del *chopper*.

La magnetización mejoró respecto de la configuración de ±0,30 pu: el RMSE del eje directo en estado estacionario bajó de 148 a 75 mA. Esa mejora no viene del filtro de 25 Hz, que actúa sobre la velocidad. En la comparación de 40 y 25 Hz, el RMSE del eje directo en la magnetización fue 56 y 60 mA, y el sobrepaso del flujo, 7,90 y 7,89 %. La mejora coincide con el paso más fino del eje directo, con $k_{\mathrm{adapt},d} = 4$ en lugar de 2. El flujo del agente sobrepasó su valor final un 8,00 %, por su lazo de flujo con las ganancias a la mitad y el alisamiento de 60 ms.

El límite a esta velocidad apareció tras la inversión de marcha. Allí el RMSE de estado estacionario fue 573 mA en el eje directo y 526 mA en el de cuadratura, frente a 389 y 423 mA después del arranque. El filtro de 25 Hz no reduce el ruido del eje directo.

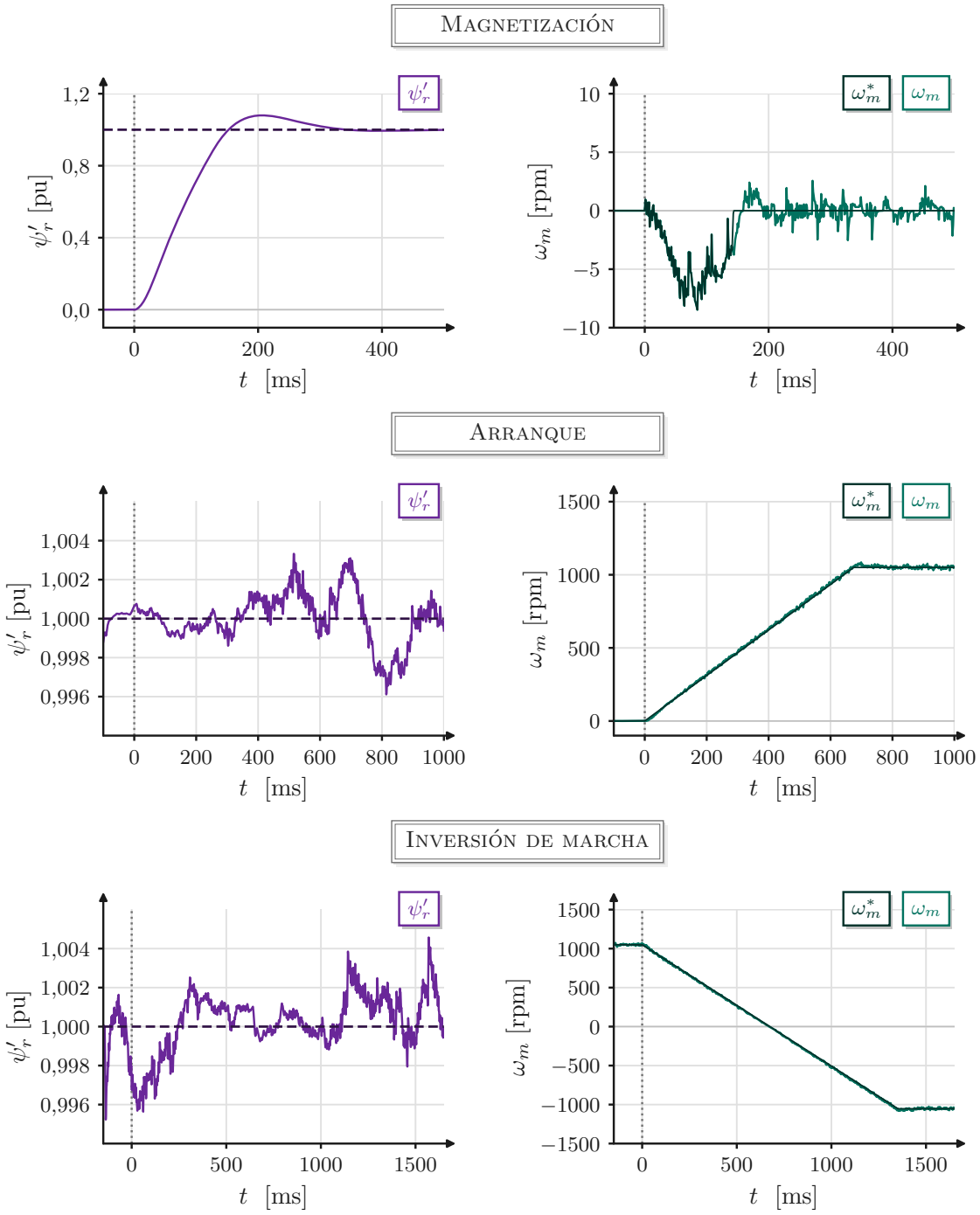


Figura 8.9: Configuración final a 1050 rpm. Una fila por maniobra: magnetización, arranque e inversión de marcha. A la izquierda, flujo de rotor estimado; la línea discontinua indica su referencia de 1 pu; a la derecha, velocidad con su referencia, en rpm, con el factor del *encoder* de cada sentido. En el arranque y la inversión se amplía la escala del flujo alrededor de 1 pu para mostrar su variación. Condiciones: las de la Tabla 8.4; el origen de tiempos es el comienzo de la referencia de i_{sd} en la magnetización y el de la rampa de velocidad en las otras maniobras.

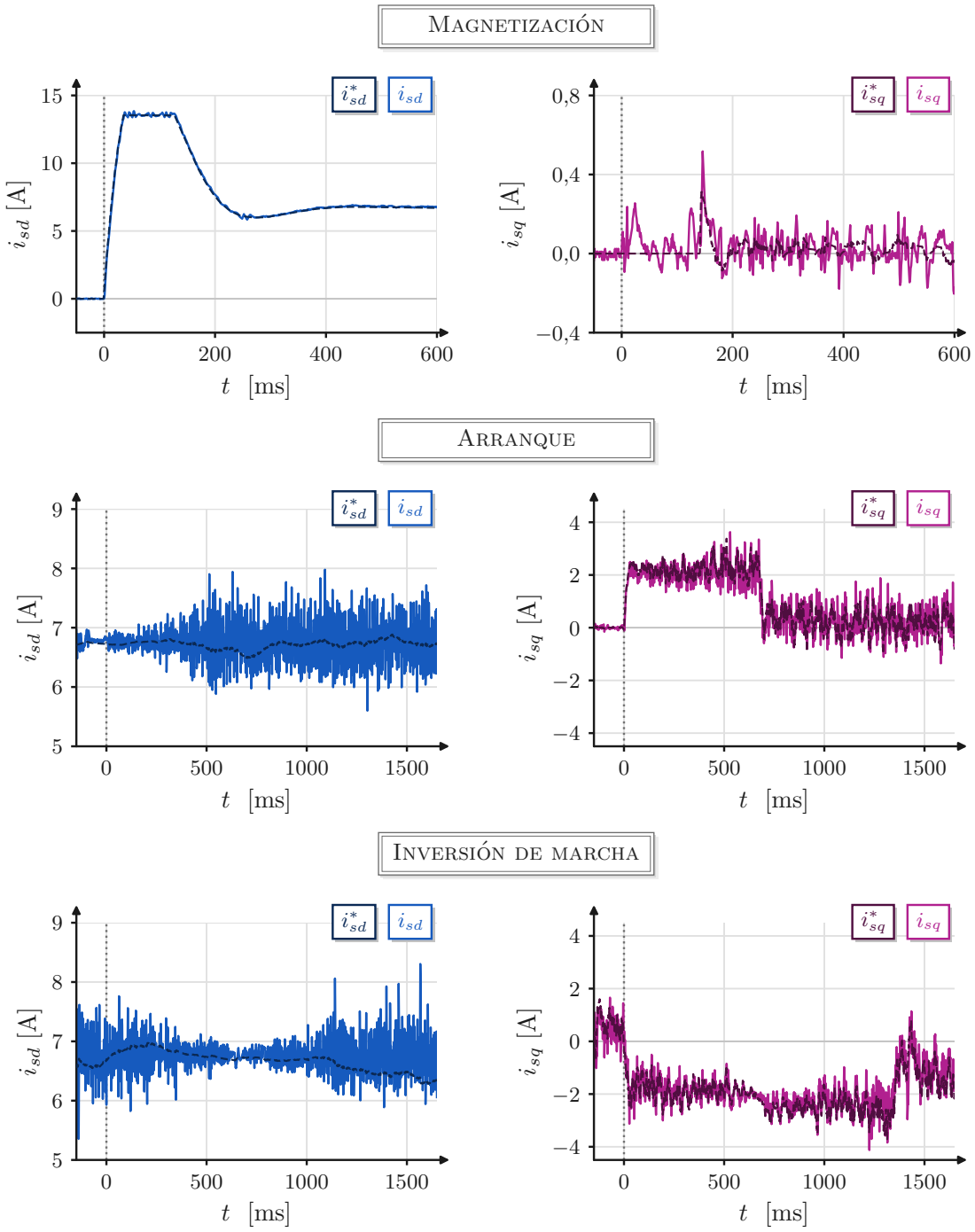


Figura 8.10: Corrientes del marco de la configuración final a 1050 rpm, medidas por la unidad de control, con sus referencias a trazos. Una fila por maniobra: a la izquierda, i_{sd}; a la derecha, i_{sq}. El arranque y la inversión comparten la escala de cada corriente. Condiciones: las de la Tabla 8.4; el origen de tiempos es el de la Figura 8.9.

Tras la inversión aparecieron sesgos de −421 y −316 mA, con aporte de la escala del *encoder* hacia atrás, 1,0459 medida con el osciloscopio y 1,0421 configurada.

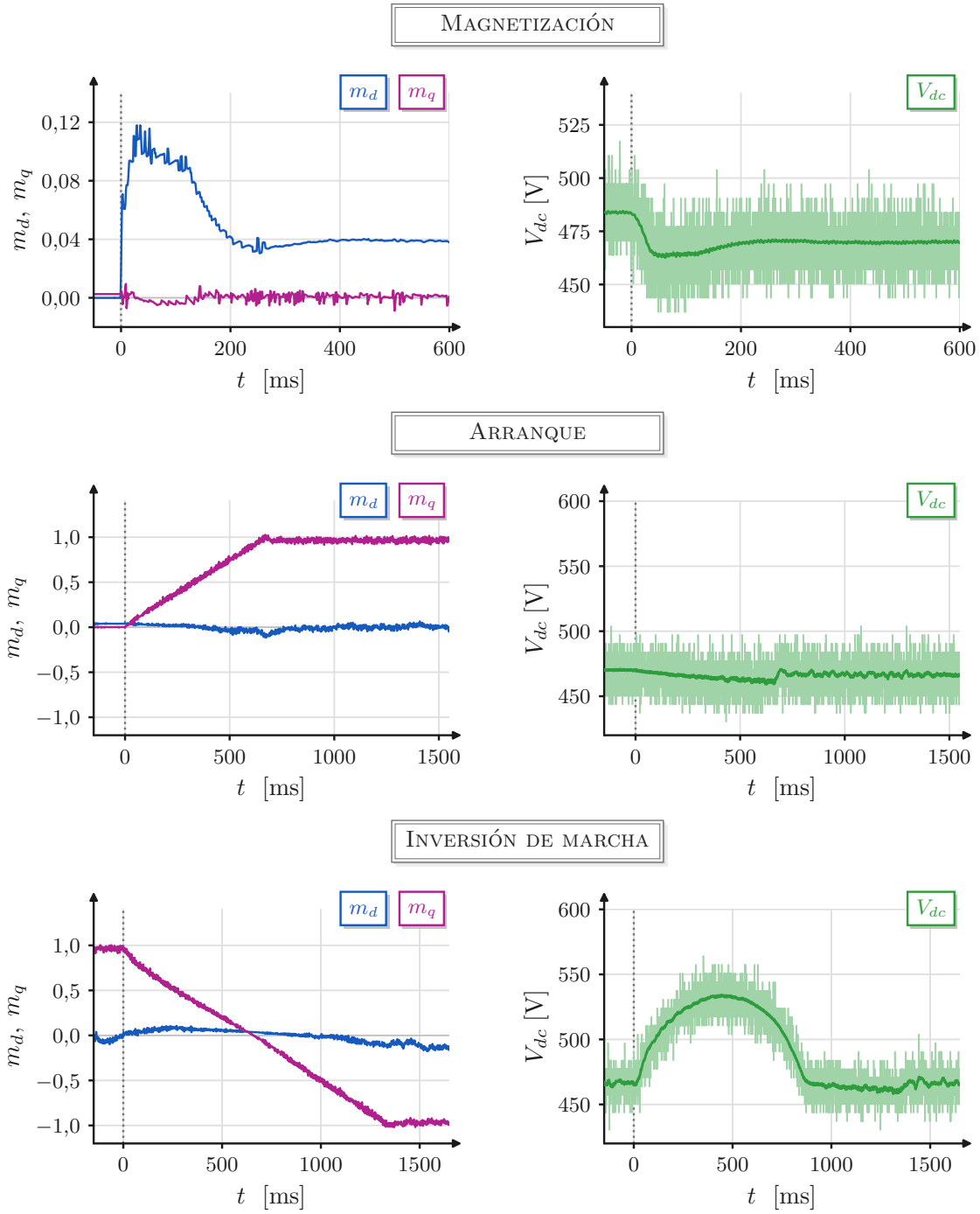


Figura 8.11: Modulantes y *DC link* de la configuración final a 1050 rpm. Una fila por maniobra: a la izquierda, m_d y m_q registradas por la unidad de control antes del modulador; a la derecha, el *DC link* medido con el osciloscopio en las capturas de la Figura 8.12, en tono claro la señal medida y en tono oscuro su media de 1 ms. El arranque y la inversión comparten la escala de cada magnitud. Condiciones: las de la Tabla 8.4; el origen de tiempos es el de la Figura 8.9 a la izquierda y el de la Figura 8.12 a la derecha.

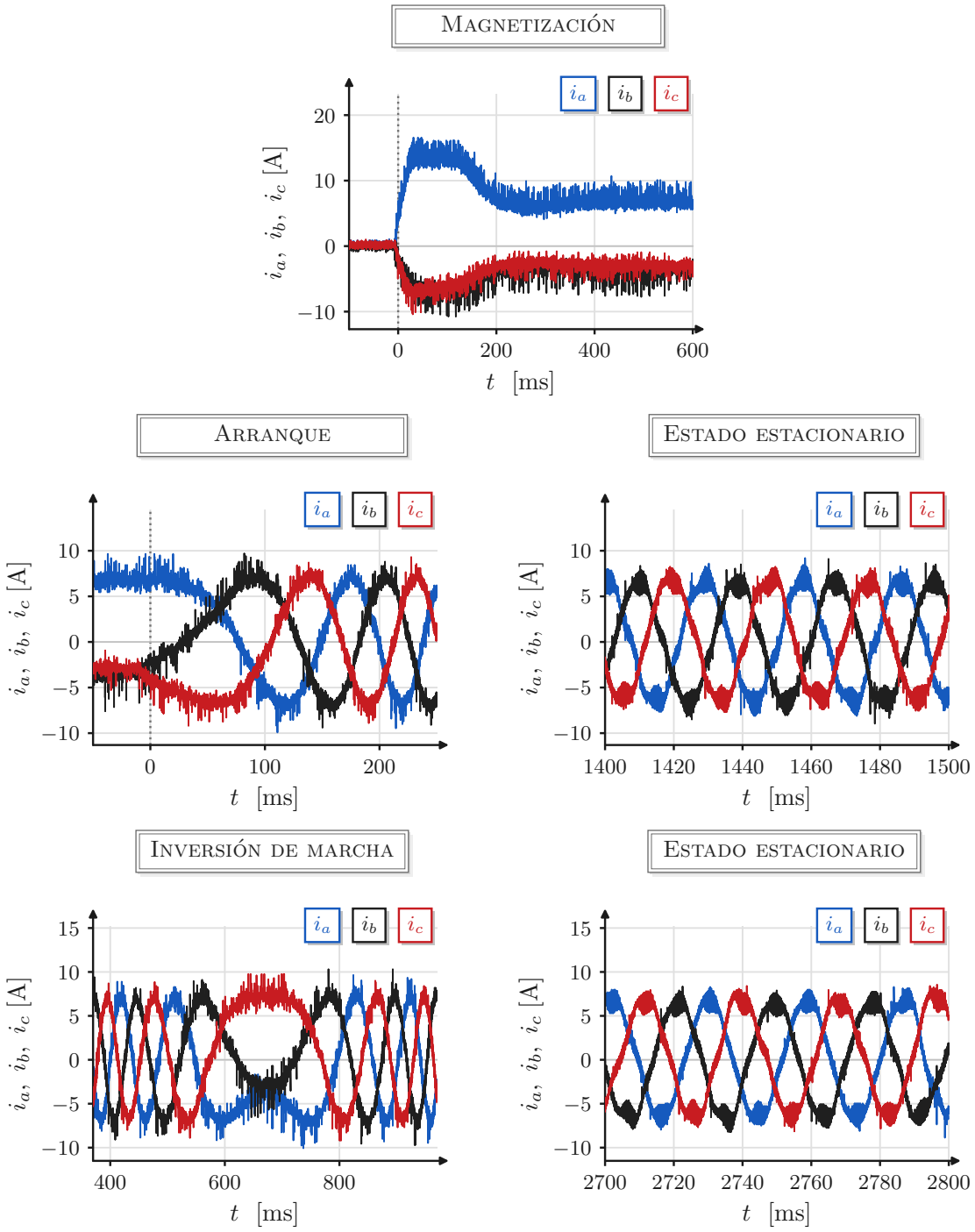


Figura 8.12: Corrientes de fase de la configuración final a 1050 rpm, medidas en bornes con el osciloscopio. Arriba, la magnetización; al medio y abajo, el arranque y la inversión de marcha, con 100 ms de estado estacionario a la derecha. La inversión pasa por velocidad cero a los 673 ms. Condiciones: las de la Tabla 8.4, *DC link* entre 458 y 534 V; el origen de tiempos es la aparición de la corriente, el despegue de su ángulo y el comienzo de la rampa.

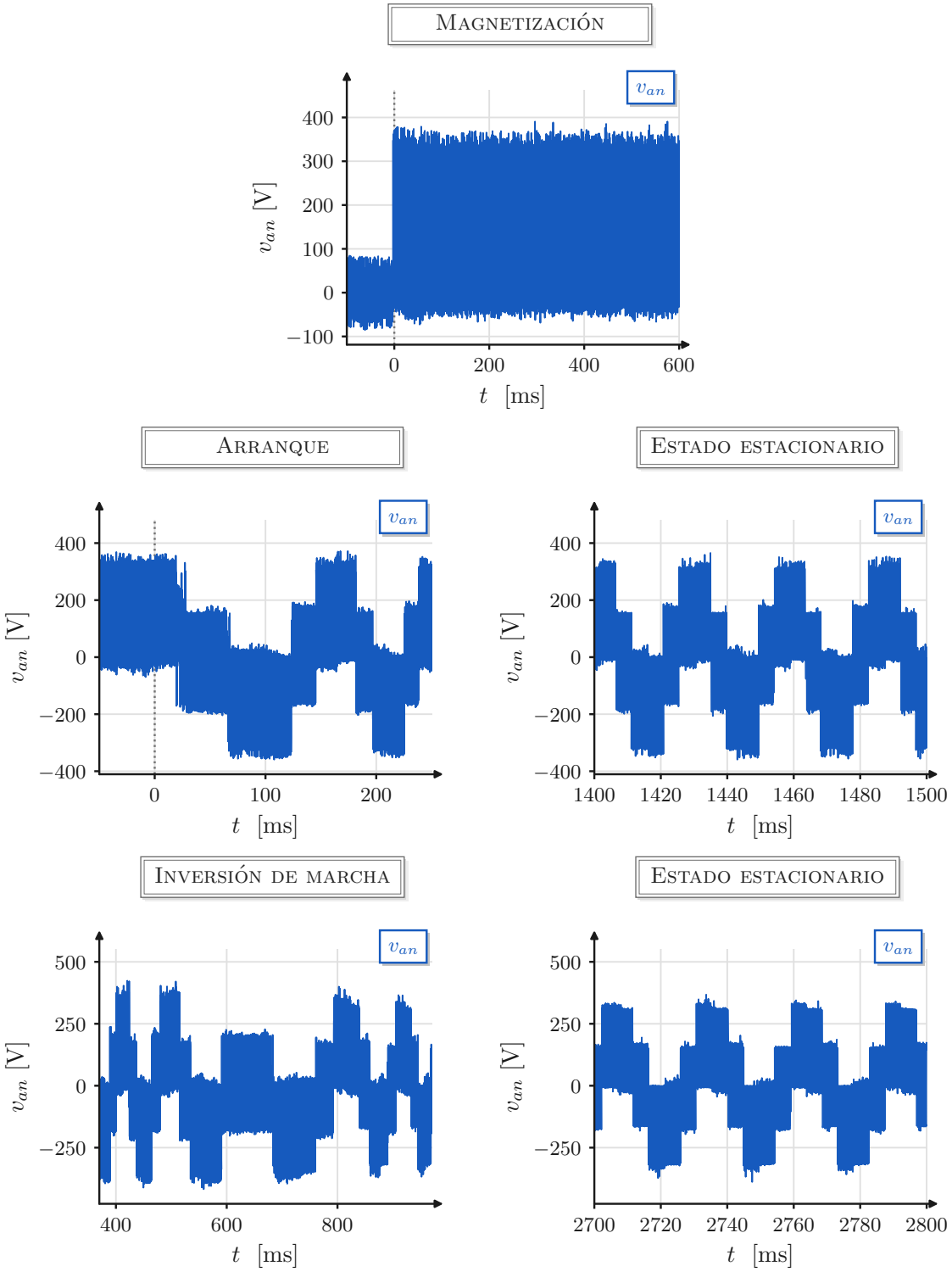


Figura 8.13: Tensión fase–neutro v_{an} de la configuración final a 1050 rpm, medida con el osciloscopio en las mismas maniobras y ventanas de la Figura 8.12. Condiciones: las de la Figura 8.12.

Capítulo 9

Conclusiones

9.1. Respuesta a la hipótesis

Los resultados respaldan la hipótesis dentro de las condiciones ensayadas. Una política aprendida con una carga $R-L$ trifásica simulada y una perturbación de tensión puede utilizarse, con sus pesos fijos, en el lazo interno de corriente de una máquina de inducción. El Agente 2 refinado completó la magnetización, el arranque hasta +0,30 pu y la inversión de marcha hasta −0,30 pu, tanto en simulación como en el banco, sin aprendizaje en línea.

Con las tres correcciones del Capítulo 8, que actúan fuera de la red y no cambian sus pesos, el mismo agente completó también esas maniobras en el banco a unos 1050 rpm, con una referencia interna de ±0,674 pu y el *DC link* en 478 V al comenzar.

La transferencia corresponde al accionamiento completo. La red genera la acción del lazo de corriente, mientras se conservan la orientación de campo, el estimador, los lazos externos y la modulación [1,2]. Por ello, completar las maniobras demuestra que la política puede integrarse en el accionamiento, pero no implica que supere al control convencional ni que la red elimine por sí sola el error en estado estacionario.

En las condiciones ensayadas, la implementación también cumplió la restricción temporal. La lectura del ciclo del agente dentro de la interrupción, tomada al terminar la campaña, fue de 239,2 μs, frente a un período de control de 333,3 μs, y el contador de ciclos tardíos quedó en cero. Son valores medidos en esas condiciones, no una cota de peor caso. Para la configuración final del Capítulo 8 no se presenta una lectura equivalente.

9.2. Cumplimiento de los objetivos

Se cumplieron el objetivo general y los seis objetivos específicos de la Sección 1.3, con el alcance que resume la Tabla 9.1. Se formuló el entorno de aprendizaje, se definieron la observación, las acciones y la recompensa, y se verificó el agente sobre la carga $R-L$ en simulación y en el banco. Luego se transfirió la política al modelo de la máquina, se ajustaron los lazos externos sin modificar sus pesos y se implementó el agente en la plataforma BRAIn.

La validación experimental sobre la máquina comparó el accionamiento con el agente y el control convencional durante la magnetización, el arranque y la inversión de marcha.

Se evaluaron el seguimiento de corriente, el flujo estimado, la velocidad, las corrientes de fase y la tensión del *DC link*. El torque se estimó con i_{sq} y el flujo estimado, sin medición independiente. El esfuerzo de conmutación se describió con el número de transiciones por pierna. La calidad de los disparos se describió con el tiempo muerto en las compuertas. Las pérdidas no se midieron.

El Capítulo 8 complementa el sexto objetivo específico. Mejora el transitorio del agente sin degradar su estado estacionario a ±0,30 pu y lleva el mismo ensayo hasta unos 1050 rpm.

Tabla 9.1: Objetivos de la memoria, evidencia que los respalda y alcance de esa evidencia.

Objetivo	Respaldo	Alcance de la evidencia
General	Capítulo 7 y Capítulo 8	Máquina de 5,5 kW sin carga mecánica, a ±0,30 pu y con el *DC link* entre 281 y 285 V. Con la configuración final del Capítulo 8, entre 281 y 286 V a ±0,30 pu, y también a unos 1050 rpm, con 478 V al comenzar. El agente sólo reemplaza el control de corriente.
1. Fundamentos	Secciones 2.1.1 y 2.2	Revisión de los fundamentos, sin evaluación de resultados.
2. Control de referencia	Capítulo 5; Figura 6.3	Modelo del PI implementado, con el factor $\sqrt{3}/2$ y sus compensaciones. Sin ensayo aislado de pequeña señal del lazo de corriente.
3. Entorno $R-L$	Capítulo 3; Sección A.3.1	Perturbación de tensión en el eje de cuadratura dimensionada con datos nominales de la máquina. La máquina se controla sin reentrenar la red.
4. Simulación	Tablas 4.3, 6.3 y 6.5; Apéndice D	Maniobras hasta ±0,30 pu sin carga mecánica. La mejora de la reflexión se demuestra sólo en simulación. Los escalones de torque se simularon sin ensayo en el banco.
5. Comparación	Tablas 7.2 y 7.3; Secciones A.7 y A.8	Torque estimado con i_{sq} y el flujo estimado. Esfuerzo de conmutación como conteo de transiciones por pierna. Calidad de los disparos como tiempo muerto medido en las compuertas. Sin medición independiente de torque ni de pérdidas.
6. Validación experimental	Secciones 4.2 y 4.3; Capítulo 7 y Capítulo 8	Sobre la carga $R-L$, sólo el agente. En la máquina, ambos controles en sesiones distintas. A unos 1050 rpm, sólo la configuración final del agente.

9.3. Resultados principales

Sobre la carga $R-L$ física, cuya inductancia equivalente se estimó en 10,41 mH, el Agente 2 obtuvo un RMSE de 164,4 mA en el eje directo y 264,7 mA en el de cuadratura. El mayor sesgo fue de 372 mA. Este ensayo confirmó el seguimiento con error en estado estacionario; no incluyó un control convencional sobre la misma carga.

Sobre la máquina, ambos accionamientos establecieron el flujo y siguieron la velocidad. En los últimos 200 ms de cada registro, el RMSE de i_{sd} varió entre 310,8 y 502,5 mA con el agente, y entre 18,3 y 106,3 mA con el control convencional. Para i_{sq}, varió entre 126,7 y

627,3 mA con el agente, y entre 16,6 y 297,5 mA con el control convencional. El valor RMS de los seis sesgos fue de 253 mA con el agente y 5,94 mA con el control convencional, según la Tabla 7.3. Estos errores corresponden al seguimiento del lazo interno. Los lazos externos de flujo y velocidad compensaron parte de las desviaciones al modificar las referencias i^*_{sd} e i^*_{sq}; por ello, el flujo estimado y la velocidad alcanzaron sus referencias aunque el agente mantuviera error de corriente en estado estacionario.

En simulación, la reflexión del eje de cuadratura redujo el error global un 46 %, aunque el error del eje directo con velocidad negativa aumentó un 35 %. Su efecto experimental no se aisló, porque no se realizaron campañas con la reflexión activada y desactivada.

El transitorio del agente llegaba tarde por causas que están fuera de la red. Al arrancar a +0,30 pu, la corriente reaccionó 6,0 ms después de su referencia, frente a 0,7 ms con el control convencional. El agente recibía la referencia de cuadratura y el ángulo del marco calculados en el período anterior. Cada corrección se comprobó en el banco. La compensación bajó la corriente media faltante de 0,59 a 0,23 A. El escalamiento $g = 2$ y $k = 0{,}5$ duplicó la resolución con $P = gk = 1$, y el asentamiento de i_{sd} bajó de 119,3 a 33,3 ms. La corrección angular llevó el sesgo de cuadratura cerca de 1050 rpm de +739 a −108 mA.

Con las tres correcciones juntas, a ±0,30 pu, el RMSE de la configuración final fue menor que el de la configuración del Capítulo 7 en las seis ventanas de estado estacionario de la Tabla 8.2. El control convencional conservó el menor RMSE, entre 17 y 297 mA. A unos 1050 rpm, la velocidad siguió la rampa de su referencia y la inversión de marcha pasó por cero a los 671 ms. El agente actuó en la primera muestra del registro, a 1,00 ms. El límite apareció tras la inversión de marcha, con un RMSE de 573 mA en el eje directo y 526 mA en el de cuadratura, frente a 389 y 423 mA después del arranque, según la Tabla 8.4.

9.4. Limitaciones

La validación se realizó sin carga mecánica, a ±0,30 pu y con el *DC link* entre 281 y 285 V. No se aplicaron escalones de carga ni se ensayaron la región de torque elevado, todo el intervalo de velocidad o la saturación de tensión. Con la configuración final del Capítulo 8, el *DC link* estuvo entre 281 y 286 V a ±0,30 pu, y a unos 1050 rpm comenzó en 478 V. La velocidad nominal pide unos 653 V de *DC link*, según la estimación de la Tabla A.6. Los escalones de carga sólo se aplicaron en simulación, en el Apéndice D.

Los accionamientos utilizan el mismo lazo de velocidad, con su filtro y su rampa, el mismo inversor y el mismo método de modulación PWM. Difieren en el control de corriente y en el ajuste del lazo de flujo. Con el agente, la referencia de flujo pasa además por un alisamiento de 60 ms, y los límites previos al modulador son distintos en cada accionamiento, como detalla la Tabla A.10. Estas diferencias también intervienen en la respuesta de cada accionamiento completo. Además, el flujo se obtiene del estimador y el torque se estima a partir de i_{sq}. Estas condiciones impiden atribuir toda diferencia al control del lazo interno.

Las tres correcciones del Capítulo 8 tienen límites propios. El giro adicional del ángulo es una calibración de esta unidad de control. La compensación del retardo actúa sólo en los primeros 50 ms de un arranque desde el reposo. La ganancia adaptativa amplifica el ruido de corriente, y su valor es una elección para este agente, sin un análisis de estabilidad. A mayor velocidad quedan errores sin una causa comprobada. Después de la inversión de

marcha a unos 1050 rpm aparecieron sesgos de -421 y -316 mA, con aporte de la escala del *encoder* hacia atrás, sin que pueda atribuirse todo el error a una sola causa.

La constante equivalente de 12 ms se identificó en pequeña señal sobre el modelo; no es una medición del banco ni describe sus transitorios de gran amplitud. La red usada en los ensayos es la del Agente 2 refinado: sus pesos están identificados. Lo que no se guardó es la huella digital del programa completo instalado en la unidad de control, que además de la red construye las entradas, aplica las escalas y los límites, acumula la acción y registra las señales. Se conserva un programa compilado antes de la campaña, con esos mismos pesos, y sus fuentes lo reproducen exactamente; sin esa huella, no puede demostrarse que fuera la copia instalada. Tampoco se archivaron todos los ajustes de las pinzas de corriente. Los Apéndices B y C documentan estas salvedades.

El circuito de frenado no intervino en las campañas presentadas. La tensión del *DC link* permaneció por debajo de su umbral de 330 V y el contador de intervenciones quedó en cero, de modo que los ensayos no caracterizan su comportamiento durante un frenado sostenido. En la configuración final a unos 1050 rpm, el umbral era de 570 V y el *DC link* llegó a 534 V.

9.5. Trabajo futuro

La validación experimental puede ampliarse mediante ensayos con carga mecánica, incorporando cambios controlados de torque durante el arranque y la inversión de marcha. Esto permitiría evaluar el rechazo de perturbaciones y el seguimiento de velocidad en condiciones distintas de la operación sin carga considerada en esta memoria. En simulación, con el ajuste de sensibilidad y sin reentrenar, el sesgo de cuadratura con el torque nominal bajó de $+1{,}20$ a $+0{,}08$ A, según el Apéndice D.

El ensayo puede extenderse hasta la velocidad nominal, con un *DC link* cercano a los 653 V, revisando primero el error en estado estacionario tras la inversión de marcha, que a unos 1050 rpm fue mayor que después del arranque. Las correcciones del Capítulo 8 pueden completarse. La compensación del retardo podría actuar también al comenzar el frenado, y la ganancia adaptativa requiere un análisis de estabilidad.

También se propone comparar otras arquitecturas de redes neuronales y diferentes cantidades de neuronas. Esta evaluación debe incluir la reducción del vector de observación, la selección de las variables que realmente aportan al control y el estudio de otras resoluciones y distribuciones del espacio de acciones. Las alternativas deben compararse utilizando el mismo procedimiento de entrenamiento y las mismas métricas de seguimiento y tiempo de ejecución.

Finalmente, se pueden estudiar nuevas funciones de recompensa y una estrategia de entrenamiento que no dependa de la perturbación adicional de tensión en el eje de cuadratura utilizada durante el refinamiento. También se propone evaluar una corrección adaptativa de la acción del agente para reducir el error en estado estacionario, con ganancias independientes para los ejes directo y de cuadratura ajustadas con apoyo de un observador adaptativo o un filtro de Kalman. Cualquier modificación debe incorporarse al entorno de entrenamiento y verificarse posteriormente sobre la carga $R-L$ y la máquina.

Apéndice A

Parámetros y configuración de la plataforma

Este apéndice reúne los valores que el cuerpo emplea sin detallar. Todos proceden del modelo validado, de los ficheros de configuración o de una medición declarada.

A.1. Banco experimental y cadena de señales

El banco se presenta siguiendo el recorrido de la energía y de las señales. El variac y el rectificador alimentan el *DC link*; el inversor aplica las tensiones a la carga $R-L$ o a la máquina; y la rama de frenado limita la elevación del *DC link* cuando se habilita. La Figura A.1 reúne esa cadena y sitúa el punto medio 0 y las salidas a, b y c. Las referencias N y n empleadas en las mediciones y en el modelo se definen en el Capítulo 2.

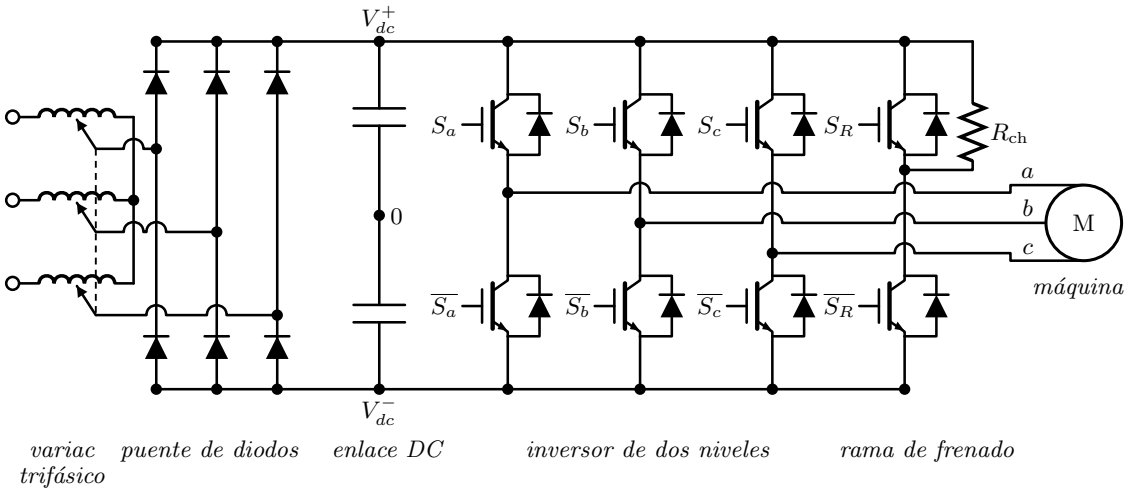


Figura A.1: Etapa de potencia del banco: variac trifásico, puente de diodos, *DC link* con punto medio 0, inversor de dos niveles, rama de frenado y máquina. Las salidas a, b y c corresponden a los puntos de salida del inversor. Elaboración propia a partir del montaje experimental.

La plataforma BRAIn cierra el lazo en tiempo real. Sus convertidores A/D reciben las mediciones; la FPGA adquiere las señales y genera la PWM con tiempo muerto; y el DSP ejecuta el IFOC con el agente o con el control convencional. La Figura A.2 distingue ese recorrido de las conexiones de supervisión, que se utilizan para programar y registrar, pero no deciden la actuación.

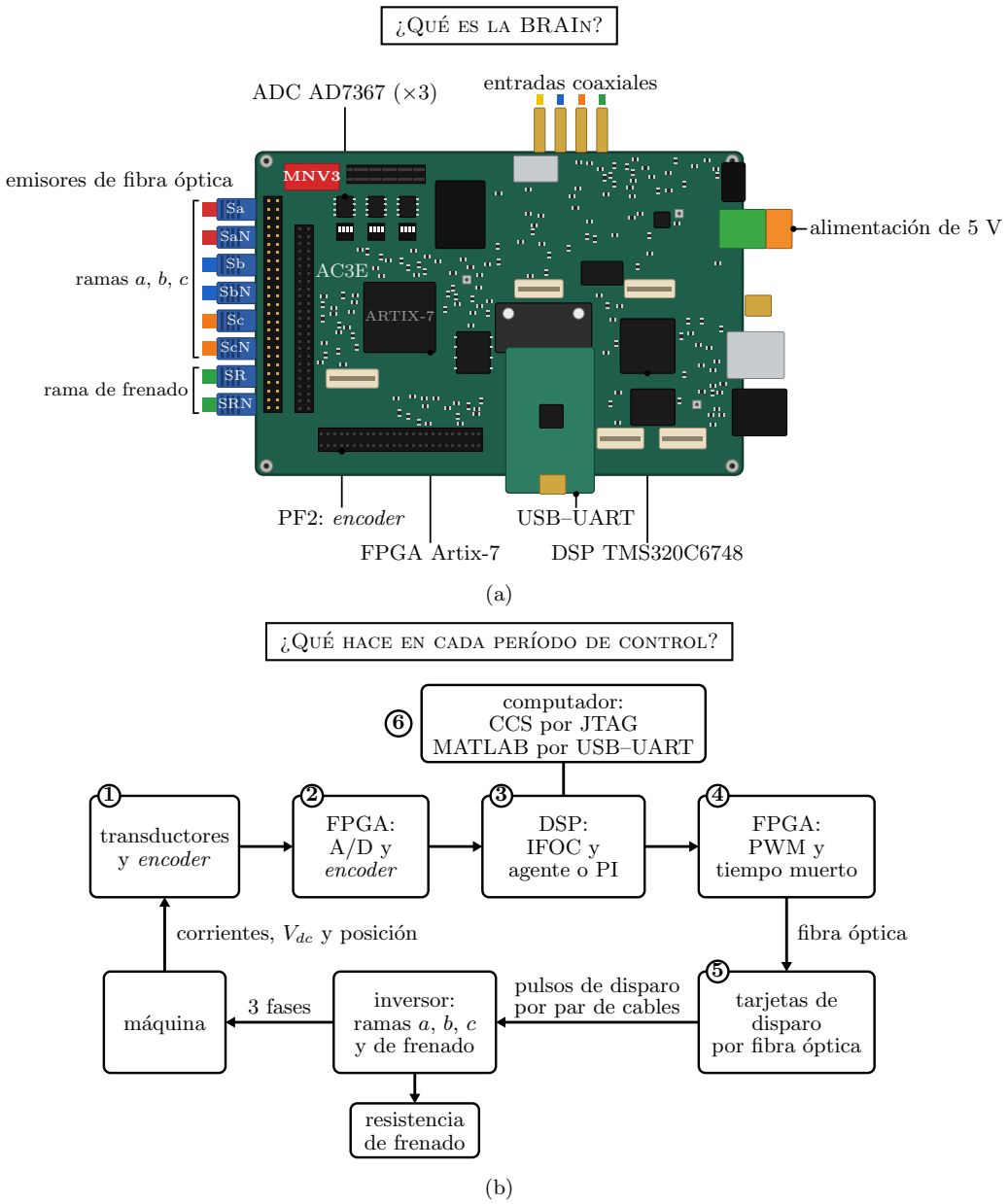


Figura A.2: Plataforma BRAIn y operaciones de un período de control. La FPGA adquiere las mediciones y genera las señales PWM; el DSP ejecuta el IFOC y el lazo de corriente seleccionado. Los pulsos llegan por fibra óptica a las tarjetas de disparo y luego al inversor. Elaboración propia a partir de la plataforma del banco [25].

Entre la unidad de control y cada módulo IGBT, una tarjeta recibe las órdenes por fibra óptica, establece la tensión compuerta–emisor y separa el control de la etapa de potencia. La Figura A.3 muestra una pierna; los demás canales repiten esa función.

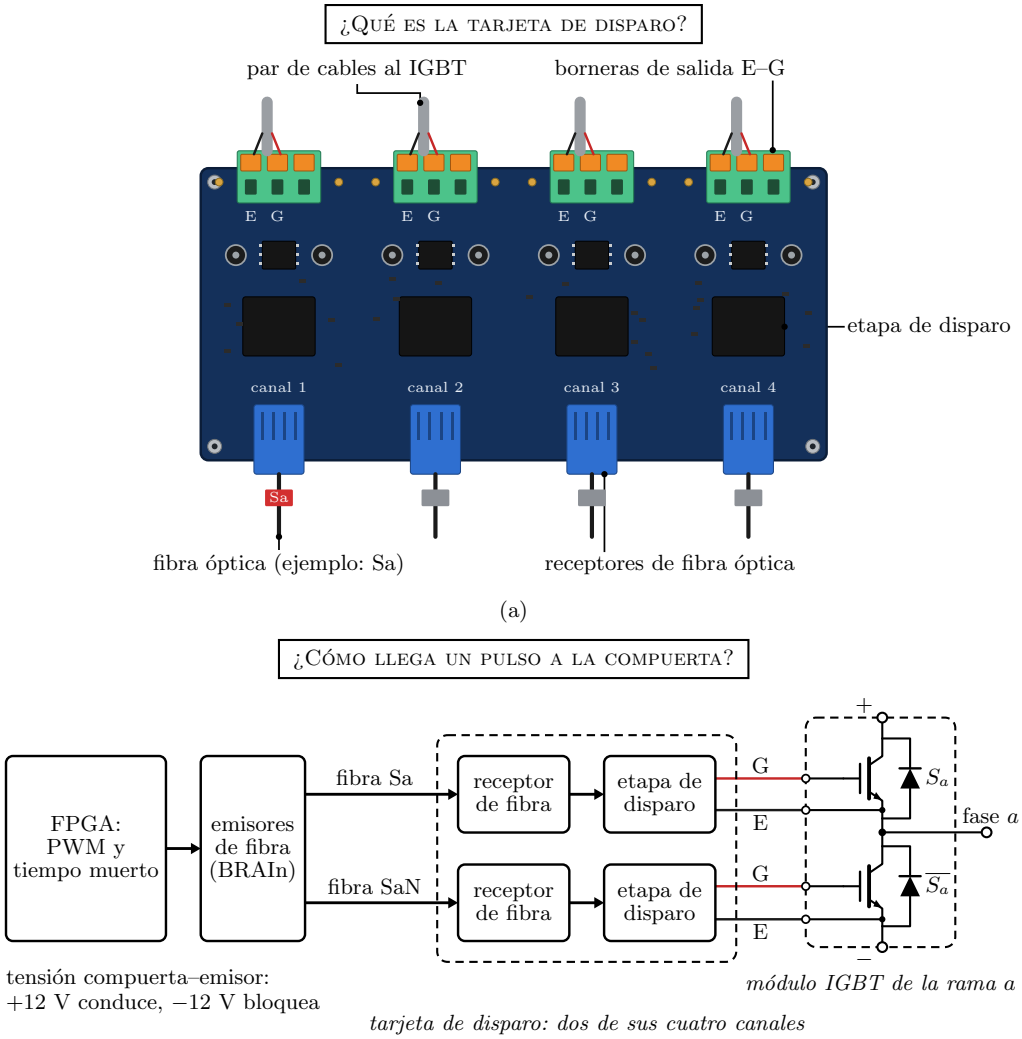
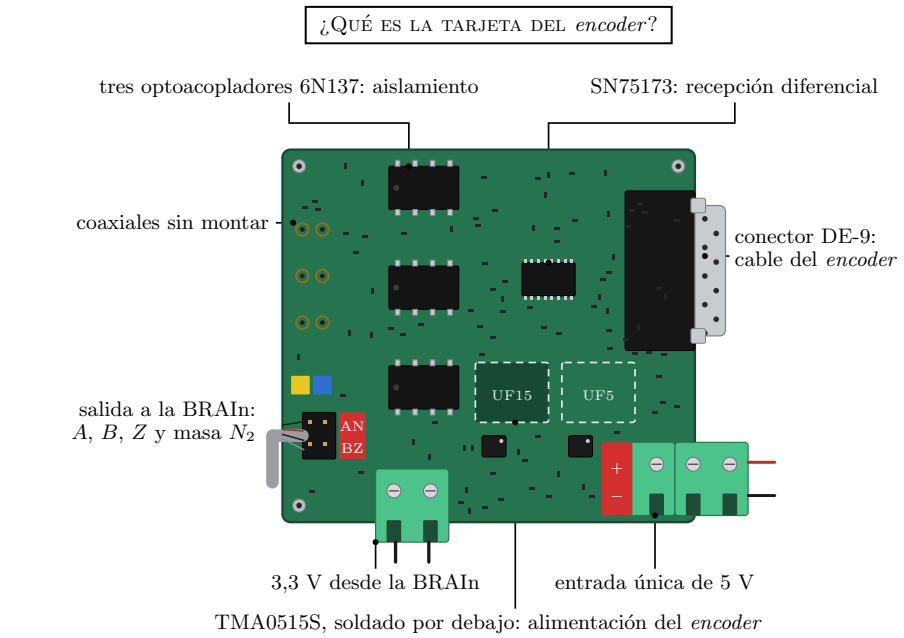
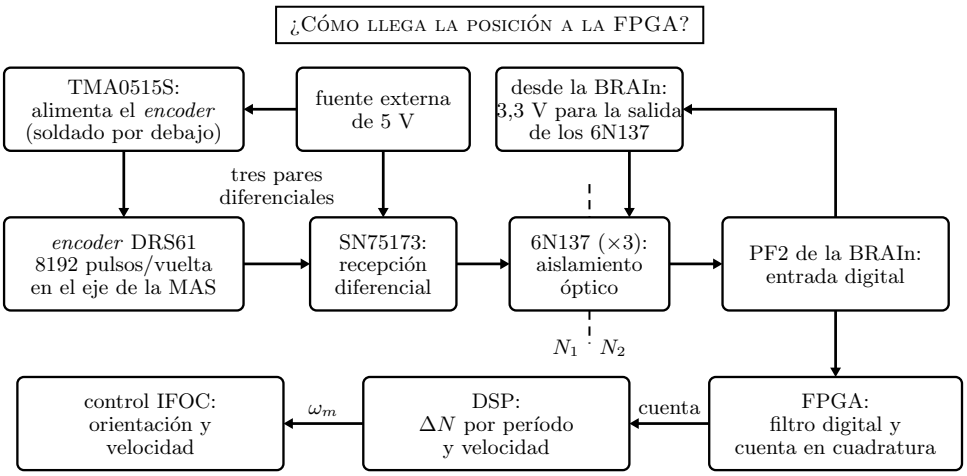


Figura A.3: Tarjeta de disparo y recorrido de las órdenes de la pierna a. Las señales complementarias S_a y $\overline{S_a}$ llegan desde la unidad de control por fibras separadas; cada canal controla la compuerta y el emisor auxiliar del IGBT correspondiente. Elaboración propia a partir de la tarjeta utilizada en el banco.

La posición se adquiere por una interfaz distinta. El *encoder* entrega las señales diferenciales A, B y Z. La tarjeta acondicionadora las convierte y las aísla antes de llevarlas a la FPGA. Allí se filtran y se cuentan en cuadratura, y el DSP obtiene el incremento de posición y la velocidad que utiliza el IFOC. La Figura A.4 describe esta cadena sin asignar números de pin que no quedaron verificados en el montaje.



(a)



El decodificador cuenta 32 768 por vuelta mecánica: 8192 pulsos y cuatro flancos por pulso.
La señal Z se recibe, pero está configurada sin efecto sobre la cuenta.
Los optoacopladores separan las dos masas: N_1, la del *encoder*, y N_2, la de la BRAIn.

(b)

Figura A.4: Tarjeta acondicionadora y recorrido de la posición desde el *encoder* DRS61 hasta el IFOC. El receptor diferencial y los optoacopladores adaptan las señales A, B y Z; la separación entre N_1 y N_2 representa el aislamiento entre el lado del sensor y la unidad de control. Elaboración propia a partir del esquemático y de la tarjeta instalada en el banco.

A.2. Parámetros de la máquina y bases de normalización

Los valores eléctricos y mecánicos provienen del modelo validado de Gálvez [22]. La Tabla A.1 los reúne junto con los datos de placa y las bases en por unidad.

Tabla A.1: Máquina de inducción del banco: placa, bases de normalización y parámetros del modelo [22]. Las derivadas salen de las anteriores: $I_b = \sqrt{2}\, I_n$, $V_b = \sqrt{2}\, V_{n,LL}/\sqrt{3}$, $Z_b = V_b/I_b$, $S_b = \frac{3}{2} V_b I_b$, $\omega_b = 2\pi f_n$, $\omega_{b,m} = \omega_b/p$, $T_b = S_b\, p/\omega_b$, $k_r = L_m/L'_r$, $R_\sigma = R_s + k_r^2 R'_r$, y σ y τ_r por (2.1) y (2.4).

Placa		
Tensión nominal l–l	$V_{n,LL}$	400 V
Corriente nominal	I_n	11,03 A
Potencia nominal	P_n	5,5 kW
Frecuencia nominal	f_n	50 Hz
Velocidad nominal	n_n	1450 rpm
Pares de polos	p	2
Bases de normalización		
Corriente, de cresta	I_b	15,6 A
Tensión, cresta f–n	V_b	326,6 V
Impedancia	Z_b	20,94 Ω
Potencia	S_b	7,64 kVA
Velocidad eléctrica	ω_b	314,2 rad/s
Velocidad mecánica	$\omega_{b,m}$	157,1 rad/s
Torque	T_b	48,65 N m

Modelo		
Resistencia de estator	R_s	1,02 Ω
Resistencia de rotor	R'_r	0,829 Ω
Ind. de magnetización	L_m	153,615 mH
Inductancias propias	$L_s = L'_r$	160,321 mH
Inercia del modelo	J	0,06 kg m^2
Acoplamiento de rotor	k_r	0,9582
Dispersión total	σ	0,0819
Inductancia transitoria	σL_s	13,131 mH
Resistencia equivalente	R_σ	1,781 Ω
Constante de rotor	τ_r	193,39 ms

La corriente de placa es $I_n = 11{,}03$ A RMS; la base instantánea de corrientes y observaciones es $I_b = \sqrt{2} I_n = 15{,}6$ A. El torque base, $T_b = 48{,}65$ N m, tampoco es el nominal: $T_n = 5500/151{,}84 \simeq 36{,}2$ N m, o 0,745 pu. Para expresar un torque en pu de T_b como múltiplo de T_n se multiplica por 1,342.

Con el flujo estimado en pu,

$$T_e = \frac{3}{2} p \frac{L_m^2}{L'_r} i_{mr} i_{sq} \simeq (2{,}989 \text{ N m/A})\, \psi'_{r,\text{pu}} i_{sq}. \tag{A.1}$$

El factor respecto de i_{sq} es constante con flujo en 1 pu; durante la magnetización se conserva el producto con el flujo estimado.

A.3. Configuración del entrenamiento

La Tabla A.2 reúne las tres etapas de entrenamiento. Se utilizan distintos valores de R y dos valores de L, 10 y 20 mH, para cubrir cargas con diferentes ganancias y constantes de tiempo; la inductancia no se barre continuamente entre ellos. El Agente 2 refinado procede del linaje del Agente 2. Su etapa final continúa el aprendizaje desde un punto de control de refinamiento previo, manteniendo la dinámica $R{-}L$ y la perturbación en q, con las condiciones de la tabla. Los intervalos de cada etapa se conservan por separado. La frecuencia de decisión es 3 kHz; el banco utiliza una portadora triangular de 3 kHz para la PWM con inyección Min–Max.

Cada episodio contiene 1500 decisiones, esto es 0,5 s, y al cerrarlo se ejecutan 375 actualizaciones. Las semillas son 20260731 para el entrenamiento y 777 para la evaluación. La procedencia de los puntos de control del refinamiento se documenta en el Apéndice C. Los pesos se actualizan durante el refinamiento y quedan fijos al terminar. En el linaje, el cambio del peso de movimiento de 0,30 a 1,50 requiere readaptar los valores Q; los cambios de normalización requieren transformar la entrada correspondiente.

Tabla A.2: Configuración con que se entrenó cada agente, verificada contra los ficheros de configuración y los registros de entrenamiento de cada etapa.

	Agente 1	Agente 2	Agente 2 refinado
Común a los tres			
Observación de 15 componentes · red 15–128–128–225 con ReLU ocultas y salida lineal · acción incremental $(\Delta m_d, \Delta m_q)$ de 225 combinaciones · pérdida de Huber con umbral 1 · memoria de $2 \cdot 10^5$ transiciones con muestreo uniforme · lote de 256 · $\gamma = 0{,}99$ · red objetivo con actualización suave $\tau_\theta = 10^{-3} \cdot T_s = 333{,}33\ \mu$s			
Dominio de entrenamiento			
Resistencia R	12 a 18 Ω	0,5 a 3,5 Ω	0,5 a 3,5 Ω
Inductancia L	10 y 20 mH	10 y 20 mH	10 y 20 mH
DC link V_{dc}	100 a 210 V	50 a 320 V	50 a 300 V
Corriente máxima	8 A	7,5 A	17 A
ω_{pu}	1,0 fija	1,0 fija	−0,45 a 0,45, sorteada
Perturbación en q	no	no	sí, hasta $\bar{p} = 0{,}85$
Ramas reservadas	no	no	sostenimiento 35 % · banco 7,8 % (12 % del resto)
Programa de entrenamiento			
Inicialización	desde cero	pesos heredados	refinamiento previo del linaje del Agente 2
$V_{dc,\text{ref}}$	210 V	320 V	320 V
Tasa de aprendizaje	$3 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$5 \cdot 10^{-5}$
Exploración ε	$1{,}00 \rightarrow 0{,}05$	$0{,}30 \rightarrow 0{,}05$	$0{,}10 \rightarrow 0{,}03$
Peso de movimiento c_m	0,30	0,30	1,50
Episodios	2000	2000	817 de 1200
Evaluación *greedy*	cada 25	cada 25	cada 25
Pasos de calentamiento	5000	5000	5000

El refinamiento final se previó para 1200 episodios y se detuvo de forma manual en el episodio 817, con la señal de parada del programa, porque el costo de evaluación había dejado de mejorar. Al detenerse, ε estaba en su valor mínimo desde el episodio 660 y las cinco evaluaciones posteriores a la del episodio 675 no mejoraron su costo. La mediana de la pérdida de Huber había bajado de 0,49 en los primeros 100 episodios a 0,32 en los dos últimos bloques de 100. Los episodios 801 a 817 no llegaron a evaluarse.

El Algoritmo A.1 reúne el procedimiento ejecutado en cada etapa. La Tabla C.3 relaciona sus líneas con el programa entregado.

Algoritmo A.1: Procedimiento de entrenamiento de una etapa, en el orden en que lo ejecuta el programa. Los valores de cada etapa están en la Tabla A.2. La correspondencia de cada línea con el programa está en la Tabla C.3.

Inicialización

1 Crear la red en línea $Q_{\boldsymbol{\theta}}$, 15–128–128–225, con pesos de Glorot y sesgos nulos.

2 Si la etapa continúa otra, copiar todos los parámetros de la red en línea de origen y multiplicar la columna de la entrada 15 de la primera capa por $V_{dc,\mathrm{ref}}/V_{dc,\mathrm{ref},0}$. La red objetivo, la memoria y los momentos de Adam no se heredan.

3 $\boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$; memoria vacía de $2 \cdot 10^5$ transiciones; momentos de Adam nulos y su contador de actualizaciones en cero; $\mathcal{C}_{\mathrm{mejor}} \leftarrow \infty$.

4 Semilla 20260731 para el entrenamiento y 777 para la evaluación.

Para cada episodio $e = 1, \ldots, N_{\mathrm{ep}}$*, salvo que se active la señal de parada*

5 $\varepsilon_e = \varepsilon_0(\varepsilon_1/\varepsilon_0)^{(e-1)/(n_\varepsilon-1)}$ si $e \le n_\varepsilon$; si no, $\varepsilon_e = \varepsilon_1$. El valor n_ε es el 60 % de N_{ep} en los Agentes 1 y 2 y el 55 % en el Agente 2 refinado.

6 Sortear el tipo de perfil, R, L, V_{dc}, ω_{pu}, p_q y las referencias según la Tabla A.5.

7 Reiniciar $i_d = i_q = 0$ y $m_d = m_q = 0$, de modo que la tensión pendiente por el retardo es nula. Los filtros de los errores absolutos, los errores previos y t_{ref} empiezan en cero.

Para cada período $k = 1, \ldots, 1500$

8 Medir las corrientes con ruido y cuantización según (A.2).

9 Calcular $e_{d,k}$ y $e_{q,k}$, sus diferencias con el período anterior, nulas en $k = 1$, sus filtros $\bar{e}_{d,k}$ y $\bar{e}_{q,k}$ y $t_{\mathrm{ref},k}$, que vuelve a cero en cada cambio de referencia. Formar $\boldsymbol{o}_k$ con las modulantes de $k-1$.

10 Con probabilidad ε_e, elegir a_k al azar entre las 225 acciones; si no, $a_k = \arg\max_a Q_{\boldsymbol{\theta}}(\boldsymbol{o}_k, a)$.

11 Acumular el par de incrementos y recortar cada componente según (3.14). Durante el período se aplica la modulante de $k-1$.

12 Evolucionar la carga con (3.4) y (3.6), con el tiempo muerto calculado con la corriente de la carga y p_q en el eje q.

13 Calcular r_k con (3.21) y el error medido en la línea 9, antes de que actúe a_k.

14 Guardar $(\boldsymbol{o}_k, a_k, r_k, \boldsymbol{o}_{k+1}, d_k)$, con $d_k = 1$ sólo en $k = 1500$.

Al cerrar el episodio

15 Agregar las 1500 transiciones a la memoria; si está llena, reemplazan a las más antiguas.

Si la memoria tiene al menos 5000 *transiciones, repetir* 375 *veces*

16 Muestrear 256 transiciones al azar, con reposición.

17 $a'_i = \arg\max_a Q_{\boldsymbol{\theta}}(\boldsymbol{o}'_i, a)$ e $y_i = r_i + \gamma(1-d_i)\, Q_{\boldsymbol{\theta}^-}(\boldsymbol{o}'_i, a'_i)$.

18 Actualizar $\boldsymbol{\theta}$ con Adam, con $\beta_1 = 0{,}9$, $\beta_2 = 0{,}999$ y constante de estabilidad 10^{-8}, sobre la pérdida de Huber de umbral 1 promediada en el lote. El gradiente no se recorta.

19 $\boldsymbol{\theta}^- \leftarrow (1-\tau_\theta)\boldsymbol{\theta}^- + \tau_\theta \boldsymbol{\theta}$.

Cada 25 *episodios*

20 Reiniciar la semilla de evaluación y simular los ocho escenarios con $a_k = \arg\max_a Q_{\boldsymbol{\theta}}(\boldsymbol{o}_k, a)$, sin exploración y sin guardar transiciones. Calcular $\mathcal{C}$ con (3.22).

21 Guardar el punto de control. Si $\mathcal{C} < \mathcal{C}_{\mathrm{mejor}}$, registrarlo como el mejor.

La Tabla A.3 define cada entrada de la observación en el entrenamiento, en la simulación y en el banco. Con la máquina, simulada o en el banco, las referencias provienen de los PI externos y cambian en cada muestra. Por eso, en la simulación y en el banco, t_{ref} vuelve a cero sólo cuando la variación filtrada de la referencia alcanza 0,05 pu. Los escalones de la carga R–L alcanzan ese umbral. Tras cada uno, el banco repite la rampa de 20 ms del entrenamiento y el modelo de la carga en PLECS mantiene t_{ref} en cero durante 100 ms. El efecto de esa diferencia no se aisló. En la simulación de la máquina no hubo reinicios. En el banco con la máquina esta entrada no se registró, por lo que los archivos no permiten determinar si el reloj se reinició durante el arranque o la inversión.

Tabla A.3: Definición de las quince entradas de la observación, con la numeración de la Figura 6.2. Simulación: los modelos en PLECS de los Capítulos 4 y 6. Banco: el programa de los Capítulos 4 y 7. Corrientes en pu de $I_b = 15{,}6$ A.

N.º	Definición	Entrenamiento	Simulación y banco
1, 2	$e_x(k) = i_x^*(k) - i_x(k)$, en cada eje	con la corriente medida de (A.2)	Simulación: la corriente del modelo. Banco: la medida, llevada a dq
3, 4	$\Delta e_x(k) = e_x(k) - e_x(k-1)$	nula en la primera muestra	el error previo parte de cero y se guarda sin reflejar
5, 6	i_d^*, i_q^*	las del escenario	Carga $R-L$: los escalones del ensayo. Máquina: las de los PI de flujo y velocidad
7	ω_{pu}	frecuencia del marco dq de la carga, fija en el episodio	Carga $R-L$, simulación y banco: 1. Simulación de la máquina: $\omega_s/\omega_b = (\omega + \omega_{sl})/\omega_b$, frecuencia del marco sincrónico, con la velocidad eléctrica del rotor ω del modelo mecánico, sin filtrar, y el deslizamiento ω_{sl} estimado. Banco con la máquina: la velocidad eléctrica del rotor medida por el *encoder*, filtrada a 80 Hz y sin el factor 1,0233, más el deslizamiento estimado, dividida por ω_b, según la Tabla B.4
8, 9	$m_d(k-1)$, $m_q(k-1)$, después de su límite, en la base $V_N/2$	límite $\pm m_{\text{lím}}$ por eje	límites de la Tabla A.12
10, 11	$\sin\theta_M$, $\cos\theta_M$, con $\theta_M = \text{atan2}(m_q(k-1), m_d(k-1))$	$\theta_M = 0$ si las dos modulantes son nulas	conserva el ángulo anterior si las dos son nulas; parte de cero
12, 13	$\bar{e}_x(k) = 0{,}98\,\bar{e}_x(k-1) + 0{,}02\,\lvert e_x(k)\rvert$	parte de cero	igual
14	$\tau_{\text{ref}} = \text{mín}(t_{\text{ref}}/20\text{ ms}, 1)$	$t_{\text{ref}} = 0$ con cualquier cambio de referencia; parte de cero	Simulación y banco: $d(k) = \sqrt{\Delta i_d^*(k)^2 + \Delta i_q^*(k)^2}$, con $\Delta i_x^*(k) = i_x^*(k) - i_x^*(k-1)$, y $f(k) = 0{,}6\,f(k-1) + 0{,}4\,d(k)$. $t_{\text{ref}} = 0$ si $f(k) \geq 0{,}05$ pu. En la primera muestra, $f(0) = 0$ y $\tau_{\text{ref}}(0) = 1$. Simulación: t_{ref} sigue en cero mientras $f(k) \geq 0{,}05$ pu y 100 ms más; en la máquina no hubo reinicios. Banco: tras un reinicio, t_{ref} avanza y se bloquea otro durante los 20 ms de la rampa; con la máquina, esta entrada no se registró
15	$V_{dc}/V_{dc,\text{ref}}$, con $V_{dc,\text{ref}} = 320$ V	V_{dc} del episodio	Simulación: 300 V. Banco: V_{dc} medido y filtrado

A.3.1. Constantes del entorno

La Tabla A.4 reúne las constantes del entorno en el mismo orden en que el Capítulo 3 enuncia cada bloque.

Tabla A.4: Constantes con que se ejecutó el entorno del Agente 2 refinado. Los símbolos son los de (A.5), (3.13) y (3.21).

Planta discreta y cadena de medida		
Límite por componente y umbral radial	m_{lim}	0,077 a 0,462
Normalización de la entrada del *DC link*	$V_{dc,\text{ref}}$	320 V
Desviación del ruido de corriente	σ_i	15,0 mA
Paso de cuantización de corriente	q_i	24,786 mA
Tiempo muerto modelado	T_d	1,0 μs
Coeficiente de las medias exponenciales		0,02
Perturbación de tensión del eje de cuadratura		
Coeficiente antes del límite	$K\xi$	$\mathcal{U}[1{,}627; 1{,}989]$
Techo de la perturbación	$\bar{p}$	0,85
Escalera de incrementos		
Niveles positivos por eje	n	7
Incremento menor	$\Delta m_{\text{mín}}$	$7{,}830 \cdot 10^{-5}$
Razón de la progresión	g_m	2,6254
Incremento mayor	$\Delta m_{\text{máx}}$	$2{,}564 \cdot 10^{-2}$
Pesos de la recompensa		
Error del eje directo	c_d	6
Error del eje de cuadratura	c_q	4
Movimiento de la modulante	c_m	1,5
Exceso sobre el límite	c_s	2
Ventana sin castigo al movimiento	τ_Δ	20 ms

Ruido y cuantización de las corrientes. La caracterización inicial de los canales de corriente de la unidad de control se realizó antes del refinamiento final. A partir de esas mediciones se adoptó una desviación de ruido de $\sigma_i = 15{,}0$ mA, mientras que la calibración del canal estableció un paso de $q_i = 24{,}786$ mA por cuenta. En el entorno, ambos efectos se aplican en cada eje según

$$\widetilde{i}^{\text{pu}}_{x,k} = q_i^{\text{pu}} \operatorname{round}\left(\frac{i^{\text{pu}}_{x,k} + n_{x,k}}{q_i^{\text{pu}}}\right), \qquad n_{x,k} \sim \mathcal{N}\big(0, (\sigma_i^{\text{pu}})^2\big), \qquad x \in \{d, q\}, \tag{A.2}$$

donde $\sigma_i^{\text{pu}} = 0{,}00096$ y $q_i^{\text{pu}} = 24{,}786\ \text{mA}/15{,}6\ \text{A} = 0{,}001589$. Primero se añade el ruido y luego se redondea al nivel más próximo. La observación y la recompensa se construyen con $\widetilde{i}_{d,k}$ y $\widetilde{i}_{q,k}$; la carga, en cambio, evoluciona con sus estados internos sin añadirles estos errores de medida. Esta representación aproxima la cadena de medida directamente en dq y no reproduce por separado la cuantización de cada corriente de fase.

La planta se actualiza una vez por período mediante la tensión promedio en dq. Por ello, el entorno no conmuta individualmente los seis semiconductores ni genera el *ripple* asociado a los pulsos PWM. El efecto del tiempo muerto se incorpora mediante la componente fundamental aproximada deducida en el apartado siguiente.

Deducción del término de tiempo muerto. Se parte del error promedio por pierna con PWM [22]. Para corrientes balanceadas $i_a = I\cos\theta$, $i_b = I\cos(\theta - 2\pi/3)$ e $i_c = I\cos(\theta +$

$2\pi/3$), el cálculo se resume en cuatro pasos:

$$\begin{aligned}
\Delta v_{x,\mathrm{dt}} &= -K_0\,\mathrm{sgn}(i_x), & K_0 &= \frac{T_d + t_{\mathrm{on}} - t_{\mathrm{off}}}{T_{\mathrm{sw}}} V_{dc},\\
b_1 &= \frac{1}{\pi}\int_{-\pi}^{\pi} |\cos\theta|\,\mathrm{d}\theta = \frac{4}{\pi},\\
\Delta \boldsymbol{v}^{(1)}_{s,\mathrm{dt}} &= \frac{2}{3}\left(\Delta v^{(1)}_{a,\mathrm{dt}} + \underline{a}\,\Delta v^{(1)}_{b,\mathrm{dt}} + \underline{a}^2 \Delta v^{(1)}_{c,\mathrm{dt}}\right) = -\frac{4}{\pi} K_0 e^{j\theta},\\
\boldsymbol{v}^{(1)}_{\mathrm{dt}} &= -\frac{4}{\pi}\frac{T_d}{T_{\mathrm{sw}}} V_{dc} \frac{\boldsymbol{i}}{\|\boldsymbol{i}\|}.
\end{aligned} \tag{A.3}$$

La transformación espacial conduce a la expresión publicada en [23, ec. (3.34)]. El entorno supone $t_{\mathrm{on}} = t_{\mathrm{off}}$, suaviza el cruce por cero con ρ de (3.7) y no representa las caídas de conducción. Con $T_d = 1$ μs, $T_{\mathrm{sw}} = 1/3000$ s y $V_{dc} = 285$ V, la magnitud es 1,09 V.

La entrada temporal se normaliza con $\tau_\Delta = 20$ ms: $\tau_{\mathrm{ref}} = \mathrm{mín}(t_{\mathrm{ref}}/\tau_\Delta, 1)$. La condición de (3.19) compara el tiempo físico t_{ref} con esa ventana; una entrada normalizada menor que uno indica que aún no transcurre.

Los siete niveles positivos que produce (3.13), redondeados a cuatro cifras, son $7{,}830 \cdot 10^{-5}$, $2{,}056 \cdot 10^{-4}$, $5{,}397 \cdot 10^{-4}$, $1{,}417 \cdot 10^{-3}$, $3{,}721 \cdot 10^{-3}$, $9{,}768 \cdot 10^{-3}$ y $2{,}564 \cdot 10^{-2}$. Cada eje añade el cero y los siete simétricos, de donde salen las 15×15 combinaciones. La coordenada d varía más rápido que q: con $j_d, j_q \in \{0, \ldots, 14\}$, el índice de salida es $a_{\mathrm{C}} = 15 j_q + j_d$ en C y $a_{\mathrm{C}} + 1$ en MATLAB, y la acción nula ocupa el índice 112 y la fila 113.

Las dos ramas reservadas imponen además sus propias condiciones. En la de sostenimiento se fijan $\omega_{\mathrm{pu}} = 0$, $p_q = 0$ y $V_{dc} \in [250; 300]$ V. La referencia del eje de cuadratura permanece en cero. Después de un refuerzo transitorio de la referencia del eje directo, ésta se mantiene en 0,434 pu. Ese valor se dimensionó con información nominal de la máquina para configurar el escenario; no es una observación del agente y no incorpora el modelo dinámico de la máquina. La del banco fija $\omega_{\mathrm{pu}} = 1$, esto es los 50 Hz de los ensayos, y $p_q = 0$. El 57,2 % restante de los episodios queda para el muestreo general, que sortea $\omega_{\mathrm{pu}} \sim \mathcal{U}[-0{,}45; 0{,}45]$.

La Tabla A.5 reúne el sorteo de cada escenario de la etapa final.

El entorno limita las referencias sorteadas mediante una cota que considera la tensión disponible y la perturbación:

$$V_{\mathrm{util}} = \mathrm{máx}\left(0{,}10\frac{V_{dc}}{2}, \frac{V_{dc}}{2} - |p_q|\right), \qquad i_{\mathrm{máx},F} = 0{,}92\,\mathrm{mín}\left(17\ \mathrm{A}, \frac{(1 - 2k_{\mathrm{dt}})V_{\mathrm{util}}}{Z}\right), \tag{A.4}$$

con $Z = \sqrt{R^2 + (\omega L)^2}$ y $k_{\mathrm{dt}} = (4/\pi)(T_d/T_{\mathrm{sw}})$ y $T_{\mathrm{sw}} = 1/3000$ s. El factor 0,92 es el margen de tensión del 8 % y los 17 A, el techo absoluto de corriente.

El incremento menor produce un cambio de corriente de 8 mA sobre la impedancia de diseño, 3,181 Ω; el mayor equivale a un tercio de la modulante disponible con el *DC link* más bajo. Fijados ambos extremos, g_m queda determinado.

Tabla A.5: Sorteo de cada escenario en la etapa del Agente 2 refinado. $\mathcal{U}[a;b]$ es una distribución uniforme e $i_{\text{máx},F}$ es la cota de (A.4). Los Agentes 1 y 2 usan el mismo reparto de perfiles, con los valores de la Tabla A.2, ω_{pu} fija y sin perturbación en q ni ramas reservadas.

Variable	Sorteo
Rama del episodio	sostenimiento, 35 %; banco, 12 % del resto; muestreo general, 57,2 %
Tipo de perfil	escalones, 55 %; mixto, 30 %; rampa, 15 %. La rama de sostenimiento lo reemplaza
R	$\mathcal{U}[0{,}5; 3{,}5]$ Ω
L	10 o 20 mH, con probabilidad 1/2 cada una
V_{dc}	$\mathcal{U}[50; 300]$ V; en el sostenimiento, $\mathcal{U}[250; 300]$ V
ω_{pu}	$\mathcal{U}[-0{,}45; 0{,}45]$ en el muestreo general; 1 en la rama del banco; 0 en el sostenimiento
p_q	según (A.6), con $\xi \sim \mathcal{U}[0{,}90; 1{,}10]$; nula en las dos ramas reservadas
Número de tramos	4 a 7, equiprobables y de igual duración
Escalones	con probabilidad 1/2 se mueve el eje directo. El eje que se mueve toma en cada tramo $\mathcal{U}[0{,}05; 0{,}90]\, i_{\text{máx},F}$; el otro queda en $\mathcal{U}[0{,}15; 0{,}65]\, i_{\text{máx},F}$ todo el episodio. La referencia de q lleva un signo sorteado en cada tramo
Mixto	en cada tramo, $i_d^* \sim \mathcal{U}[0{,}05; 0{,}95]\, i_{\text{máx},F}$ e $i_q^* \sim \mathcal{U}[0; 0{,}90]\, i_{\text{máx},F}$, con signo sorteado
Rampa	interpolación lineal entre valores sorteados como en el perfil mixto, en los bordes de los tramos
Sostenimiento	$i_q^* = 0$; $i_d^* = 0$ en el primer 5 %, 0,868 pu hasta el 20 % y 0,434 pu después, sin superar $i_{\text{máx},F}$
Recorte final	si $\lvert \boldsymbol{i}^* \rvert > i_{\text{máx},F}$, las dos referencias se escalan por igual

Cálculo de la perturbación del eje q. En cada episodio del muestreo general se sortea $\xi \sim \mathcal{U}[0{,}90; 1{,}10]$, que introduce una variación de ±10 % antes de limitar la amplitud:

$$\begin{aligned} \alpha_0 &= -K\omega_{\text{pu}}\xi, \\ a_{\text{máx}} &= \min(\bar{p}, K|\omega_{\text{pu}}|), \\ \alpha &= \begin{cases} -a_{\text{máx}}, & \alpha_0 < -a_{\text{máx}}, \\ \alpha_0, & -a_{\text{máx}} \le \alpha_0 \le a_{\text{máx}}, \\ a_{\text{máx}}, & \alpha_0 > a_{\text{máx}}, \end{cases} \\ p_q &= \alpha \frac{V_{dc}}{\sqrt{3}}. \end{aligned} \tag{A.5}$$

Aquí ω_{pu} es la frecuencia normalizada del marco, K fija la pendiente y $\bar{p}$ limita la amplitud antes de convertirla en la tensión p_q.

El agente seleccionado conserva $K = 1{,}808$, $\bar{p} = 0{,}85$ y $|\omega_{\text{pu}}| \le 0{,}45$. Como $K|\omega_{\text{pu}}| \le 0{,}8136 < \bar{p}$, en el muestreo general se obtiene

$$p_q = -K\omega_{\text{pu}} \min(\xi, 1) \frac{V_{dc}}{\sqrt{3}}. \tag{A.6}$$

Así, la amplitud final queda entre el 90 y el 100 % de la nominal.

El valor de K no se obtuvo durante el aprendizaje. Los datos nominales de la Tabla A.1 y la condición de magnetización de 0,434 pu entregan una tensión inducida en el eje de cuadratura de 313,1 V por unidad de velocidad eléctrica del rotor. Al expresarla respecto de la base $V_{dc}/\sqrt{3}$, con $V_{dc} = 300$ V, se obtiene

$$K = \frac{313{,}1}{300/\sqrt{3}} = 1{,}808. \tag{A.7}$$

La identificación realizada sobre el modelo de la máquina en PLECS entregó $K = 1{,}810$, con una correlación de 0,9998. Esta comprobación respalda la escala utilizada. La relación física depende de la velocidad eléctrica del rotor; el entorno representa esa dependencia mediante la frecuencia del marco utilizada para construir p_q. Como la perturbación se escala con el *DC link* de cada episodio, la correspondencia numérica con la tensión inducida se establece para 300 V. En los demás casos, p_q sigue siendo una perturbación externa acotada y no un modelo de la máquina.

Esta elección utiliza información previa de la exigencia de tensión de la aplicación únicamente para dimensionar una perturbación externa; no incorpora al entorno las ecuaciones del flujo de rotor ni la dinámica mecánica de la máquina. Tampoco constituye un parámetro físico de la carga ni una variable observada por el agente. La planta de entrenamiento siguió siendo la carga $R-L$: la perturbación no modificó R, L, las matrices ni los estados. En otra aplicación debe recalcularse o verificarse esta escala. La Tabla A.6 estima la tensión requerida a flujo constante con un presupuesto conservador, $V_{dc}/2$ de amplitud de fase, y no demuestra que el control funcione a esas velocidades. La PWM con inyección Min–Max admite $V_{dc}/\sqrt{3}$ de fase, o $V_{dc}/\sqrt{2}$ RMS entre fases, y los 400 V nominales exigen un *DC link* ideal de $400\sqrt{2} = 565{,}7$ V. La base de normalización $V_N = 650$ V se respalda en los 400 V RMS nominales de la máquina y queda un 14,9 % por encima del valor mínimo ideal del *DC link* con Min–Max. Esa base no es el *DC link* medido en los ensayos, entre 281 y 285 V: con ella el agente opera también a baja tensión, porque $m_{\text{lím}}$ reduce su rango con V_{dc}, y se entrenó entre 50 y 300 V.

Tabla A.6: Tensión en bornes y *DC link* estimado para cada velocidad, a flujo constante y con una amplitud de fase de $V_{dc}/2$. A 0,30 pu, el banco entregó 120,9 V RMS entre fases, un 0,8 % por encima de los 120 V estimados. La proporcionalidad desprecia la caída resistiva, $R_s I_n = 11{,}3$ V RMS por fase.

ω_{pu}	V_{LL} en bornes	V_{dc} estimado	
0,30	120 V	196 V	tramo del banco
0,45	180 V	294 V	extremo del muestreo general
0,50	200 V	327 V	
0,75	300 V	490 V	
1,00	400 V	653 V	velocidad nominal

Con 300 V, esta estimación da aproximadamente 0,46 pu, cerca del extremo de 0,45 pu del muestreo general. Los ensayos sólo validan ±0,30 pu. Aumentar el *DC link* aporta tensión disponible, pero ampliar el rango de operación exige nuevos ensayos y, si se amplía el dominio de entrenamiento, más exploración y episodios.

Los ocho escenarios de selección del punto de control se generan con la misma semilla al comienzo de cada evaluación, de modo que todas las evaluaciones de una etapa usan los mismos casos, incluido el ruido de medida. Cada escenario dura 0,5 s y fija R, L, V_{dc}, ω_{pu}, p_q y las referencias. Su término $\mathcal{C}_j$ en (3.22) se calcula en las muestras 900 a 1500, de 0,30 a 0,50 s, con las corrientes de la carga, sin el ruido ni la cuantización que recibe la observación. $\mathcal{C}$ es el promedio de los ocho $\mathcal{C}_j$. Las Tablas A.7, A.8 y A.9 describen los escenarios de la etapa final, sus referencias en cada tramo y el resultado del punto de control seleccionado.

En el episodio 675, $\mathcal{C} = 2{,}74$ A es ese promedio de ocho criterios y no un error de corriente. De ese valor, $\mathcal{C}^* = 2{,}27$ A corresponde al seguimiento perfecto, porque en seis escenarios la referencia cambia dentro de la ventana. El resto, $\Delta\mathcal{C} = 0{,}47$ A, es lo que agrega el agente, con 0,41 A de sesgo y 0,06 A de desviación estándar adicional. Como estos escenarios se usaron en cada evaluación para escoger los pesos, sus resultados no constituyen una evaluación independiente de generalización. Esa evaluación corresponde a los Capítulos 4, 6 y 7.

Tabla A.7: Escenarios de selección del punto de control en la etapa final. Los genera el sorteo de la Tabla A.5 con la semilla 777, reiniciada antes de cada evaluación. Cada uno dura 0,5 s. La Tabla A.8 da sus referencias en cada tramo.

Esc.	Tipo	R [Ω]	L [mH]	V_{dc} [V]	ω_{pu}	p_q [V]	Tramos	$i_{\text{máx},F}$ [A]
1	mixto	1,23	20	65,9	−0,427	29,4	5	1,10
2	escalones	1,21	10	163,6	0,050	−8,6	7	15,64
3	sostenimiento	1,55	10	266,5	0,000	0,0	3	15,64
4	escalones	0,90	10	264,0	−0,040	9,9	5	15,64
5	mixto	3,20	10	132,9	0,314	−43,6	6	6,23
6	escalones	1,83	20	206,7	−0,049	9,7	5	15,64
7	mixto	1,07	10	123,6	0,237	−30,6	4	15,64
8	sostenimiento	2,40	20	296,7	0,000	0,0	3	15,64

Tabla A.8: Referencias de corriente de los escenarios de selección de la Tabla A.7, en amperios. En cada tramo, i_d^* e i_q^* son constantes. El tiempo se mide desde el inicio del escenario. En los de sostenimiento, 13,54 y 6,77 A son 0,868 y 0,434 pu.

Esc.	Tramo [ms]	i_d^* [A]	i_q^* [A]
1	0 a 100	0,95	0,55
	100 a 200	1,00	0,03
	200 a 300	0,67	−0,88
	300 a 400	0,41	−0,18
	400 a 500	0,43	0,03
2	0 a 71,3	4,70	8,11
	71,3 a 143	4,84	8,11
	143 a 214,3	5,16	−8,11
	214,3 a 285,7	3,11	8,11
	285,7 a 357	7,41	−8,11
	357 a 428,7	8,36	8,11
	428,7 a 500	5,18	8,11
3	0 a 25	0,00	0,00
	25 a 100	13,54	0,00
	100 a 500	6,77	0,00
4	0 a 100	2,63	−6,39
	100 a 200	10,48	−6,39
	200 a 300	11,78	−6,39
	300 a 400	13,85	−6,39
	400 a 500	13,12	−6,39
5	0 a 83,3	2,07	−2,47
	83,3 a 166,7	5,46	−1,42
	166,7 a 250	1,91	−3,23
	250 a 333,3	3,23	0,55
	333,3 a 416,7	5,18	0,57
	416,7 a 500	4,05	4,74
6	0 a 100	13,07	−8,58
	100 a 200	1,32	8,75
	200 a 300	13,00	−8,69
	300 a 400	2,98	−8,75
	400 a 500	9,82	8,75
7	0 a 125	4,95	−4,40
	125 a 250	9,03	9,72
	250 a 375	0,98	7,29
	375 a 500	11,73	−6,97
8	0 a 25	0,00	0,00
	25 a 100	13,54	0,00
	100 a 500	6,77	0,00

Tabla A.9: Resultado del punto de control seleccionado, episodio 675, en los escenarios de la Tabla A.7, en amperios. Cada $\mathcal{C}_j$ es el término del escenario j en (3.22), calculado en las muestras 900 a 1500, de 0,30 a 0,50 s, con las corrientes de la carga. El sesgo es $I_b|\mu_e|$ por eje y σ es la desviación estándar de la corriente. La desviación estándar adicional, $\frac{1}{2}(\sigma_{i_d} + \sigma_{i_q}) - \mathcal{C}_j^*$, es una diferencia y no una desviación estándar. Es negativa cuando la corriente varía menos que su referencia. Los resultados se reprodujeron después del entrenamiento con la red y la configuración guardadas en el punto de control.

	Sesgo		σ					$\Delta\mathcal{C}_j$, de ellos	
Esc.	d	q	i_d	i_q	$\mathcal{C}_j$	$\mathcal{C}_j^*$	$\Delta\mathcal{C}_j$	sesgo	desv. estándar adicional
1	0,139	0,392	0,193	0,344	0,800	0,061	0,739	0,531	0,207
2	0,177	0,180	1,465	7,068	4,624	4,358	0,266	0,358	−0,091
3	0,159	0,062	0,066	0,037	0,272	0,000	0,272	0,221	0,052
4	0,581	0,029	0,497	0,498	1,107	0,184	0,924	0,610	0,314
5	0,537	0,169	0,615	2,033	2,030	1,395	0,635	0,706	−0,071
6	0,218	0,104	3,346	8,349	6,169	6,092	0,077	0,322	−0,245
7	0,036	0,308	5,615	7,056	6,679	6,062	0,617	0,343	0,273
8	0,048	0,148	0,006	0,005	0,201	0,000	0,201	0,196	0,006
media	0,237	0,174	1,475	3,174	2,735	2,269	0,466	0,411	0,056

A.4. Ajuste de los controladores

El Capítulo 5 desarrolla los criterios de Fröhr [27] y las ganancias del control convencional. Aquí se reúnen las constantes utilizadas y la identificación del lazo con el agente, necesaria para ajustar los lazos externos con los pesos de la red fijos.

A.4.1. Lazo interno de corriente del control convencional

Para la planta de (5.2), $\tau_\sigma = 7{,}3727$ ms y los retardos de adquisición, actuación y retención se agrupan en $\tau_c = 1{,}5T_s = 0{,}50$ ms. El ajuste por óptimo del módulo de (5.3) conduce a una constante de tiempo equivalente $\tau_{eqi} = 2\tau_c = 1{,}0$ ms. El tiempo muerto medido, 1,003 μs, se trata por separado en la Sección A.7.

A.4.2. Realización del PI y compensaciones del equipo

Este apartado desarrolla las ecuaciones y los coeficientes de la implementación descrita en la Sección 5.4. La forma de estado que sigue es equivalente al PI programado con una integral y una corrección por saturación; la función $C(z)$ describe su comportamiento cuando no actúan los límites.

La realización desplegada discretiza (5.1) por Euler hacia adelante con período T_s y realimenta la salida saturada u_{sat} hacia el estado interno [1, 29],

$$C(z) = k\,\frac{z - z_0}{z - 1}, \qquad x[n] = z_0\,x[n-1] + c_2\,u_{\text{sat}}[n-1], \qquad u[n] = k\big(e[n] - x[n]\big), \tag{A.8}$$

donde $k = k_p$, $z_0 = 1 - T_s/T_i$, $c_2 = (z_0 - 1)/k$, n indexa las muestras y x es el estado que acumula la acción integral. Esa realimentación es local: informa del recorte del propio bloque, y no del límite vectorial posterior al lazo de corriente ni de la prioridad del eje directo posterior al de velocidad. La actuación interna de esos límites posteriores no quedó registrada en las campañas del Capítulo 7.

En la implementación experimental se añade a la salida de los controladores de corriente la prealimentación de dos términos: el acoplamiento entre ejes, cuya compensación propone [2] cuando el ancho de banda del lazo es reducido, y la tensión inducida del eje de cuadratura, que es un añadido propio de este trabajo,

$$v_{sd}^{*} = u_d - \omega_s\,\sigma L_s\,i_{sq}, \qquad v_{sq}^{*} = u_q + \omega_s\,\sigma L_s\,i_{sd} + k_r\,\omega_s\,|\boldsymbol{\psi}'_r|, \tag{A.9}$$

donde u_d y u_q son las salidas de los controladores y ω_s la frecuencia del marco sincrónico. Los términos se calculan con las corrientes medidas y con el flujo y el deslizamiento del estimador. La compensación es una acción adicional a la salida del PI y no modifica el criterio empleado para calcular sus parámetros [2].

La frecuencia del marco empleada en (A.9) corresponde al programa ensayado. Al agrupar la planta con $R_\sigma = R_s + k_r^2 R'_r$, el término inducido del eje q contiene la velocidad eléctrica del rotor ω. Como $\omega_s = \omega + \omega_{\mathrm{sl}}$ y $\omega_{\mathrm{sl}}|\boldsymbol{\psi}'_r| = R'_r k_r i_{sq}$,

$$k_r(\omega_s - \omega)|\boldsymbol{\psi}'_r| = k_r^2 R'_r i_{sq} \simeq (0{,}761\ \Omega) i_{sq}. \tag{A.10}$$

Este término se puede agrupar con la resistencia si se escribe la planta con R_s. Aquí se conserva la prealimentación implementada; no se identifica con una cancelación exacta de la planta escrita con R_σ.

Los coeficientes compilados son los de (5.3), pero la cadena de normalización posterior del equipo introduce un factor $\sqrt{3}/2$: la tensión que ordenan los controladores y la prealimentación se normaliza con $V_{dc}/\sqrt{3}$, y el modulador la aplica con $V_{dc}/2$. Por tanto, el lazo físico completo no reproduce exactamente el óptimo del módulo ideal, y su ganancia proporcional efectiva es de 11,37 V/A. En estado estacionario el integrador lo compensa, y el factor no se corrigió en el programa ensayado. La ganancia del lazo de velocidad está compilada como constante normalizada, que el controlador convierte a 131,42 A/pu.

El modelo del PI del Capítulo 6 usa las ganancias del programa ensayado: $k_p = 13{,}1315$ V/A y $k_i = 1781{,}1$ V/(A s). Cada PI limita su salida a $\pm V_{dc}/\sqrt{3}$ y aplica el antienrollamiento con ese recorte. Después se suman los términos de prealimentación de (A.9). La orden total se normaliza por $V_{dc}/\sqrt{3}$ y se limita en módulo a uno; este último recorte no se realimenta al PI. Después se aplica Min–Max y se calcula el ciclo de trabajo en la escala $V_{dc}/2$, por lo que se reproduce el factor $\sqrt{3}/2$ y la ganancia proporcional efectiva de 11,37 V/A.

En los lazos externos, el modelo conserva el PI de velocidad y el filtro de su medida empleados con el agente, de 80 Hz nominales. El PI de flujo utiliza el ajuste convencional del Capítulo 5, distinto del aplicado con el agente (Tabla A.10). Para la prealimentación usa el filtro del *firmware*, discretizado por Euler hacia adelante. El *DC link* es ideal, de 300 V; no se reproducen la cuantización ni la corrección del *encoder*, y se conserva el cálculo del ángulo IFOC del modelo.

A.4.3. Lazos externos del control convencional

Los dos lazos externos ven una planta distinta y les corresponde otro criterio. El lazo de flujo ve la dinámica del rotor (2.4), de constante $\tau_r = 193{,}39$ ms, en serie con el lazo cerrado de corriente, cuya constante equivalente es τ_{eq}. El cociente entre la constante grande y la suma de las pequeñas vale

$$\frac{\tau_r}{\tau_{eq}} = \frac{193{,}39}{1} \simeq 193 > 4, \tag{A.11}$$

y a ese caso le corresponde el óptimo simétrico realista, que además ofrece mejor rechazo a las perturbaciones. El lazo de velocidad ve la ecuación mecánica, que es un integrador, en serie con el mismo lazo cerrado de corriente, y una planta con acción integral cae también en el óptimo simétrico. Con ese método la constante del controlador se relaciona con la equivalente del lazo interno por (5.4), con $a = 2$, de modo que $T_i = 4\,\tau_{eq}$. Para $\tau_{eq} = 1$ ms resulta $T_i = 4$ ms y para $\tau_{eq} = 12$ ms, $T_i = 48$ ms.

El lazo de flujo del control convencional se ajustó tomando $\tau_{eq} = 1$ ms, la aproximación de primer orden $2\,\tau_c$ del lazo de corriente ajustado por óptimo del módulo, y a ese valor corresponden la ganancia de la Sección 5.3 y el cociente de (A.11). El de velocidad emplea en los dos accionamientos la sintonía experimental de la Sección 5.3, la que el criterio entrega con $\tau_{eq} = 12$ ms sin sumar el filtro de la medida. Esa constante es la del lazo de corriente con el Agente 2 refinado, identificada en la Sección A.4.4. Su ganancia está compilada en el equipo como constante normalizada, y el controlador la convierte a 131,42 A/pu.

A.4.4. Ajuste de los lazos externos con el Agente 2 refinado

Los pesos del agente permanecen fijos; el ajuste siguiente sólo modifica los lazos externos. Como la red no tiene una función de transferencia conocida, su lazo cerrado se identificó en PLECS con pequeñas variaciones cuadradas de ±0,23 A a 7 y 10 Hz sobre las referencias de corriente. Un modelo de primer orden ajustado y validado con datos distintos entregó, en cuadratura, τ_{eq} entre 11,8 y 13,4 ms, ganancia de continua entre 0,95 y 0,98 y varianza explicada del 94 al 95 %. Al repetir el procedimiento con la red utilizada en la campaña se obtuvieron 10,8 a 13,2 ms y aproximadamente un 94 % de varianza explicada. Por ello se adoptó $\tau_{eq} = 12$ ms. Esta constante describe la dinámica de pequeña señal en simulación, no el tiempo de inferencia ni los transitorios de gran amplitud.

Con este valor,

$$\frac{\tau_r}{\tau_{eq}} = \frac{193{,}39}{12} \simeq 16{,}1 > 4, \tag{A.12}$$

por lo que se mantiene el criterio de ajuste del lazo de flujo. La aproximación de (5.4) entrega $k_{p,\mathrm{imr}} = 8{,}0580$ y $T_i = 48$ ms. En PLECS, esa ganancia produjo una oscilación de 9,2 rpm y una desviación típica de 220 mA en i^*_{sq}: las variaciones lentas de i_{sd} hacían variar el flujo y el torque, y el lazo de velocidad respondía modificando i^*_{sq}. Un barrido por factores de dos mostró que la menor reducción que eliminaba esta interacción era 0,5. Se adoptó entonces $k_{p,\mathrm{imr}} = 4{,}0290$, manteniendo $T_i = 48$ ms; la misma prueba bajó a 0,48 rpm y 10 mA. Como T_i quedó fijo, el coeficiente integral $k_i = k_p/T_i$ también bajó a la mitad, a 83,9 s^{-1}. La reducción se aplicó a este diseño del agente, no al PI de flujo convencional.

La reducción de ganancia aumentó el sobrepaso de la magnetización de aproximadamente 7,7 % a 12,1 %. Para moderarlo se ensayó el alisamiento de la referencia i^*_{mr} del lazo de flujo, que genera i^*_{sd}, dentro del intervalo de $4\tau_{eq}$ a $6\tau_{eq}$ propuesto por Fröhr [27]. El barrido entre 0 y 100 ms seleccionó 60 ms: en PLECS el sobrepaso bajó a 8,2 %, frente al 8,24 % medido en el banco. Estos dos cambios constituyen el ajuste experimental utilizado con el agente.

La identificación no fue válida en el eje directo ni con excitaciones de gran amplitud. En esos casos la acción incremental limita la pendiente de la modulante y la respuesta

deja de parecerse a un primer orden. Por tanto, $\tau_{eq} = 12$ ms sólo se emplea para sintonizar los lazos externos alrededor del punto de operación; no se usa para interpretar los tiempos experimentales de la Figura 7.10.

A.4.5. Cadena de ganancias intermedias

La medida de corriente conserva una cadena física de sensores y adaptación de señal antes del convertidor analógico–digital. Sus ganancias se incorporan a la calibración con que el programa convierte las cuentas a amperes y luego normaliza por I_b. La velocidad utiliza una escala calibrada aparte. Que el controlador sea digital no elimina estas ganancias: permite expresar las señales que recibe en unidades físicas o en por unidad, sin volver a introducir las ganancias de la cadena ya calibrada en el ajuste de los lazos.

En la salida, la modulante se convierte en tensión media mediante (2.7). La comprobación de esta conversión y la de las escalas de entrada son necesarias para relacionar el controlador implementado con el modelo empleado en su ajuste.

El contraste entre la modulante publicada y la componente fundamental de la tensión fase–neutro comprueba la conversión de salida. La Tabla B.3 presenta diferencias del 0,78 % con el control convencional y del 1,9 % con el agente. Esta comprobación no verifica las escalas de corriente o velocidad, que requieren sus propios contrastes y la calibración descrita en el Apéndice B.

A.4.6. Lazo de velocidad común a ambos accionamientos

El lazo de velocidad y el filtro de su medida son comunes a ambos accionamientos. Con 32 768 cuentas por vuelta y $T_s = 333{,}333$ μs, una cuenta del *encoder* equivale a 0,003662 pu. La sintonía estricta amplificó esa cuantización y produjo en el banco un ciclo límite de 75 a 80 Hz, observado también con controladores PI en el lazo interno. La sintonía conservadora basada en $\tau_{eq} = 12$ ms eliminó ese comportamiento.

La medida de velocidad se filtra con un polo de 80 Hz, elegido mediante un barrido: a 50 Hz el rezago ya afectaba la respuesta y a 200 Hz el *ripple* reaparecía en la referencia de corriente. El filtro añade aproximadamente 2 ms a la dinámica vista por el controlador y se conserva en ambos accionamientos.

Como comprobación adicional, el ajuste teórico que incluye ese filtro se probó sobre el banco con el lazo de corriente convencional. La magnetización respondió según el diseño, pero al girar a 0,30 pu apareció una oscilación sostenida de 50 a 65 Hz y de 10 a 11 A *peak to peak* en i^*_{sq}, seguida por i_{sq}. Esta prueba confirmó que el banco tenía menos margen que el modelo en esa banda y respaldó la sintonía conservadora de 12 ms. Fue un ensayo auxiliar del 11 de septiembre de 2026 y no forma parte de la comparación experimental del Capítulo 7. Las Figuras A.5 y A.6 muestran el resultado.

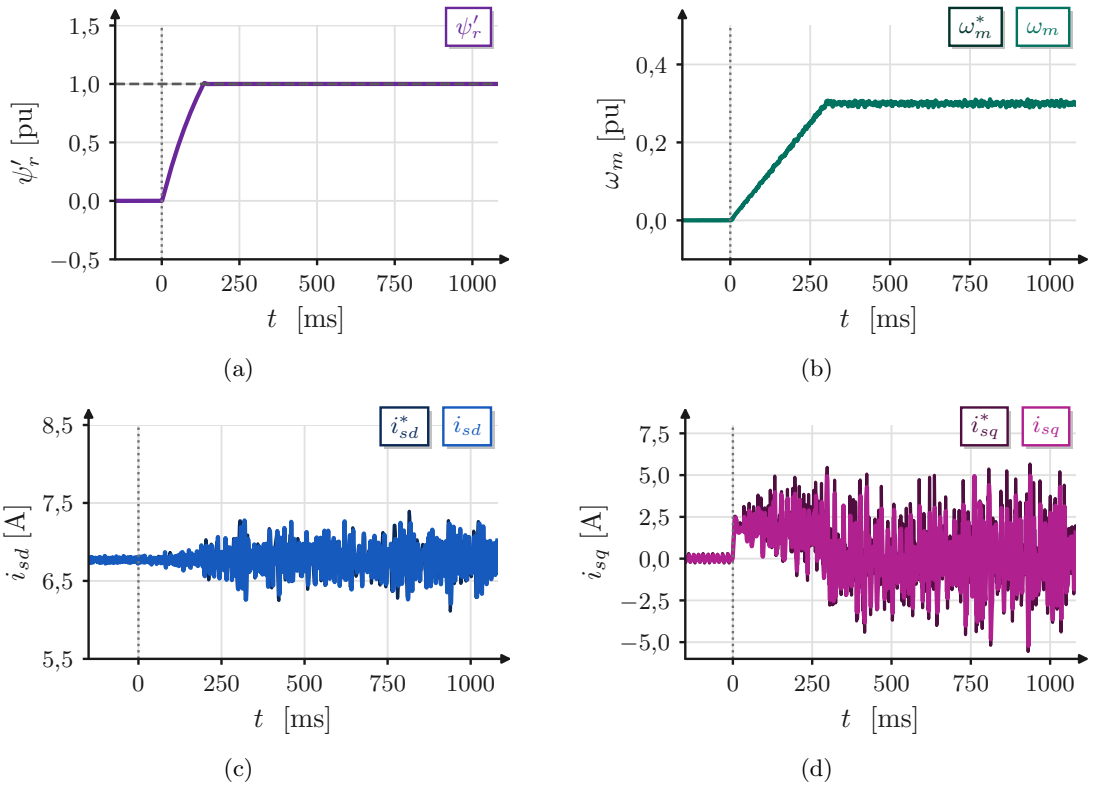


Figura A.5: Verificación experimental de la sintonía teórica del lazo de velocidad, con $k_{p,\omega} = 33{,}82$ y $T_{i,\omega} = 11{,}96$ ms. (a) flujo de rotor estimado en la magnetización; (b) velocidad mecánica y su referencia en el arranque; (c) corriente del eje directo con su referencia; (d) corriente del eje de cuadratura con la suya. La magnetización y el flujo permanecen operativos; al entrar la máquina en giro aparece una oscilación sostenida en i^*_{sq}, seguida por i_{sq}, por lo que esta sintonía no se empleó en las campañas definitivas. Condiciones: referencia de 0,30 pu; sin carga mecánica; control y PWM a 3 kHz; *DC link* de 300 V; el tiempo se mide desde el inicio del cambio de referencia.

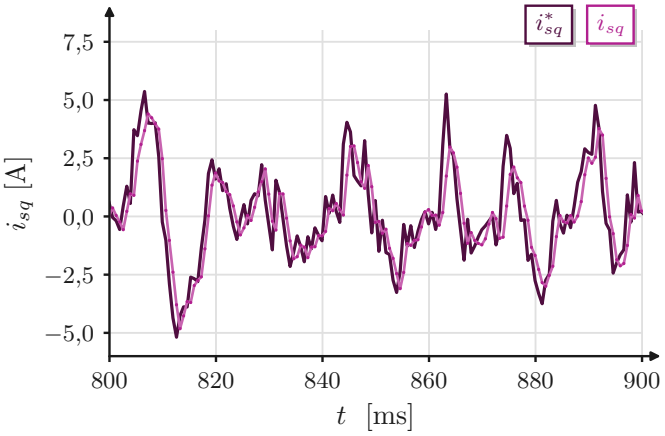


Figura A.6: Detalle del ciclo límite observado con la sintonía teórica del lazo de velocidad. La oscilación se origina en la referencia i^*_{sq} que genera el lazo externo y la corriente medida la sigue. La frecuencia observada se encuentra entre 50 y 65 Hz. Condiciones: las de la Figura A.5.

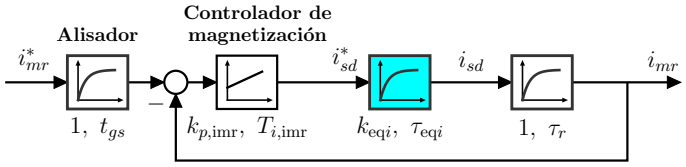
A.4.7. Configuración de las campañas experimentales

El control convencional emplea el ajuste del lazo de corriente por óptimo del módulo y el del lazo de flujo por óptimo simétrico realista [1, 27], descritos en el Capítulo 5. El lazo de velocidad conserva el ajuste común de ambas campañas. La Figura A.7 muestra los lazos externos con el agente, que se comparan con los del control convencional de la Figura 5.2, y la Tabla A.10 reúne la configuración efectiva del entrenamiento, de las dos simulaciones y de las dos campañas.

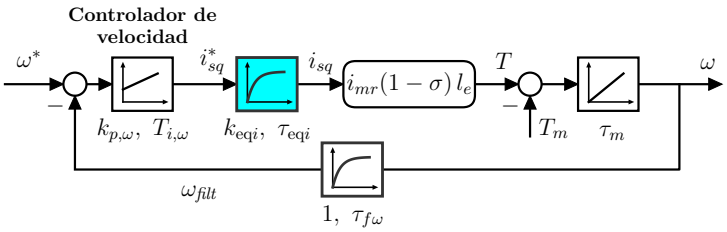
La unidad registra referencias, corrientes en el marco dq, flujo estimado, velocidad y modulantes. Los osciloscopios miden corrientes de fase, las tensiones v_{xN} y la tensión del *DC link*. El flujo se obtiene del estimador y el torque se analiza a partir de i_{sq}; ninguno dispone de una medición independiente.

El Agente 2 refinado sustituye el control convencional del lazo interno de corriente y conserva los pesos aprendidos en el entorno $R-L$. La reflexión del eje de cuadratura permanece activa. No se añade ningún *trim*, integrador ni compensador destinado a eliminar su error en estado estacionario.

El lazo de velocidad es el común, ajustado con los 12 ms del lazo con el Agente 2 refinado. El PI de flujo, en cambio, no usa con el agente el ajuste convencional de $k_{p,\mathrm{imr}} = 96{,}6955$ y $T_{i,\mathrm{imr}} = 4$ ms. Parte del diseño asociado a la constante equivalente de 12 ms, $k_{p,\mathrm{imr}} = 8{,}0580$ con $T_{i,\mathrm{imr}} = 48$ ms, y reduce sus ganancias a la mitad: $k_{p,\mathrm{imr}} = 4{,}0290$ con el mismo $T_{i,\mathrm{imr}}$. Además, un alisamiento de primer orden de 60 ms actúa sobre la referencia i^*_{mr} del lazo de flujo. La referencia alisada se utiliza para calcular el error del PI y como prealimentación al formar i^*_{sd}; el alisamiento no filtra la corriente medida, la velocidad ni la salida de la red. El control convencional no lo utiliza. Por ello, la comparación evalúa los accionamientos completos y no aísla el efecto del control del lazo de corriente.



(a) flujo



(b) velocidad

Figura A.7: Lazos externos con el Agente 2 refinado, en la misma forma que la Figura 5.2. En (a), el alisador de constante t_{gs} actúa sobre la referencia i^*_{mr} antes del controlador de magnetización; (b) es el lazo de velocidad común a los dos accionamientos. El bloque cian representa el lazo de corriente cerrado con el agente como un primer orden equivalente, empleado sólo para ajustar los controladores externos. Bajo cada bloque se indican su ganancia y su constante de tiempo; los valores están en la Tabla A.10. Elaboración propia.

Tabla A.10: Configuración efectiva del entrenamiento, de las simulaciones del Capítulo 6 y de las campañas del Capítulo 7, en cinco bloques. El entrenamiento sólo concierne al agente. En los bloques 2 y 3 cada columna corresponde a un control, y una celda que abarca las dos indica un valor común. Cuando la simulación y el banco difieren, se indican ambos. Los PI de flujo y velocidad trabajan en por unidad.

1. Entrenamiento del Agente 2 refinado, sobre la carga $R-L$		
DC link	50 a 300 V, sorteado en cada episodio	
Base de tensión	$V_N/2$ por componente, $V_N = 650$ V	
Límite de la modulante	$\pm m_{\text{lím}}$ por componente y umbral radial	
Referencias de corriente	hasta 17 A	
ω_{pu} observada	frecuencia del marco, fija en cada episodio	
2. Lazo interno de corriente, en la máquina		
	Control convencional	Agente 2 refinado
Controlador	PI en los ejes d y q: $k_{p,i} = 13{,}1315$ V/A; $T_{i,i} = 7{,}3727$ ms	red DDQN con pesos fijos
Compensaciones	acoplamiento entre ejes y tensión inducida en q	ninguna; reflexión del eje q con $\omega_{\text{pu}} < -0{,}02$
ω_{pu} observada	no aplica	simulación: frecuencia del marco; banco: velocidad eléctrica medida y filtrada más el deslizamiento estimado
Base de tensión	orden normalizada por $V_{dc}/\sqrt{3}$ y aplicada con $V_{dc}/2$	escala V_N, convertida a $V_{dc}/2$ con V_N/V_{dc}
Límite de la modulante	cada PI, $\pm V_{dc}/\sqrt{3}$; orden total, $\lvert\boldsymbol{m}\rvert \leq 1$	simulación: $\lvert m_d\rvert \leq 2/\sqrt{3}$ y $\lvert\boldsymbol{m}\rvert^2 \leq 4/3$; banco: $\pm m_{\text{lím}}$ en el estado y $\lvert\boldsymbol{m}\rvert \leq 1$ en la orden
3. Lazo de flujo		
	Control convencional	Agente 2 refinado
PI de flujo: k_p; T_i	96,6955; 4 ms	4,0290; 48 ms (k_p reducido a la mitad desde el diseño de 8,0580; T_i sin cambio)
Alisamiento de i^*_{mr}	no	primer orden de 60 ms
Límite de i^*_{sd}	común: 0 a 13,54 A, el doble de la corriente de magnetización nominal	
4. Lazo de velocidad, común a los dos accionamientos		
PI de velocidad: k_p; T_i	8,4243, equivalente a 131,42 A/pu con I_b; 48 ms. Ajustado para la constante equivalente de 12 ms del lazo de corriente con el Agente 2 refinado	
Filtro de la medida	primer orden de 80 Hz nominales; Euler hacia atrás en simulación y hacia adelante en el banco	
Rampa de ω^*	1 pu/s	
5. Ejecución y medidas, comunes salvo el *DC link* del banco		
Frecuencia de control	3 kHz, con un período de 333,33 μs, también en el entrenamiento	
Bases	$I_b = 15{,}6$ A; $\omega_b = 314{,}16$ rad/s eléctricos, también en el entrenamiento	
DC link	simulación: 300 V ideal; banco: medido, 282 a 285 V con el control convencional y 281 a 284 V con el Agente 2 refinado	
Límite de $\lvert\boldsymbol{i}^*_s\rvert$ con prioridad de d	simulación: 15,6 A; banco: 15,0 A	
Medida de velocidad	simulación: ideal, sin cuantización; banco: *encoder* de 32 768 cuentas, con el ángulo corregido por 1,0233	

A.5. La red dentro del procesador y su tiempo de inferencia

Se exportan los pesos, sesgos y la tabla de incrementos de modulante del punto de control seleccionado a encabezados de C en precisión simple [25]. Estos parámetros quedan fijos en el banco; el entorno de entrenamiento no se ejecuta en la placa.

En cada período, el programa construye las quince entradas de la observación, evalúa las dos capas ocultas con ReLU y la salida lineal de 225 valores Q, y elige la acción de mayor valor. Su índice identifica una combinación de incrementos de modulante. Se conservan el orden de entradas, las escalas de normalización y el orden de esas combinaciones; los pesos permanecen fijos. No intervienen la red objetivo, la memoria de experiencias ni la exploración.

A 3 kHz, el período de 333,3 μs debe alojar adquisición, construcción de la observación, inferencia, selección de la acción y actualización de la modulante. Por ello se midieron por separado la red sola y el ciclo completo dentro de la interrupción.

La red suma 47 585 parámetros (186 KiB en precisión simple) y requiere 47 104 multiplicaciones con acumulación por evaluación. Los pesos se compilan como constantes y residen en la memoria interna de nivel dos, de 256 kB. Al no disponer de sistema de archivos ni memoria externa para leerlos, cambiar de agente requiere recompilar y programar el procesador.

La evaluación con un acumulador tarda 304,1 μs. La versión utilizada reparte los productos entre ocho acumuladores y tarda 227,9 μs en promedio, con un máximo de 228,5 μs. Las 32 observaciones de prueba produjeron la misma acción que la red de referencia. Esta prueba verifica la conversión en esos casos; la limitación sobre la identificación del *firmware* exacto de la campaña se declara en el Apéndice C.

La lectura tomada al terminar la campaña dio 239,2 μs para el ciclo completo dentro de la interrupción. Si excede el período, el programa conserva el ciclo de trabajo anterior, registra el incumplimiento y pasa a estado seguro tras varios incumplimientos consecutivos. En esa lectura, el contador de ciclos tardíos valía cero. El contador se reinicia cada vez que el agente toma el lazo. Estos valores se midieron en las condiciones de la campaña y no constituyen una cota de peor caso. Estos tiempos describen la ejecución y no la constante de tiempo del lazo cerrado de la Sección A.4.

A.5.1. Comprobación de ejecución en las campañas

La Figura A.8 muestra el ciclo del agente dentro del período de control. Durante los ensayos no se registraron ciclos tardíos y las decisiones de prueba coincidieron con las de la red de referencia (Tabla A.11).

Tabla A.11: Ejecución del agente en la unidad de control. Las 32 decisiones de prueba coincidieron con la red de referencia; CCS registró cero fallos.

Indicador	Valor
Tiempo de evaluación de la política [μs]	227,9
Máximo registrado [μs]	228,5
Ciclo del agente dentro de la interrupción [μs]	239,2
Período de control [μs]	333,3
Decisiones coincidentes en la prueba del procesador	32/32

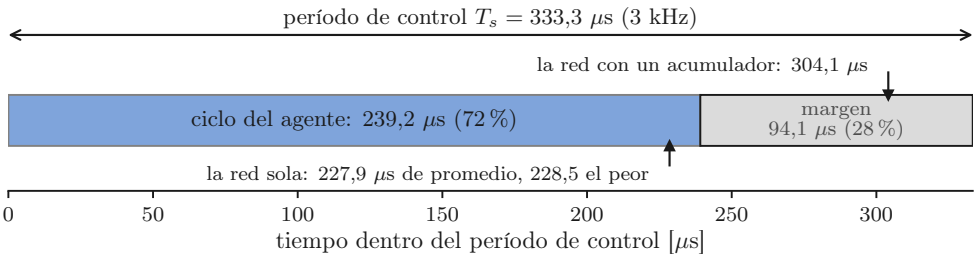


Figura A.8: Tiempo de ejecución del agente: ciclo medido dentro de la interrupción (azul) y tiempo restante del período (gris). Las marcas muestran la red implementada y la misma red evaluada sin la optimización de ocho acumuladores; ambas se midieron sin interrupciones.

A.6. Cadena de actuación del agente

La acción del agente está expresada en la base fija $V_N/2 = 325$ V con que se entrenó, justificada en la Sección A.3.1. La tensión disponible es $V_{dc}/2$, y con $V_{dc} = 300$ V, $V_{dc}/V_N = 0{,}461538$. Fuera de la red, el modelo de la máquina multiplica el incremento por V_N/V_{dc}, lo suma a la modulante anterior en la base $V_{dc}/2$ y aplica un límite vectorial a $2/\sqrt{3}$, con prioridad del eje directo. Ese límite se fijó según la zona lineal que permite el modulador con Min–Max y no pertenece al agente. El resultado es su memoria, y la observación siguiente lo recibe multiplicado por V_{dc}/V_N, en la base $V_N/2$. La Tabla A.12 ordena estas operaciones en el entrenamiento, el modelo y el banco. En los tres, la memoria es la modulante acumulada después de su límite.

Tabla A.12: Operaciones entre la acción elegida y la tensión aplicada, en orden. Tensiones en amplitud de fase, con $V_{dc} = 300$ V en el modelo y 282 V en el banco.

Operación	Entrenamiento	Modelo de la máquina	Banco
1. Acción	mayor Q o exploratoria	mayor Q; reflexión si $\omega_{\text{pu}} < -0{,}02$	como el modelo
2. Acumulación	$\boldsymbol{m}_{dq}(k-1) + \Delta\boldsymbol{m}_{dq}(k)$	la misma, con $\Delta\boldsymbol{m}_{dq} \times V_N/V_{dc}$ (base $V_{dc}/2$)	como el entrenamiento
3. Recorte de la modulante acumulada	$\pm m_{\text{lím}}$ por eje en la base $V_N/2$ (equivale a $\pm V_{dc}/2$)	vectorial, fuera de la red: $\lvert m_d\rvert \le 2/\sqrt{3}$ y $m_d^2 + m_q^2 \le 4/3$, con prioridad de d ($V_{dc}/\sqrt{3}$: 173 V)	$\pm m_{\text{lím}}$ por eje en la base $V_N/2$ (141 V)
4. Memoria (entradas 8 y 9)	resultado de 3	resultado de 3; a la observación, $\times V_{dc}/V_N$	resultado de 3
5. Cambio de base	–	hecho en 2	$\times V_N/V_{dc}$, V_{dc} filtrado
6. Límite de módulo	–	–	$\lvert\boldsymbol{m}\rvert \le 1$ (141 V), sobre el vector
7. Paso a *abc*	–	ángulo del marco	ángulo del marco
8. Min–Max	–	sí	sí
9. Ciclo de trabajo	–	0 a 100 %; no actúa	12,5 a 87,5 % (122 V)
Límite que actúa al final	3	3	9

La simulación sobre la carga $R-L$ del Capítulo 4 sigue la columna del banco, sin el paso 6 y con el ciclo de trabajo entre 0 y 100 %. La campaña $R-L$ usa $\lvert\boldsymbol{m}\rvert \le 0{,}95$ en el paso 6.

La Figura A.9 sigue un paso del modelo de la máquina, en $t = 2{,}000$ s del arranque con $V_{dc} = 300$ V. Es una reproducción posterior de una observación archivada. La acción elegida es la 128 en la numeración de C, que es la 129 en MATLAB. Los ciclos de trabajo que resultan son iguales a los registrados por el modelo.

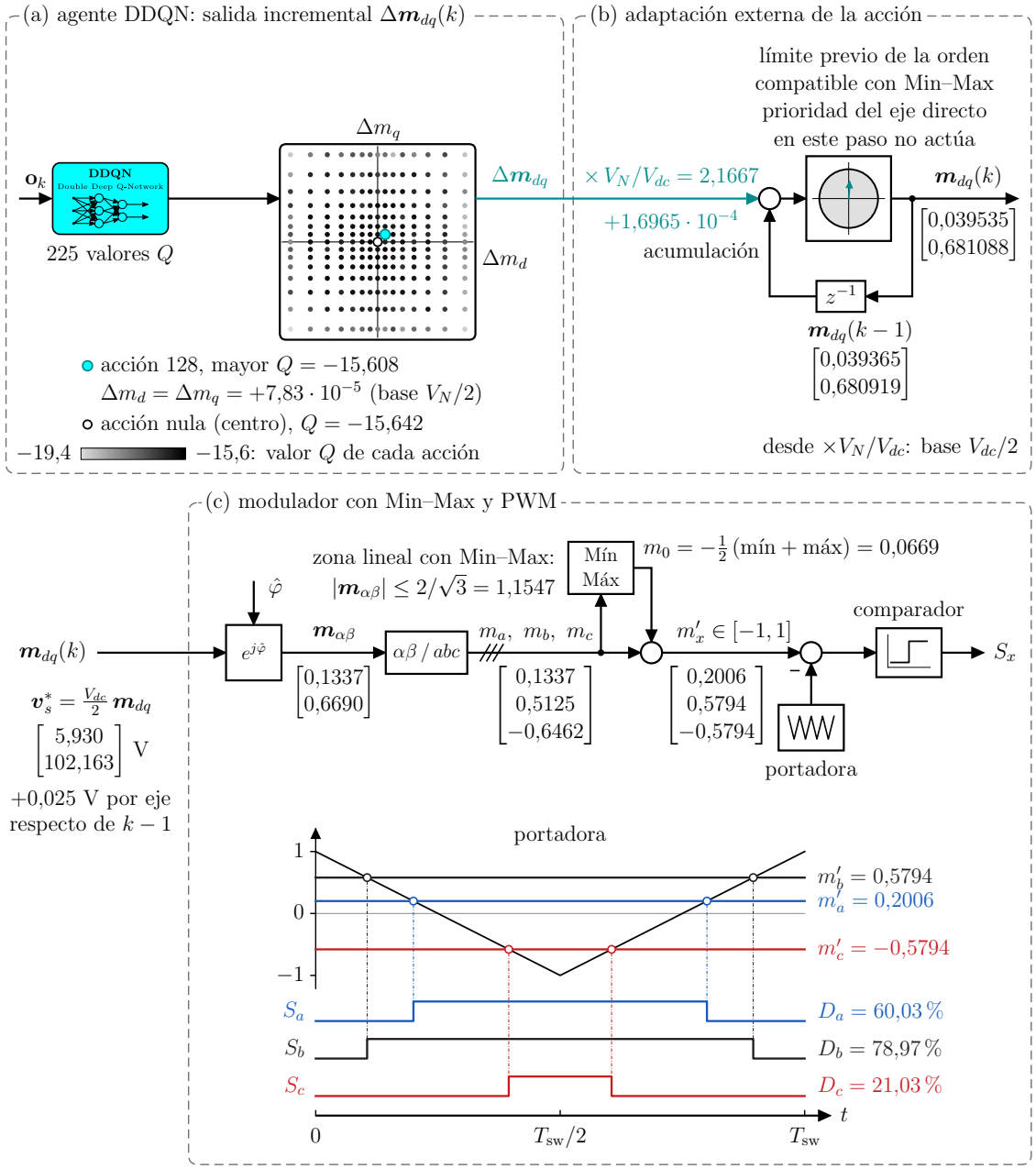


Figura A.9: Un paso de actuación del modelo de la máquina, en $t = 2{,}000$ s del arranque, con $V_{dc} = 300$ V. (a) El agente selecciona el incremento $\Delta \boldsymbol{m}_{dq}(k)$, la acción de mayor Q. Cada punto de la grilla es una acción, en gris según su valor, con separación no a escala. (b) Fuera de la red, el modelo cambia la base del incremento, lo acumula y limita la orden según la zona lineal del modulador con Min–Max. (c) Min–Max agrega la componente común y deja las tres modulantes de fase dentro de $[-1, 1]$ antes de compararlas con la portadora. Elaboración propia.

Min–Max no amplía la orden, sino que suma a las tres fases una componente común que las deja entre -1 y 1. Por eso el límite previo a $2/\sqrt{3}$ aprovecha todo el margen lineal.

La observación de ese paso se evaluó con el archivo de pesos de la simulación con la máquina identificado en la Tabla C.2. Con las corrientes en pu de I_b y la modulante en la base $V_N/2$, sus quince entradas, en el orden de la Figura 6.2, son

$$\boldsymbol{o}_k = \big[\,-0{,}039024;\ 0{,}019121;\ -0{,}000477;\ 0{,}000472;\ 0{,}404894;\ 0{,}012906;\ 0{,}299761;\ 0{,}018169;\ 0{,}314270;\ 0{,}998333;\ 0{,}057715;\ 0{,}029605;\ 0{,}014551;\ 1;\ 0{,}9375\,\big]^{\mathsf{T}}.$$

El recorte del paso 9 en el banco se fijó para la carga $R-L$ como un tope de comando calculado con 10 A, 15 Ω y 300 V, y permaneció activo en las campañas de la máquina. En la máquina definió un límite de tensión, no una cota fija de 10 A.

A.7. Tiempo muerto medido en el banco

El tiempo muerto lo fija un contador de la lógica programable y no la portadora. Se midió en las condiciones de las campañas, con la portadora a 3 kHz, sobre tres capturas independientes, con el *DC link* desconectado y descargado y los ciclos de trabajo forzados al 50 %, de manera que la pierna conmuta sin intervención del lazo. La Figura A.10 muestra las dos transiciones.

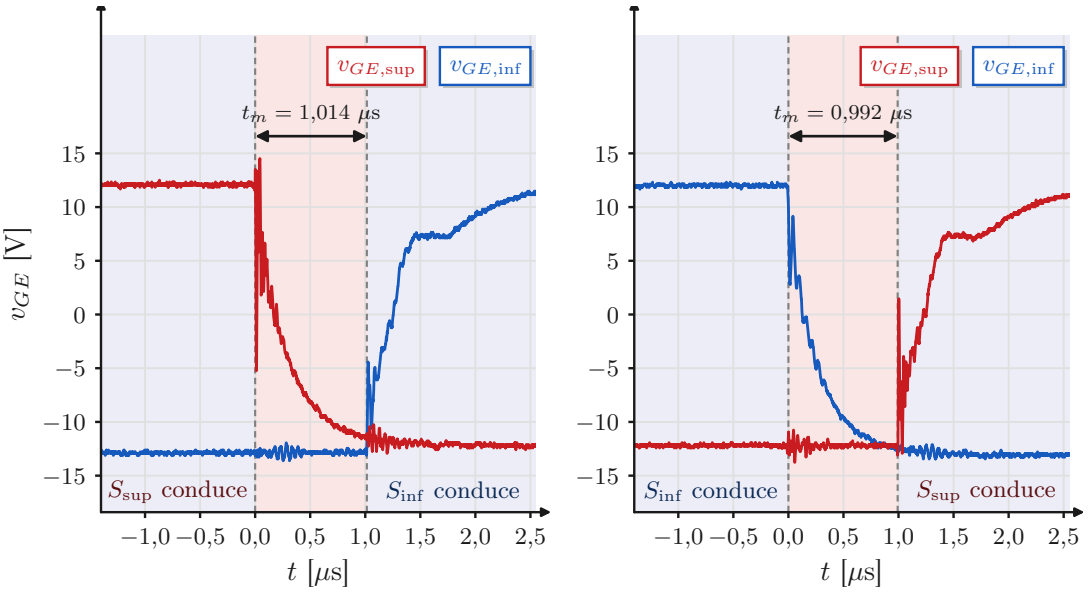


Figura A.10: Tensiones de compuerta de los dos IGBT de un mismo semipuente durante las dos transiciones. Las bandas y los rótulos de conducción representan el estado de las señales de compuerta, no una medición de corriente en los interruptores. Los instantes señalados son los cruces por 0 V de cada señal. Condiciones: portadora de 3 kHz, ciclos de trabajo forzados al 50 %, *DC link* desconectado y descargado.

Las tres capturas dan 1,014, 0,992 y 1,002 μs, con media de 1,003 μs y un recorrido entre extremos de 22 ns, frente a 1,000 μs comandados. A esa portadora la duración representa el 0,30 % del período de conmutación.

La separación media medida por los cruces de compuerta, 1,003 μs, se distingue del valor comandado, 1,000 μs. Al adoptar 5,5 V como criterio de habilitación, se obtiene una ventana de 1,30 μs. Esta segunda estimación depende del criterio elegido y no mide directamente la conducción del diodo bajo carga ni la distorsión de tensión de la pierna.

El entorno de entrenamiento modela esa distorsión con el término $(4/\pi)(T_d/T_{\mathrm{sw}})$ tomando $T_d = 1{,}0$ μs. La ventana estimada con el criterio de 5,5 V es un 30 % mayor que ese tiempo. Esa diferencia entre tiempos no constituye una medida del error de tensión del modelo. No se corrige, porque el agente que controla el banco es el que se entrenó con ese modelo, y queda como una más de las diferencias declaradas entre el entorno de entrenamiento y la planta real.

A.8. La modulación en un período de portadora

La Figura A.11 abre un período de portadora de 333,3 μs en el arranque. Entre flancos de subida de la pierna a median 332,0 μs con el control convencional y 332,5 μs con el agente; el ciclo de trabajo desplaza esos flancos. La media entre centros de pulso en 60 períodos es 333,3 μs en ambas campañas. Los ciclos de trabajo de a, b y c son 78,8, 37,8 y 21,1 % con el control convencional, y 76,8, 29,3 y 22,7 % con el agente.

La tensión fase–neutro se reconstruye con (2.5); la tensión del *DC link* se contrasta mediante su medición directa (Tabla B.3).

En ventanas de un segundo de magnetización y arranque, cada pierna registra aproximadamente 6000 transiciones, sin pulsos perdidos ni añadidos. La separación entre subidas es de 0,97 a 1,03 períodos. El intervalo mínimo entre transiciones durante el arranque es 57 μs con el control convencional y 49 μs con el agente. Este recuento describe la conmutación; las pérdidas no se calcularon.

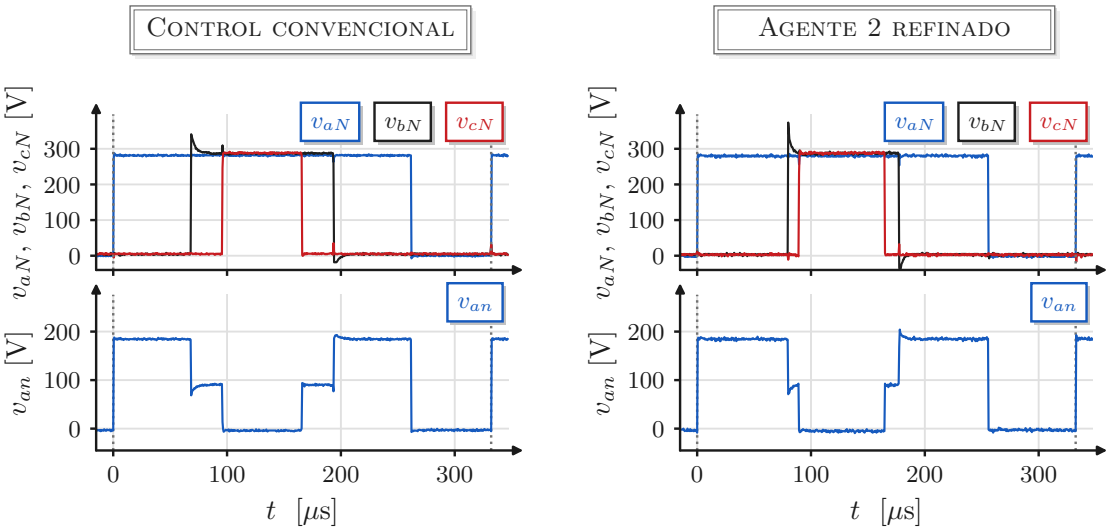


Figura A.11: Arranque: (a) control convencional; (b) Agente 2 refinado. Arriba, tensiones v_{xN} medidas entre cada salida del inversor y N, el negativo del *DC link*; abajo, tensión fase–neutro reconstruida. Las verticales marcan dos subidas consecutivas de v_{aN}. Osciloscopio B, 2 MS/s, acoplado en continua; condiciones de la Tabla B.1.

A.9. Circuito de frenado

La inversión de marcha devuelve energía al *DC link*; el circuito de frenado limita su tensión. La Tabla A.13 da el ajuste común a ambos controles y el Apéndice B, su verificación experimental.

Tabla A.13: Circuito de frenado: resistencia medida, formada por tres elementos de 15 Ω en serie por rama y dos ramas en paralelo.

Entrada en conducción, V_{on}	330 V	Resistencia medida, R_{ch}	22,5 Ω
Salida, V_{off}	315 V	Potencia continua, P_{cont}	1000 W
Disparo de la protección, V_{trip}	550 V	Conducción continua máxima, t_{on}	0,5 s

Con carga, la máquina auxiliar devuelve energía al *DC link* de su variador. Para una inversión sin carga desde 0,40 pu, la energía cinética calculada con la inercia del modelo es $E = J\omega_m^2/2 = 118{,}4$ J. Esta cifra no mide la energía devuelta al *DC link* ni establece una velocidad máxima del banco. El aumento de velocidad requiere revisar la energía y los límites térmicos de la Tabla A.13; la conducción durante el frenado tiene duración limitada.

Apéndice B

Validez de las campañas y límites experimentales

Este apéndice presenta los criterios de validez de las capturas y las comprobaciones que complementan los resultados del cuerpo. Las dos campañas sobre la máquina utilizan la escala de velocidad calibrada; las anteriores se conservan en el archivo experimental y no respaldan las comparaciones publicadas. Los criterios siguientes respaldan la interpretación de las magnitudes y la selección de los registros.

B.1. Cómo se miden las métricas dinámicas

Para una transición de y_0 a y_f se define $z(t) = [y(t) - y_0]/[y_f - y_0]$. El tiempo de subida separa los primeros cruces de $z = 0{,}1$ y $z = 0{,}9$; el sobrepaso es $M_p = 100\,\text{máx}(0, \text{máx}\, z - 1)$; y el asentamiento comienza cuando la señal entra en $y_f \pm 0{,}05|y_f - y_0|$ y permanece allí hasta el final de la ventana. La misma definición sirve para aumentos y disminuciones [30, cap. 5]. Como la referencia de velocidad es una rampa de 1 pu/s y el accionamiento tiene límites, estos valores sólo describen las maniobras ensayadas.

En la Tabla 6.2, los cruces se interpolan entre muestras. Para i_{sd}, la ventana termina antes de que disminuya la referencia: a 134,7 ms con el control convencional y 124,0 ms con el agente. Para i_{sq} se promedian las muestras de los 10 ms anteriores y se evalúan los 280 ms posteriores al arranque. El flujo se evalúa durante 900 ms y la velocidad desde 250 ms antes hasta 1100 ms después del inicio de la rampa, promediando las muestras de los 20 ms anteriores.

En los ensayos, la velocidad se procesa con ese promedio de 20 ms; las figuras conservan la señal sin suavizar. En la inversión, el tiempo de subida corresponde al recorrido aproximado entre $+0{,}24$ y $-0{,}24$ pu, y no a la duración completa de la maniobra. Los indicadores de corriente se calculan con los registros dq de la unidad de control. En magnetización se usa i_{sd} sin suavizar. En el arranque y al iniciar el frenado, i^*_{sq} continúa variando porque procede del lazo de velocidad; por eso referencia y corriente se promedian con las 15 muestras anteriores (10 ms), y el RMSE y el sesgo se calculan en los primeros 100 ms. Este procedimiento compara el seguimiento de cada corriente respecto de su propia referencia, no la respuesta de ambos controles ante un mismo escalón.

No se informa sobrepaso ni asentamiento de i_{sq} porque su referencia no permanece fija. En la magnetización del agente tampoco se informa el asentamiento de i_{sd}: la corriente vuelve a la banda de ±5 % sólo 6,7 ms antes de que la referencia comience a disminuir, intervalo insuficiente para comprobar permanencia (Figura B.1).

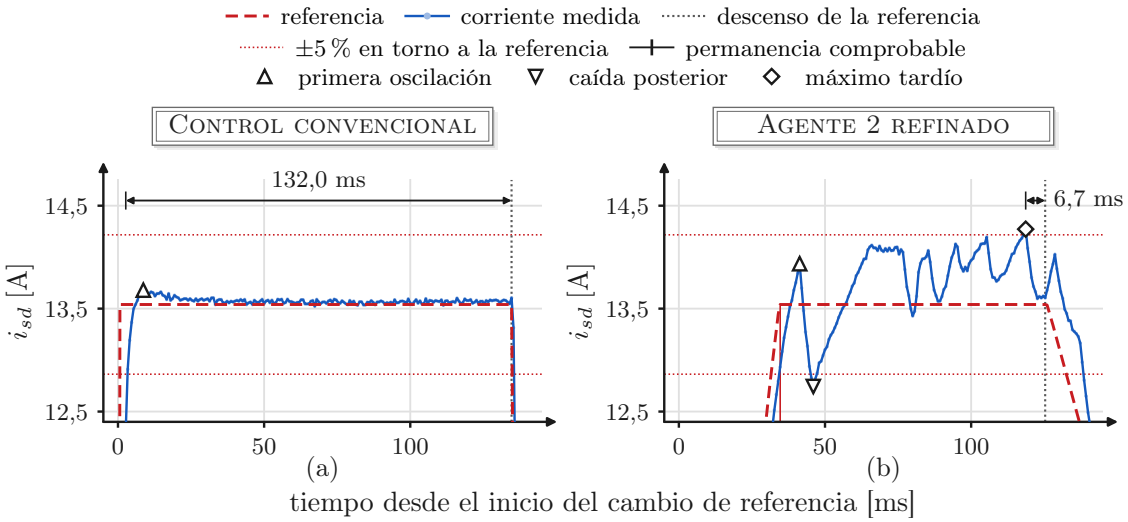


Figura B.1: Ampliación de i_{sd} durante la magnetización. La banda de ±5 % se centra en la referencia de 13,54 A. Con el agente, la corriente vuelve a entrar en la banda apenas 6,7 ms antes de que la referencia comience a disminuir; el registro no permite comprobar su tiempo de asentamiento.

El contraste de la Sección 7.4 aplica estos procedimientos a las simulaciones del Capítulo 6, recortadas con la duración de los registros de la unidad de control: la magnetización hasta el inicio del arranque, y el arranque y la inversión de marcha desde 250 ms antes hasta 1,1 s después del inicio de la rampa. En el flujo, y_f es la referencia de 1 pu, los cruces se interpolan entre muestras y el asentamiento se informa también para la banda de ±2 %. El error de corriente, $e_x = i_x^* - i_x$, se resume con el RMSE, el sesgo y la desviación estándar del error definidos en el Capítulo 4, en los últimos 200 ms de cada registro. Se utiliza la desviación estándar del error, no la de la corriente, y se verifica $\mathrm{RMSE}^2 = \bar{e}^2 + \sigma_e^2$ antes del redondeo. Los intervalos finales pueden empezar en instantes distintos respecto del inicio de la maniobra; no se presentan como ventanas temporalmente idénticas.

B.2. Intervalos de medición durante las maniobras

Los intervalos de la Figura B.2 se utilizan para evaluar el seguimiento cuando la referencia de corriente del eje de cuadratura es aproximadamente constante. Se excluye la subida inicial. La Tabla B.1 conserva las referencias, las corrientes medidas y los errores de cada tramo. El estado estacionario corresponde a los últimos 200 ms de cada registro.

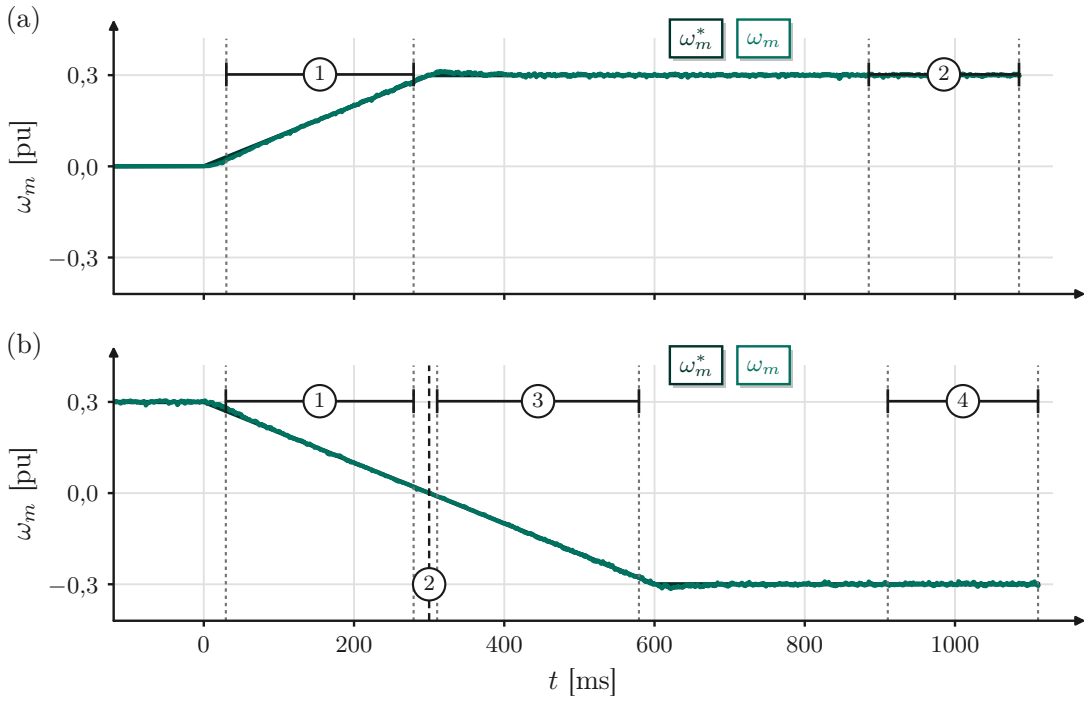


Figura B.2: Ventanas de medición sobre los registros del control convencional, con la velocidad de referencia y la medida. Las líneas verticales punteadas delimitan los tramos de la Tabla B.1 y los últimos 200 ms de cada registro; la línea vertical discontinua marca el paso por velocidad cero. (a) arranque: ① arranque y ② estado estacionario. (b) inversión de marcha: ① frenado, ② paso por velocidad cero, ③ giro en sentido contrario y ④ estado estacionario. Condiciones: las de la Tabla 7.1.

Tabla B.1: Seguimiento de corriente durante los tramos en que la referencia de cuadratura es aproximadamente constante, excluyendo la subida inicial. La inversión se separa en frenado y giro en sentido contrario por el paso por velocidad cero. Las ventanas se determinan a partir del lazo de velocidad, común a ambas campañas. Condiciones: las de la Tabla 7.1.

lazo interno	tramo	i_{sq}^* [A]	i_{sq} [A]	error [mA]	i_{sd}^* [A]	i_{sd} [A]	error [mA]
control convencional	arranque	+1,807	+1,785	+21	6,777	6,771	+6
	frenado	−1,826	−1,815	−11	6,755	6,770	−14
	giro en sentido contrario	−1,832	−1,803	−29	6,770	6,770	+1
Agente 2 refinado	arranque	+2,469	+2,192	+277	6,549	6,795	−246
	frenado	−1,531	−1,782	+251	6,590	6,770	−180
	giro en sentido contrario	−2,714	−2,436	−278	6,475	6,798	−323

Las medias finales de cuadratura se conservan en la Tabla B.2. Recalculadas desde los registros de la unidad de control con sus últimas 301 muestras, dan los mismos seis valores. En el arranque son −0,3564 A con el control convencional y +0,0100 A con el agente. Con la desviación estándar de cada registro, el error estándar de esas dos medias es de unos 0,03 y 0,07 A, y la del agente no se distingue de cero. Permiten describir el punto de operación

sin mezclarlas con las estadísticas del error. En el código y en los registros se revisaron el signo de giro, la transformación de coordenadas, la asignación de las pinzas y el perfil de monitorización, sin encontrar errores. Un desplazamiento constante en la medición de las corrientes de fase tiende a convertirse en una componente alterna cuando el marco *dq* gira, y con el marco prácticamente detenido puede aparecer como valor medio. En la magnetización, con el eje detenido, las medias son de apenas −0,020 y −0,007 A. En esas ventanas la velocidad se mantiene en ±0,30 pu y varía menos de 0,006 pu/s. Por tanto, la velocidad es prácticamente constante en esas ventanas. Con el control convencional, i_{sq} sigue a su referencia, −0,356 A frente a −0,345 A en el arranque, de modo que ese signo lo pidió el PI de velocidad.

Con la convención de (2.2), −0,36 A corresponde a un torque electromagnético negativo. La ausencia de carga mecánica externa no elimina el roce ni las pérdidas del conjunto acoplado, que se oponen al giro. En las dos campañas, el variador de la máquina de carga permaneció apagado. Ese estado no quedó registrado en los archivos de los ensayos. En estado estacionario, el torque electromagnético equilibra el torque resistente y tiene el signo del giro.

La respuesta durante el arranque confirma la polaridad del marco. Mientras la velocidad sube con la rampa de 1 pu/s, i_{sq} es de +1,75 A con el control convencional y de +2,20 A con el agente. Así, una corriente de cuadratura positiva produce una aceleración positiva, en el sentido horario, observado desde el lado por donde sale el cable del *encoder*. La media negativa se obtuvo después, entre 0,59 y 0,79 s tras el fin de la rampa. En esa ventana, la velocidad supera a su referencia en sólo 0,00005 pu y la acción proporcional del PI de velocidad aporta −0,007 A. Ese término no explica por sí solo la referencia negativa de −0,345 A, que proviene casi por completo de la acción integral. Esa acción siguió bajando y pasó de −0,35 a −0,38 A dentro de la ventana. Como la integral todavía variaba, esa media no basta para identificar el valor final de estado estacionario y no se descarta un transitorio residual.

Además, la media no se repite entre sesiones. El control convencional, con el mismo lazo de velocidad, dio a +0,30 pu +0,16 y +0,30 A el 29 y el 30 de agosto y −0,36 A el 10 de septiembre. A −0,30 pu dio −0,55 A el 30 de agosto y −0,11 A el 10 de septiembre. Entre esas sesiones, el programa sólo cambió las ganancias del lazo de corriente, que no determinan por sí solas el valor medio de i_{sq} en estado estacionario.

Esa diferencia entre sesiones, de hasta 0,66 A, es compatible con un error de orientación del marco que cambie de una sesión a otra. Con i_{sd} de unos 6,8 A, cada grado de error traslada unos 0,12 A a i_{sq}, y 0,66 A equivalen a unos 6°. El ángulo exacto del marco estimado no se midió con un instrumento independiente, de modo que esa causa no queda demostrada. La media de una sola sesión no se usa por sí sola para inferir el torque ni para atribuir la diferencia al controlador.

Tabla B.2: Corriente media de cuadratura en los últimos 200 ms de cada registro, en el marco *dq* que estima el control. No mide directamente el torque, porque depende de la orientación de ese marco. Los errores se presentan en la Tabla 7.3.

Maniobra	Control convencional [A]	Agente 2 refinado [A]
Magnetización	−0,0204	−0,0073
Arranque	−0,3564	+0,0100
Inversión de marcha	−0,1125	−0,5691

B.3. Inicio de las maniobras en las capturas

El tiempo de las figuras se mide desde el inicio de cada maniobra, no desde el disparo del osciloscopio. Las corrientes y las tensiones se capturan con instrumentos distintos; por eso cada registro se alinea por separado. En la carga $R{-}L$, el inicio corresponde al cambio de referencia.

Con el agente, se detectan la aparición de corriente en la magnetización, el inicio del giro del vector de corriente en el arranque y la intersección del tramo de frecuencia constante con la rampa en la inversión. Los desplazamientos de las capturas de corriente y tensión son, respectivamente, 6 y 2 ms en la magnetización, 82 y 86 ms en el arranque, y 1652,4 y 1629,1 ms en la inversión. La inversión completa dura 601,4 y 600,7 ms; estos tiempos incluyen el frenado y el giro en sentido contrario.

Con el control convencional, las capturas se alinean con el registro de la unidad de control. Se utiliza la frecuencia eléctrica de la corriente y el módulo de la tensión, porque cerca de velocidad cero la tensión cae a unos 7 V y no permite medir bien su frecuencia. La magnetización no se desplaza; los desplazamientos son 40,0 y 48,6 ms en el arranque, y 2438,9 y 2438,0 ms en la inversión.

B.4. Validez de captura: modo sombra y perfiles de monitorización

En modo sombra el programa reutiliza los canales de las referencias para publicar las corrientes transformadas por la rama del agente, y una referencia que reproduce el *ripple* de la medida no permite calcular error de seguimiento. Cada archivo se comprobó con el cociente $\mathrm{rms}[\Delta i^{*}_{sd}]/\,\mathrm{rms}[\Delta i_{sd}]$, entre 0,97 y 1,02 en las capturas en sombra y entre 0,01 y 0,03 en lazo cerrado. La prueba invalida la campaña a 0,40 pu, que tampoco sirve para fijar un umbral de saturación de tensión; el punto de 0,05 pu se excluye por la ausencia del registro homólogo del agente.

La segunda condición de validez es la coincidencia entre `g_mon_perfil` y `PERFIL`: tras programar el procesador la primera puede volver a cero sin que cambie la segunda, y el archivo conserva el rótulo de un perfil con el contenido de otro. En el perfil 0 los canales 3 y 9 llevan m_q y m_d, y en el perfil 2, el integrador del controlador de flujo e i_{mr}. El mapa `D.nombres`, el perfil efectivo y la prueba de *ripple* se verifican juntos antes de atribuir significado físico a una traza.

B.5. Comprobación de las mediciones externas

La escala del osciloscopio se contrastó con la unidad de control [25]; los resultados se reúnen en la Tabla B.3. La comparación utiliza el valor simultáneo del *DC link*, que durante las maniobras baja de unos 301 V a 282–285 V, y la velocidad corregida con el factor de calibración del *encoder*. Sin una evaluación de la incertidumbre, este contraste comprueba proximidad entre cadenas, pero no mide con precisión el deslizamiento.

Los rótulos del instrumento no identifican correctamente todas las señales: los canales de corriente aparecen en porcentaje, el cuarto canal figura como `I_rf` aunque contiene V_{dc} y los canales de tensión se denominan “1”, “2” y “3”. La revisión del cableado y una

magnetización del control convencional fijan el mapa de tensión como $1 \to v_{aN}$, $2 \to v_{cN}$ y $3 \to v_{bN}$. En esa prueba, v_{aN} tiene valor medio positivo, 2,64 V, y el mayor ciclo de trabajo, como corresponde al marco detenido en el origen.

Las pinzas rotuladas `IA`, `IB` e `IC` tampoco siguen el orden de fases. El reparto esperado $(+I, -I/2, -I/2)$ y el sentido común de giro de los vectores de tensión y corriente determinan la asignación $a \to$ `IC`, $b \to$ `IB` y $c \to$ `IA`. Con ella, ambas frecuencias coinciden en 15,36 Hz y el vector de corriente queda a 0,03° de la fase a. Las figuras aplican estas correcciones sin modificar los datos originales.

Tabla B.3: Contraste de las mediciones durante el estado estacionario del arranque. Para el agente se emplea la misma ventana de 200 ms en ambas cadenas. La tensión fase–neutro de la unidad se deduce de la modulante; la frecuencia se obtiene de la velocidad corregida del *encoder*.

Magnitud	Cadena	control convencional	Agente 2 refinado
Módulo de corriente [A]	Osciloscopio	6,720	6,744
	Unidad de control	6,784	6,815
	Diferencia [%]	0,94	1,04
Tensión del *DC link* [V]	Osciloscopio	282,2	281,6
	Unidad de control	282,8	281,1
	Diferencia [%]	0,2	0,2
Fundamental fase–neutro [V RMS]	Osciloscopio	70,03	71,00
	Unidad de control	69,49	69,70
	Diferencia [%]	0,78	1,9
Frecuencia eléctrica [Hz]	Osciloscopio	15,30	15,33
	Velocidad corregida	15,35	15,35
	Diferencia [Hz]	−0,05	−0,02

Durante la magnetización, el reparto de las corrientes de fase depende del ángulo en que se establece el marco sincrónico. Con el eje detenido, ese origen angular es relativo; una diferencia entre campañas no implica por sí sola un error de orientación [2].

La Tabla B.4 identifica la velocidad que usa cada parte del accionamiento en las campañas del Capítulo 7. La corrección de 1,0233 se obtuvo al orientar el marco. Sin ella, la orientación medida se desviaba unos +21° con velocidad positiva y −25° con velocidad negativa. En simulación, cada 1 % de error en la velocidad medida desvía el marco unos 10°, de modo que se estimó una lectura del *encoder* un 2,3 % baja. Con la corrección, la desviación bajó a −0,8° y −1,3°.

La referencia interna es 0,30 pu en base de 1500 rpm, equivalente a 450 rpm internas. Al aplicar el factor 1,0233 se obtiene una velocidad física estimada de 0,30699 pu, unos 460,5 rpm. Según la Tabla B.3, la frecuencia eléctrica medida con el osciloscopio, 15,30 y 15,33 Hz, queda a 0,05 Hz o menos de la velocidad corregida. De los 15,00 Hz de la medida sin corregir queda a unos 0,3 Hz. La velocidad física no se verificó con un instrumento independiente de velocidad.

Tabla B.4: Velocidad que usa cada parte del accionamiento en las campañas del Capítulo 7, su base y si lleva el factor 1,0233. $\omega_b = 314{,}16$ rad/s eléctricos y 1 pu equivale a 1500 rpm.

Señal o bloque	Velocidad que usa	Base	Factor
Figuras del Capítulo 7	medida del *encoder*, filtrada a 80 Hz	ω_b	no
PI de velocidad	la misma	ω_b	no
Protección de sobrevelocidad	la misma	ω_b	no
Entrada ω_{pu} del agente	la misma, más el deslizamiento estimado	ω_b	no
Umbral de la reflexión	la entrada ω_{pu}, contra $-0{,}02$ pu	ω_b	no
Deslizamiento estimado	ninguna. Sale de i_{sq} y de $\hat{i}_{mr}$, según (2.4)	rad/s	no aplica
Ángulo de orientación $\hat{\varphi}$	integral de la velocidad del *encoder* por 1,0233 más el deslizamiento estimado	rad	sí
≪Velocidad corregida≫ de la Tabla B.3	velocidad del *encoder* por 1,0233	Hz	sí
Simulación del Capítulo 6	la del modelo mecánico, sin error de medida	ω_b	no existe
Entrenamiento	frecuencia del marco de la carga $R-L$	ω_b	no existe

En el banco, la entrada ω_{pu} del agente suma el deslizamiento estimado a la medida sin corregir. El ángulo integra la medida multiplicada por 1,0233 más ese mismo deslizamiento. En el entrenamiento ambas velocidades coinciden.

B.6. Tensión del DC link y circuito de frenado

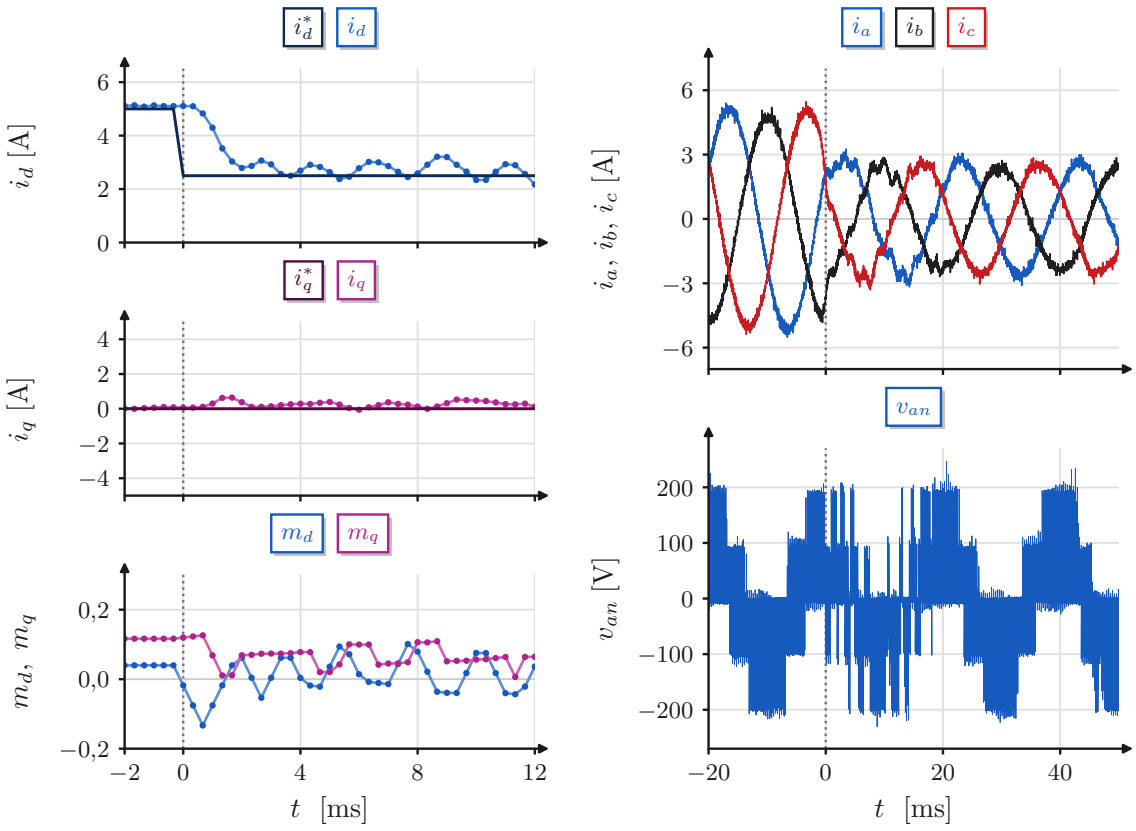
Durante la inversión de marcha la máquina devuelve energía al *DC link*, por lo que esta maniobra se examina frente al umbral del circuito de frenado. Con el control convencional la unidad de control registró un máximo de 294,8 V y el osciloscopio, remuestreado a la cadencia de aquélla, 294,9 V, un 0,03 % de diferencia entre dos cadenas independientes. En la captura cruda, muestreada a 2,5 μs, ninguna muestra alcanza los 315 V. Con el Agente 2 refinado la unidad registró un máximo de 295,2 V, y el osciloscopio, 295,9 V al promediar el milisegundo anterior, con un máximo instantáneo de 313,8 V. En las dos campañas la tensión que registra el control permaneció por debajo del umbral de 330 V y el contador `g_chop_n_eventos` quedó en cero, de modo que ninguna ejercita el circuito de frenado. Los máximos instantáneos de la captura cruda no se emplean para ese dictamen: el mayor de todos, 327,2 V, aparece en la magnetización del control convencional y dura una muestra de 2,5 μs.

El funcionamiento del circuito se verificó en un ensayo separado, con la máquina detenida y el control desconectado. Ajustado a un umbral seguro de 75 V, contuvo el *DC link* en 76,2 V, registró 575 intervenciones, acumuló 1,01 s de conducción y disipó 252 J. El ensayo demuestra que el circuito conduce y limita la tensión en esa configuración; no demuestra que haya intervenido durante la inversión ni valida su operación al umbral de 330 V.

La tensión del *DC link* con el agente durante el frenado se evalúa desde el inicio de la inversión de marcha hasta el paso por velocidad cero, de 0 a 306 ms. La media anterior abarca los 200 ms previos; el máximo suavizado promedia, en cada instante, las muestras del milisegundo anterior.

B.7. Bajada de la corriente del eje directo

La cuarta maniobra, la bajada del eje directo, tiene fila propia en las Tablas 4.3 y 4.4. Lo único que el cuerpo no publica de ella es la lámina, que recoge la Figura B.3.



(a) unidad de control (b) medida en bornes

Figura B.3: Bajada del eje directo de 5,0 a 2,5 A por las dos cadenas, con las escalas verticales de las Figuras 4.5 y 4.3, comparable con las tres del cuerpo. La amplitud del vector pasa de 5,1 a 2,5 A. Condiciones: las de la Figura 4.3, con $i_q^* = 0$ y *DC link* de 292,2 V de media en estado estacionario.

B.8. Perfil completo de las maniobras

Las Figuras B.4 y B.5 muestran las maniobras completas en simulación. Las modulantes acumulan las acciones del agente y se normalizan por $V_{dc}/2$. La última fila indica la acción seleccionada; la horizontal discontinua marca la acción nula, utilizada el 2,3 % del tiempo sobre la carga y el 8,6 % sobre la máquina. Este patrón no permite atribuir por sí solo el sesgo de corriente a la selección de acciones.

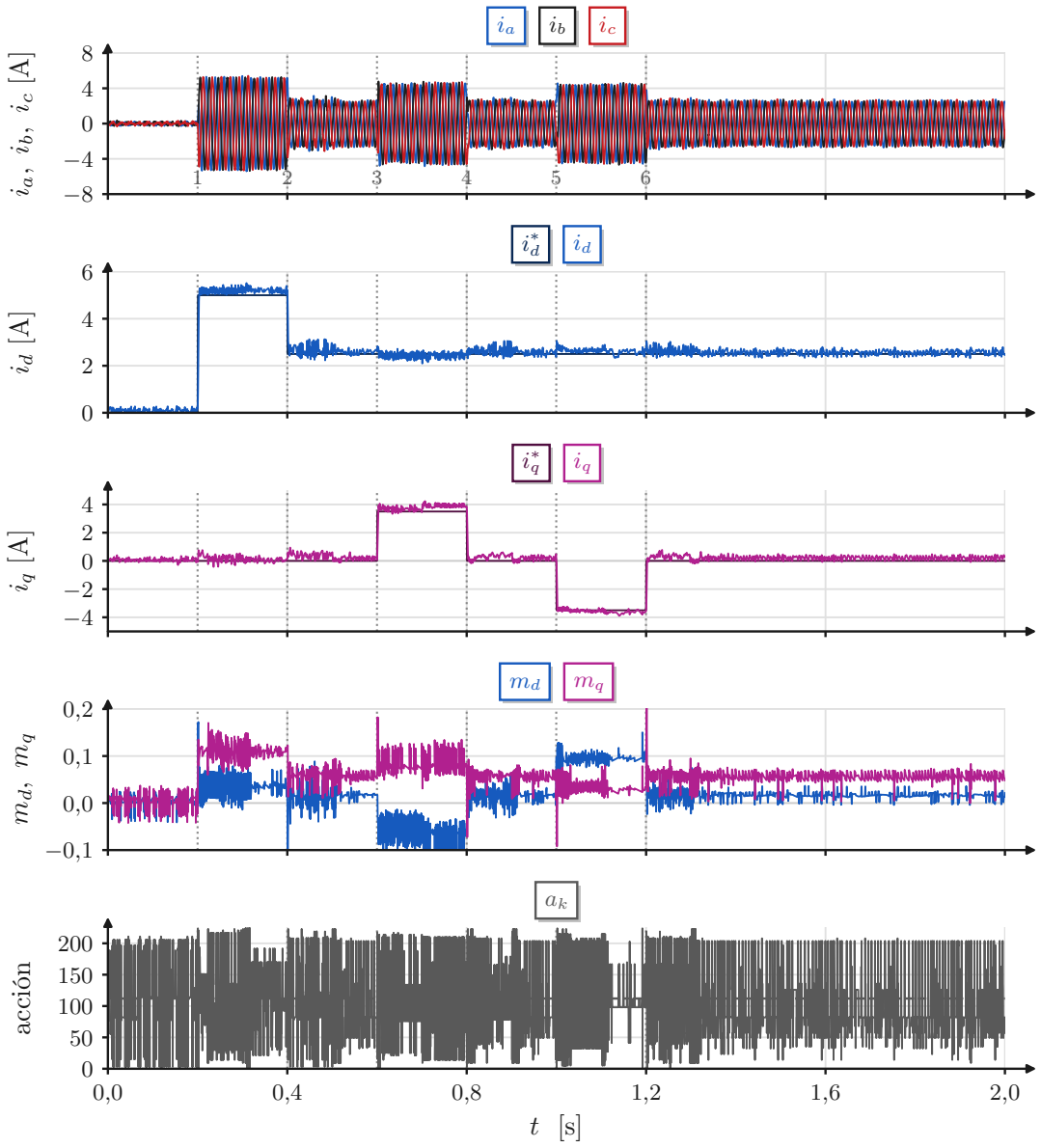


Figura B.4: Perfil completo sobre la carga $R-L$ con el agente en el lazo interno, en simulación, en una corrida con tramos de 200 ms, distinta de la del Capítulo 4, que usa tramos de 400 ms. Condiciones: $R = 1{,}00\,\Omega$, $L = 10{,}00$ mH, *DC link* de 300 V y lazo interno a 3 kHz. Las verticales punteadas numeran los seis cambios de referencia, detectados sobre el propio registro: aumento de i_d, reducción de i_d, aumento positivo de i_q, anulación de i_q, aumento negativo de i_q y anulación final de i_q.

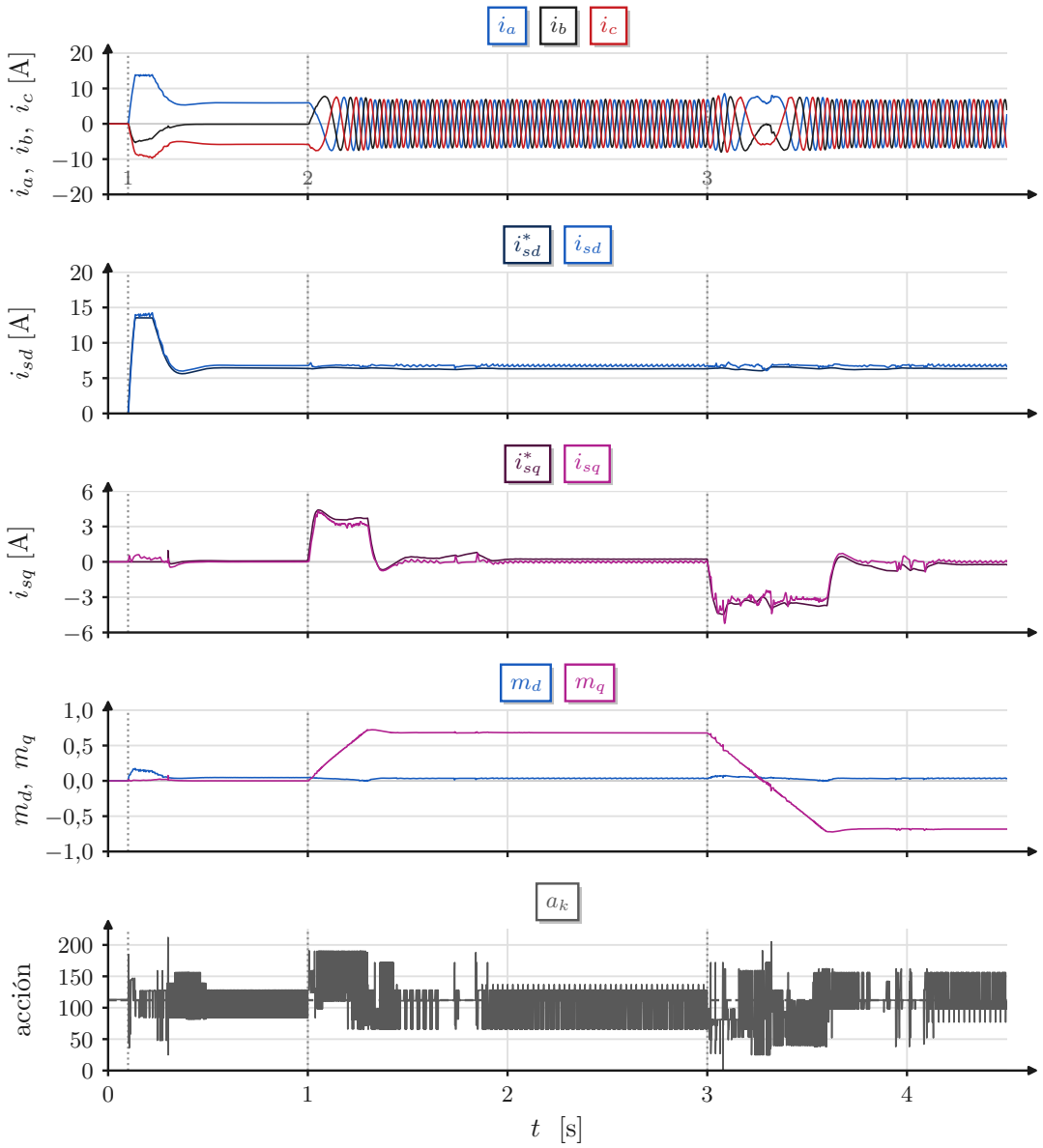


Figura B.5: Perfil completo sobre la máquina con el Agente 2 refinado en el lazo interno, en simulación. Condiciones: sin carga mecánica y *DC link* de 300 V. Las verticales punteadas marcan la magnetización, el arranque a +0,30 pu y la inversión a −0,30 pu. Las corrientes de fase se reconstruyen desde el marco sincrónico [2]; la comprobación de esa transformación es numérica y no una medida de precisión del instrumento. m_q alcanza 0,72 aquí y 0,17 en la Figura B.4.

B.9. Procedencia de las métricas

La Tabla B.5 reúne la procedencia de las métricas principales. Los registros *dq* de la máquina se guardan a 1500,15 Hz, una muestra de cada dos períodos de control, y la PWM conmuta a 3 kHz. Por eso el error, su desviación estándar y el *ripple peak to peak* calculados con esos registros no caracterizan la ondulación de conmutación. Las afirmaciones sobre esa ondulación, el *DC link*, la fundamental de v_{an} y el tiempo muerto se apoyan en las capturas de los osciloscopios, a 400 kSa/s o más.

Tabla B.5: Procedencia de las métricas principales: registro, maniobra, señal, ventana, número de muestras y filtrado. Los registros R1 a R7 se identifican al pie. Los modelos y programas de cada ensayo están en la Tabla C.2.

Métrica	Registro	Maniobra	Señal	Ventana	Muestras	Filtrado
Tabla 4.3, Exp.	R1	cuatro escalones	e_d, e_q	desde 30 ms tras el cambio hasta el fin del registro, 310 a 347 ms	950, 946, 856 y 840	ninguno
Tabla 4.3, Sim.	R2	los mismos	los mismos	la misma duración	949, 945, 855 y 839	ninguno
Tabla 6.3	R3	arranque e inversión	i_{sd}, i_{sq}	2,80 a 3,00 s y 4,30 a 4,50 s	601 por ventana	ninguno
Tabla 6.5	R4	inversión	$\lvert e_{sq}\rvert$, $\lvert e_{sd}\rvert$	1,60 a 3,00 s y 4,5 a 8 s; el global, desde 0,50 s	4200, 10 501 y 22 501	ninguno
Tablas 7.3 y B.2	R5 y R6	las tres	e_{sd}, e_{sq}, i_{sq}	últimos 200,6 ms de cada registro	301	ninguno
Tabla 7.2, i_{sq}	R5 y R6	arranque y comienzo del frenado	i_{sq}, i_{sq}^{*}	primeros 100 ms	150	media de 15 muestras, 10 ms
Tabla 7.2, velocidad	R5 y R6	arranque e inversión	ω	-250 a $+1100$ ms del inicio de la rampa	2025	media de 20 ms
Tabla B.3	R5, R6 y R7	arranque	$\lvert\boldsymbol{i}\rvert$, V_{dc}, v_{an}, frecuencia	convencional, 880 a 1080 ms; agente, 920 a 1120 ms tras el escalón	300 y 301 en la unidad	fundamental por proyección

Archivos de datos

R1: unidad de control en el banco con la carga $R-L$, 1 de septiembre, a 3000,3 Hz; carpeta `MANIOBRA_300V_SIN_TRIM\BRAIN`: `ventana_1_magnetizacion.mat`, `ventana_2_bajada_id_doble_a_nominal.mat`, `ventana_3_arranque.mat` y `ventana_5_inversion_de_marcha.mat`.

R2: modelo de la carga $R-L$, a 3000 Hz: `ACT1\mediciones.mat`.

R3: modelos de la máquina de 4,5 s, a 3000 Hz: `20_MAS_SIN_TRIM_PIF_050\mediciones.mat`, agente, y `3P_MAS_PI_FIRMWARE_10_09\mediciones.mat`, control convencional.

R4: modelo de la máquina de 8 s, a 3000 Hz: `67_MAS_SIN_TRIM_PIF_050_8s\mediciones.mat`, con reflexión, y `60_MAS_SIN_TRIM_SIN_ESPEJO\mediciones.mat`, sin reflexión.

R5: unidad de control con el Agente 2 refinado, 30 de agosto, a 1500,15 Hz; carpeta `07_AGENTE_SIN_TRIM\DATOS`: `ARRANQUE_AGENTE_300V_w300_20260830_233557.mat`, magnetización y arranque, y `ARRANQUE_AGENTE_300V_w300_20260830_234141.mat`, inversión.

R6: unidad de control con el control convencional, 10 de septiembre, a 1500,15 Hz; carpeta `09_PI_OM_10-09\DATOS`: `OPERAR_PI_300V_20260910_193235.mat`, magnetización; `ARRANQUE_PI_300V_w300_20260910_192755.mat`, arranque, y `ARRANQUE_PI_300V_w300_20260910_193225.mat`, inversión.

R7: los dos osciloscopios, en el arranque: `07_AGENTE_SIN_TRIM\OSCILOSCOPIO_DEFINITIVO\3.bin` y `4.bin`, agente; `09_PI_OM_10-09\OSC\a1.bin` y `b1.bin`, control convencional.

Apéndice C

Trazabilidad de agentes y registros

Este apéndice identifica los agentes y las sesiones que respaldan los resultados, y declara las limitaciones de reproducción. El inventario completo de archivos y huellas acompaña a los datos.

El archivo conserva veinte registros binarios de los osciloscopios: ocho de la carga $R{-}L$ y seis de cada campaña sobre la máquina. Las escalas se contrastan en la Sección B.5.

C.1. Cómo se identifica un agente

Cada agente se identifica por la huella SHA–256 de su contenido, porque existen encabezados con nombres iguales y pesos distintos. El paquete de datos conserva esa huella y los encabezados derivados en precisión simple, utilizados en la unidad de control [25], y doble, utilizados en simulación.

La carga $R{-}L$ se ensaya con el Agente 2; la máquina, en simulación y en el banco, con el Agente 2 refinado. Este último se obtiene al continuar el aprendizaje sobre la dinámica $R{-}L$ con la perturbación de tensión en q. La red seleccionada corresponde al episodio 675 de la etapa final y procede del episodio 1050 del refinamiento previo, según la configuración guardada en el propio punto de control. Los pesos permanecen fijos durante cada ensayo y ningún resultado físico de esta memoria contiene aprendizaje en línea.

La Tabla C.1 reúne la cadena completa de entrenamiento. El Agente 1 y la etapa de resistencia baja parten desde cero. Cada etapa siguiente continúa desde el punto de control de la etapa que indica la tercera columna, según la configuración guardada en cada punto de control. Las etapas intermedias no se ensayaron en los resultados de esta memoria. El entrenamiento del Agente 2 refinado se detuvo en el episodio 817 de los 1200 previstos.

La Tabla C.2 identifica el archivo de pesos y el modelo o el ejecutable de cada ensayo. Los pesos y sesgos de cada encabezado se compararon con los de su punto de control: coinciden número a número en los 47 585 parámetros, y las 225 acciones coinciden fila por fila. El encabezado del Agente 1, `ddqn_e6_ep1775_obs15.h`, de huella `623ec8a171862aea`, cumple la misma comparación.

En los encabezados de doble precisión del Agente 2 y en la copia del Agente 2 refinado que usan los modelos, la constante que designa la acción nula vale 113, su número de fila en MATLAB. En C, con numeración desde cero, la acción nula ocupa la posición 112, y la 113

corresponde al menor incremento positivo del eje directo. Esa constante sólo interviene si la red entrega un índice fuera de rango, lo que no ocurre, y para rotular el índice registrado cuando el control fija la modulante en cero. En las simulaciones conservadas no cambió ninguna acción aplicada. Los encabezados de precisión simple de la unidad de control usan 112. El encabezado exportado del Agente 2 refinado, de huella `f2b664a914a90f52`, se corrigió a 112 después de esas simulaciones, de modo que su huella difiere de la de la copia del modelo, que contiene los mismos pesos.

Tabla C.1: Cadena de entrenamiento de los agentes. La tercera columna indica la etapa cuyo punto de control se continúa. La carpeta es relativa a la raíz de los datos de entrenamiento y la huella son los 16 primeros caracteres del SHA–256 del punto de control.

	Etapa y cambio introducido	Parte de	Episodio	Carpeta y punto de control	Huella
1	Agente 1. R de 12 a 18 Ω, L de 10 a 20 mH, *DC link* de 100 a 210 V, corriente hasta 8 A y ω fija en 1 pu	cero	1775 de 2000	`11_ENTRENAMIENTO_CERO_E6\02_CHECKPOINTS\ckpt_ep1775.mat`	`5d4860905e315a38`
2	Resistencia baja. R de 0,5 a 3,5 Ω, *DC link* de 50 a 210 V, corriente hasta 5 A y nuevo conjunto de incrementos	cero	1425 de 2000	`13_ENTRENAMIENTO_E7_RBAJA\02_CHECKPOINTS\ckpt_ep1425.mat`	`47e49b7818bb73dd`
3	Agente 2. *DC link* de 50 a 320 V y corriente hasta 7,5 A	2	1650 de 2000	`15_ENTRENAMIENTO_E9_VALTA\02_CHECKPOINTS\ckpt_ep1650.mat`	`f3d8e6eaff9a6ac6`
4	Refinamiento, fase 1. Perturbación de tensión en q con $K = 1{,}9$ y tope $\bar{p} = 0{,}75$; ω de 0 a 1 pu y *DC link* de 50 a 300 V	3	525, el mejor de 3000	`17_AGENTE2_REFINADO\FASE1\02_CHECKPOINTS\best_e6.mat`	`cd67c8c1a4c9c7a1`
5	Refinamiento, fase 2. 12 % de los episodios en el punto del banco $R-L$; menor exploración y menor tasa de aprendizaje	4	400 de 800	`17_AGENTE2_REFINADO\02_CHECKPOINTS\ckpt_ep0400.mat`	`791c9ec0598d53de`
6	Corriente ampliada. Corriente hasta 17 A y tope $\bar{p}$ de 0,75 a 0,85	5	950 de 1200	`19_AGENTE2REF_MAQUINA\02_CHECKPOINTS\ckpt_ep0950.mat`	`4cf1bf6194db7a62`
7	Movimiento de la modulante. Peso de movimiento c_m de 0,30 a 1,50	6	675 de 1200	`21_ANTICHAT\02_CHECKPOINTS\ckpt_ep0675.mat`	`c694f836f520b44e`
8	Perturbación. K de 1,9 a 1,808, dimensionado con información nominal de la máquina, y p_q con signo opuesto al de ω	7	1050 de 1200	`23_PERTURBACION\02_CHECKPOINTS\ckpt_ep1050.mat`	`9e7bb8354f094f69`
9	Agente 2 refinado. ω de $-0{,}45$ a 0,45 pu; 35 % de episodios con $\omega = 0$, $p_q = 0$ y referencia de corriente sostenida	8	675 de 1200; detenido en 817	`25_ACOPLAMIENTO_D\02_CHECKPOINTS\ckpt_ep0675.mat`	`352671d16a28611e`

Tabla C.2: Archivos de pesos, modelos y ejecutables de cada ensayo. La huella son los 16 primeros caracteres del SHA–256. La simulación usa los encabezados de doble precisión y la unidad de control, los de precisión simple.

Ensayo	Archivo de pesos y huella	Modelo o ejecutable y huella	Comprobación
Capítulo 4, simulación con la carga $R-L$	Agente 2: `ddqn_e9_ep1650_obs15.h`, `adc8b4109e165a46`	Copia del modelo `RL_T400.plecs` que aplica el incremento del agente sin reducirlo, `42a7fdf7111a479b`	El modelo incluye el encabezado; pesos idénticos a los del punto de control
Capítulo 4, banco con la carga $R-L$, 1 de septiembre de 2026	Agente 2: `ddqn_obs15_weights_f32_E9.h`, `d67450d5fc06143d`	Ejecutable `BRAIn_PWM_LEZANA_3K15K.out`, `e11b14073f4ba0a4`	El manifiesto de la campaña anota `e11b1407`; el ejecutable contiene los pesos del Agente 2
Capítulos 6 y 7, simulación con la máquina	Agente 2 refinado: copia del modelo `ddqn_e9_ep0675_obs15.h`, `2b66e65529d0ead6`	Modelos `MAS_SIN_TRIM_PIF_050.plecs` y `MAS_PIF_050_8S.plecs`, idénticos, `3259401814d08c02`	El modelo incluye el encabezado; pesos idénticos a los del punto de control
Capítulo 7, banco con la máquina, 30 de agosto de 2026	Agente 2 refinado: `ddqn_obs15_weights_f32_E9.h`, `4add400227868612`	Proyecto de la unidad de control del agente; huella del ejecutable no conservada	Agente confirmado, según se indica en la Sección C.2
Capítulo 7, banco con la máquina, control convencional, 10 de septiembre de 2026	No aplica	Ejecutable `BRAIn_IFOC_PI_3K.out`, `3ce29cb66cb220a8`	Ejecutable lanzado y archivado durante la sesión

La Tabla C.3 relaciona cada línea del Algoritmo A.1 con el programa entregado, y la Tabla C.4 identifica la versión del programa con que se entrenó cada agente. La parte del objetivo DDQN, la pérdida, Adam, la actualización suave, el indicador terminal, el programa de exploración y la evaluación es igual en todas las versiones. El archivo del escenario de las etapas 4 a 6 se modificó después de sus puntos de control. El archivo del escenario de la etapa 9 es anterior a su punto de control.

Tabla C.3: Correspondencia de las líneas del Algoritmo A.1 con el programa de la etapa final, en la carpeta `25_ACOPLAMIENTO_D\01_CODIGO`.

Líneas	Archivo	Función y líneas del código
1	`train_e6.m`	`crearRed_`, l. 285-293
2	`train_e6.m`	`warmRescale_`, l. 218-283, llamada en l. 42-43
3 y 4	`train_e6.m`	l. 47-48 y 62-63
5	`train_e6.m`	l. 65 y 74-78; señal de parada, l. 69-72
6	`makeScenario_e6.m`	completo
7 a 14	`runEpisode_e6.m`	l. 71-146
15	`train_e6.m`	`pushBuf_`, l. 370-380
16 a 19	`train_e6.m`	l. 87-94; `ddqnUpdate_`, `lossFcn_`, `huber_` y `softUpdate_`, l. 316-362
20 y 21	`train_e6.m`	l. 100-125; `evalua_`, l. 295-306

Tabla C.4: Versión del programa de entrenamiento de cada etapa. La huella son los 16 primeros caracteres del SHA–256 del archivo.

Etapas de la Tabla C.1	`train_e6.m`	`runEpisode_e6.m`	Cambios de las reglas
1 y 2	`83573eedddeaed33`	`ec9851c85a89094b`	desde cero; frecuencia del marco fija en 1 pu; sin p_q
3, Agente 2	`43ddaa75a01589d1`	`ec9851c85a89094b`	parte de la etapa 2; entrada 15 multiplicada por $320/210 = 1{,}5238$
4 a 8	`6c3646c631f42445`	`3582fcb9e4d77493`	frecuencia del marco sorteada y p_q; en la etapa 4, la entrada 7 se absorbe en el sesgo al copiar los pesos; la etapa 8 se reanudó en el episodio 300
9, Agente 2 refinado	`6c3646c631f42445`	`3582fcb9e4d77493`	como las etapas 4 a 8, sin absorción ni reanudación; escenario de `makeScenario_e6.m` `803a8087846b4f1c`

C.2. Procedencia de los registros

La campaña principal sobre la carga R–L, del 1 de septiembre de 2026, consta de cuatro registros de 2000×10 valores, uno por maniobra. Cada uno se tomó como una captura independiente, después de alcanzar el punto de partida, aplicar el cambio, registrar y volver al reposo, y no como una secuencia continua. Las láminas del cuerpo muestran tres de ellos: el aumento de i_d desde cero y los aumentos positivo y negativo de i_q desde cero; el cuarto, la reducción del eje directo, se recoge en el Apéndice B. Las métricas agregadas incluyen las cuatro: ninguna es un dato descartado. La unidad de control registra diez canales a 3 kHz. En bornes, una cadena registra las corrientes de fase y la tensión del *DC link* a 400 kSa/s y la otra las tensiones v_{xN} a 800 kSa/s, con 4 y 8 MSa/s en la magnetización, de ventana más corta.

Los registros que sostienen el Capítulo 7 sobre la máquina proceden de dos campañas, ambas con la escala del canal de velocidad ya calibrada. La campaña con el control convencional descrito en [1] se realizó el 10 de septiembre de 2026; la del Agente 2 refinado, la noche del 30 de agosto.

El manifiesto de la campaña sobre la carga $R-L$ anota los ocho primeros caracteres de la huella del ejecutable cargado, `e11b1407`. Coinciden con el ejecutable conservado del proyecto, que contiene los pesos del Agente 2. De la campaña del control convencional se archivó el ejecutable lanzado durante la sesión. Se confirma que la campaña del 30 de agosto utilizó el Agente 2 refinado. Esta identificación es coherente con el proyecto conservado, cuyo encabezado contiene los pesos verificados de E25 episodio 675. No se conserva un manifiesto con la huella del ejecutable exacto cargado durante esa sesión.

La huella de los pesos identifica la red, pero no el programa que construyó la observación, aplicó los límites y registró las señales. De ese programa se conserva un ejecutable enlazado el 30 de agosto de 2026, antes de la campaña, que contiene los pesos del Agente 2 refinado. Sus fuentes se reunieron después a partir de respaldos del proyecto. Compiladas con el mismo compilador y las mismas opciones, reproducen byte a byte las secciones cargables de ese ejecutable, con su código, sus datos y sus pesos. El paquete complementario las entrega rotuladas como reconstrucción posterior. Como no se guardó la huella del programa instalado en la unidad de control, no puede demostrarse que fuera esa copia.

La campaña del control convencional es una sola sesión continua de cinco minutos. Un registro contiene la magnetización y el arranque y otro la inversión. El de la parada quedó incompleto, porque el registro circular dio la vuelta antes de guardarse, y no se usa. El registro de cierre recoge las tres capturas juntas, de modo que duplica el contenido de los anteriores y permite comprobarlos. Las seis capturas del osciloscopio se asignaron a su maniobra por su contenido, con la tensión del *DC link* y la frecuencia contrastadas con la unidad, y no por la fecha que llevan, que es la del reloj de cada instrumento.

La campaña del agente es también una sola sesión continua, de seis minutos. Un registro contiene la magnetización y el arranque. La inversión se registró dos veces durante la misma sesión; en esta memoria se presenta la segunda captura y ambas se conservan en el archivo experimental. El registro de cierre recoge las tres maniobras juntas, lo que permite comprobar las capturas anteriores. Las láminas de la magnetización proceden de ese registro de cierre, porque es el único que contiene la maniobra completa desde flujo nulo.

Las campañas anteriores del agente sobre la máquina no forman parte de esta memoria: se registraron antes de calibrar la escala del canal de velocidad que orienta el marco de referencia [2]. Se conservan íntegras, con sus huellas, en el archivo experimental.

Las seis capturas de osciloscopio de cada campaña sobre la máquina se conservaron completas. Para compararlas con los registros de la unidad de control, su inicio se alinea por separado con el de cada maniobra.

C.3. Material para reproducir y ampliar el trabajo

Junto con esta memoria se entrega un paquete complementario con un archivo de instrucciones y un manifiesto con la huella SHA–256 de cada archivo. Contiene los modelos de PLECS de la carga $R-L$ y de la máquina, los agentes con sus pesos y configuraciones, y los proyectos de Code Composer Studio del agente y del control convencional, con sus ejecutables archivados. También incluye los registros experimentales, los datos utilizados en los resultados y los guiones de las figuras y tablas. El archivo de instrucciones indica qué modelo, pesos, registros y guion corresponden a cada figura y tabla, y qué vínculos no se recuperaron. La Tabla C.5 reúne las versiones de los programas. Este material permite repetir las simulaciones, recompilar los controles y continuar el trabajo en otro computador.

Tabla C.5: Programas y versiones recuperadas. Las versiones de Python, Inkscape y MiKTeX son las instaladas al preparar el paquete.

Programa	Versión	Uso
MATLAB	R2025b en la reejecución de la evaluación; la del entrenamiento no se registró	entrenamiento y evaluación
PLECS Standalone	5.0.3; modelos guardados en el formato 5.0	simulación
Code Composer Studio	9.2.0	proyectos de la unidad de control
Compilador TI C6000	8.3.4, con `-mv6740 -O2 --define=c6748 --define=BENCH_FAST`	programa de la unidad
StarterWare C6748	1.20.04.01	bibliotecas de la unidad
Python	3.12.10, con NumPy 2.5.1, SciPy 1.18.0, Matplotlib 3.11.1 y PyMuPDF 1.28.2	métricas, figuras y comprobaciones
Inkscape	1.4.2	esquemas vectoriales
MiKTeX	26.5	memoria y rótulos de las figuras

La comprobación básica del paquete se ejecuta con Python y no requiere MATLAB, PLECS ni la unidad de control. Primero carga el Agente 2 refinado desde sus cabeceras y reproduce las 32 decisiones de referencia del procesador y una decisión archivada de la simulación. Después regenera desde los registros las Tablas 7.2, 7.3, 7.4 y B.2, que resultan idénticas a las de esta memoria. Las dos pruebas se superaron desde dos carpetas distintas.

No se recuperaron la huella del ejecutable cargado en la campaña del agente ni un registro independiente del estado eléctrico de la máquina de carga durante las campañas. El archivo de instrucciones declara, además, las figuras cuyo guion se identificó sólo por su contenido o por su fecha.

Durante los ensayos, la unidad de control registró las referencias, las corrientes en el marco dq, el flujo estimado, la velocidad y las modulantes. Los osciloscopios registraron las corrientes de fase, las tensiones v_{xN} y la tensión del *DC link*, con las frecuencias de adquisición indicadas en la sección anterior. Los archivos de las capturas de tensión conservan la configuración del instrumento. No se archivaron por separado los ajustes de las pinzas de corriente, pero la escala utilizada se comprobó mediante una medida cruzada con la unidad de control.

Los registros originales y sus huellas permiten identificar los datos empleados en cada resultado. De este modo, el material entregado constituye un punto de partida para verificar los ensayos, reproducir el análisis y desarrollar nuevas extensiones de esta memoria.

Apéndice D

Comprobaciones complementarias

Este apéndice reúne las comprobaciones que no pertenecen a los ensayos experimentales del Capítulo 8. Ninguna es una validación en el banco. La Tabla D.1 indica de qué cadena sale cada una y qué agente usa. El error y el sesgo se toman como en la memoria, $e = i^* - i$, de modo que son positivos cuando la corriente queda bajo su referencia.

Tabla D.1: Procedencia de las comprobaciones de este apéndice. El modelo en PLECS es el de la Figura 6.4 de la memoria, con sensores ideales. El gemelo del banco es ese modelo con el ruido de medida, la cuantización del *encoder* y los retardos del banco.

Comprobación	Cadena	Agente	Condiciones
Escalones de torque y rechazo de la perturbación	Simulación en PLECS	Agente 2 refinado, sin ajuste y con un ajuste previo de sensibilidad	+0,30 pu, 300 V, hasta el torque nominal
Agente 2 con las perturbaciones compensadas	Simulación en PLECS y gemelo del banco	Agente 2, entrenado sólo con la carga $R-L$, con sus pesos fijos; sin ganancia adaptativa ni compensación del retardo	±0,30 pu, 300 V, sin carga
Barrido de k_{adapt}	Gemelo del banco	Agente 2 refinado con ángulo al día	±0,30 pu, 284 V, sin carga
Umbral del primer paso grande	Cálculo con datos del banco	Agente 2 refinado con la configuración del Capítulo 7 y con ganancia adaptativa	Arranque a +0,30 pu

D.1. Escalones de torque y rechazo de la perturbación

El banco no dispone de una carga mecánica controlada ni de una medida de torque, de modo que los escalones de carga sólo se evaluaron en PLECS. La máquina se magnetiza, arranca a +0,30 pu en 0,7 s y recibe cuatro escalones de $0{,}25\,T_n$ en 1,5, 2,0, 2,5 y 3,0 s, que se retiran juntos en 3,5 s. El torque nominal es $T_n = 36{,}22$ N m.

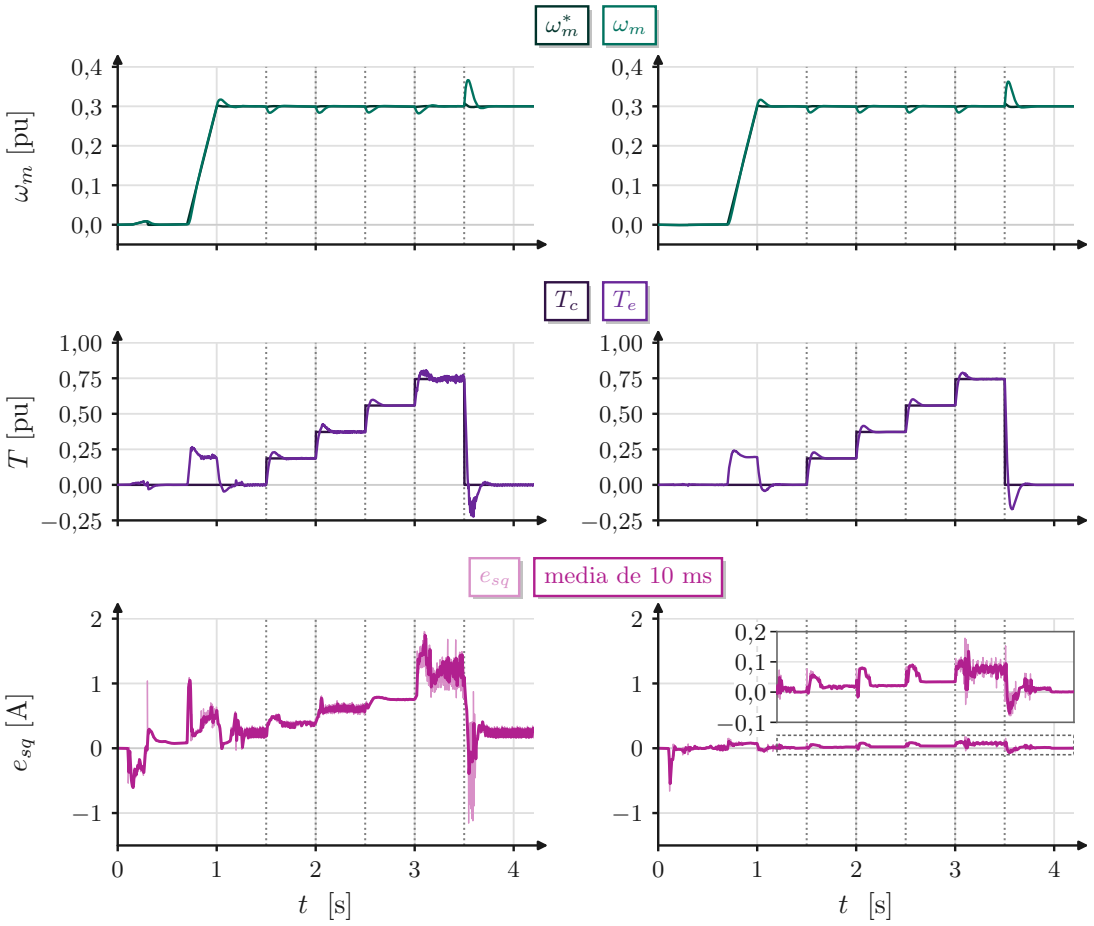
El torque de carga es una perturbación para el lazo de velocidad. La Tabla D.2 muestra que el Agente 2 refinado la rechazó en todos los escalones. La velocidad cayó entre 23,7 y 27,0 rpm y volvió a la banda de ±0,5 % en 124 a 218 ms. El torque electromagnético igualó al de carga con una diferencia menor que 0,2 %. La corriente máxima fue 14,8 A, sin saturar la modulante.

El sesgo de i_{sq} creció con la carga, hasta +1,20 A con T_n: la corriente quedó bajo su referencia. El lazo de velocidad lo absorbió elevando la referencia, de modo que la velocidad no quedó con error. La comprobación de sensibilidad conservó los pesos de la red y usó $g_d = 48$, $g_q = 16$, $k_d = 1/32$ y $k_q = 1/8$. Los productos fueron $P_d = 1{,}5$ y $P_q = 2$. Con ese ajuste, el sesgo quedó entre +0,02 y +0,08 A en los cuatro escalones. La corriente máxima fue 14,5 A y el flujo estimado no bajó de 1,00 pu.

Tabla D.2: Sesgo de la corriente de cuadratura en los cuatro escalones de torque, con el Agente 2 refinado sin ajuste y con el ajuste de sensibilidad mostrado en la presentación. Sesgo: media de $e_{sq} = i^*_{sq} - i_{sq}$ en los últimos 200 ms del escalón, positivo cuando la corriente queda bajo su referencia. Condiciones: PLECS con sensores ideales, *DC link* de 300 V y referencia de +0,30 pu.

Escalón	Sin ajuste [A]	Con ajuste de sensibilidad [A]
$0{,}25\,T_n$	+0,39	+0,02
$0{,}50\,T_n$	+0,61	+0,02
$0{,}75\,T_n$	+0,75	+0,03
$1{,}00\,T_n$	+1,20	+0,08

La Figura D.1 compara la maniobra completa en los dos casos que se usaron en la presentación. El ajuste redujo el error de cuadratura en los cuatro niveles de torque sin reentrenar la red, mientras la velocidad rechazó cada escalón. También bajó la desviación estándar del error. En los últimos 200 ms de cada escalón, pasó de valores entre 0,005 y 0,190 A a 0,016 A o menos. Esta prueba se limita a PLECS con sensores ideales y no corresponde a la configuración final ensayada en el banco.



(a) sin ajuste de sensibilidad (b) con ajuste de sensibilidad

Figura D.1: Escalones de torque simulados en PLECS con el Agente 2 refinado: (a) sin ajuste; (b) con el ajuste de sensibilidad que se mostró en la presentación. De arriba hacia abajo: velocidad con su referencia, torque electromagnético con el torque de carga y error de cuadratura $e_{sq} = i_{sq}^* - i_{sq}$. En el error, la línea clara es la señal de PLECS y la oscura, su media móvil de 10 ms; en (b), el recuadro amplía el error entre 1,2 y 4,2 s con su propia escala. Las líneas verticales marcan los cuatro escalones de torque y el retiro de la carga. Condiciones: las de la Tabla D.2; el origen de tiempos es el comienzo de la simulación. Escala común por fila.

D.2. El Agente 2 con las perturbaciones compensadas

El Agente 2 se entrenó sólo con la carga $R-L$, sin la perturbación de tensión que se agregó durante el refinamiento. Aplicado directamente a la máquina no la controla: el flujo estimado llega a 4,8 pu. La alternativa al refinamiento es quitar la perturbación fuera de la red. Una compensación suma a la tensión del agente la tensión inducida por el flujo de rotor, el término de flujo del eje directo y la diferencia entre el acoplamiento de la máquina y el que el agente aprendió en su carga. Con el flujo constante, el agente ve en cada eje

$$v_{d,\mathrm{ag}} = R_s\, i_{sd} + \sigma L_s \frac{\mathrm{d}i_{sd}}{\mathrm{d}t} \mp \omega_b\, \sigma L_s\, i_{sq}, \qquad v_{q,\mathrm{ag}} = R_s\, i_{sq} + \sigma L_s \frac{\mathrm{d}i_{sq}}{\mathrm{d}t} \pm \omega_b\, \sigma L_s\, i_{sd}, \quad \text{(D.1)}$$

donde $v_{d,\mathrm{ag}}$ y $v_{q,\mathrm{ag}}$ son las tensiones que pone el agente y $\omega_b = 2\pi \cdot 50$ rad/s es la frecuencia del marco de su entrenamiento. Los signos superiores corresponden al giro hacia adelante y los inferiores, al giro en sentido contrario. Es una carga $R-L$ con $L = \sigma L_s = 13{,}1$ mH, entre los dos valores de 10 y 20 mH usados en el entrenamiento.

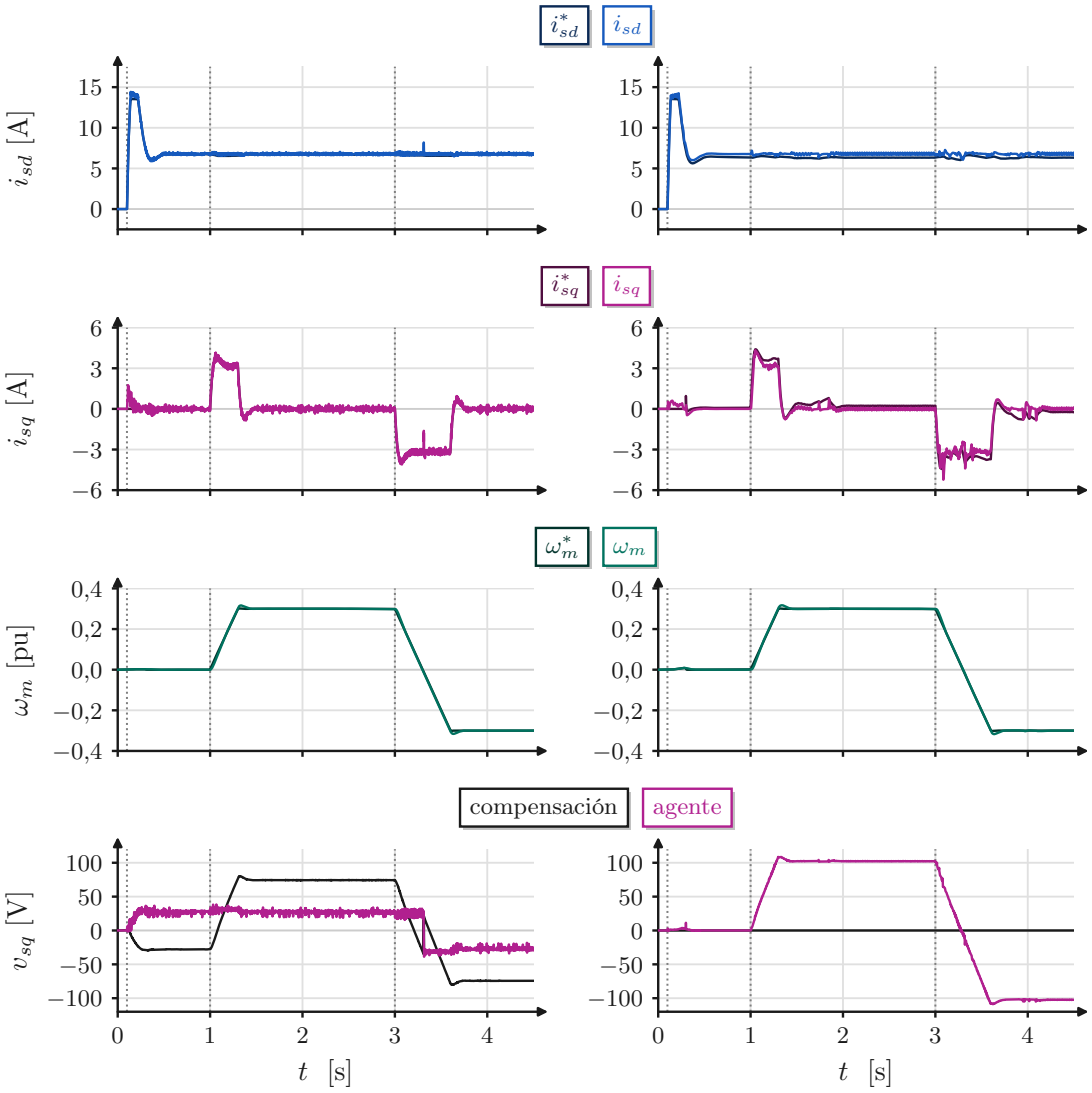
El Agente 2 compensado no es la configuración final. Tiene compensadas las perturbaciones, pero no se refinó con la ganancia adaptativa: sus errores observados se multiplican por 2 y su paso no se reduce, de modo que $k_{\mathrm{adapt}} = 1$ y $P = 2$. Tampoco usa la compensación del retardo. En PLECS, la orden de tensión se arma con medio período de adelanto.

La Tabla D.3 compara el Agente 2 compensado con el Agente 2 refinado, en PLECS y en el gemelo del banco. En PLECS, el sesgo del Agente 2 compensado fue de 0,06 A o menos, unas siete veces menor que el del Agente 2 refinado. En el gemelo esa ventaja desapareció: el sesgo subió a 0,23 A en el eje directo, y el del Agente 2 refinado bajó a 0,11 A. El punto débil es el paso por velocidad cero, donde el cambio del sentido de giro produjo errores de hasta 2,8 A.

Tabla D.3: El Agente 2 con las perturbaciones compensadas frente al Agente 2 refinado sin ajuste, en PLECS con sensores ideales y en el gemelo del banco. Sesgo: media de $e = i^* - i$ en las mesetas de +0,30 y −0,30 pu; RMS del error en el arranque, de 1,0 a 1,4 s, y en la inversión de marcha, de 3,0 a 3,7 s. Condiciones: *DC link* de 300 V, ±0,30 pu, sin carga, una semilla.

		PLECS		Gemelo del banco	
Indicador, d / q	Unidad	Agente 2 refinado	Agente 2 compensado	Agente 2 refinado	Agente 2 compensado
Sesgo a +0,30 pu	A	−0,44 / +0,23	−0,06 / −0,03	−0,11 / −0,02	−0,23 / −0,12
Sesgo a −0,30 pu	A	−0,44 / −0,23	−0,06 / +0,01	−0,04 / +0,11	−0,24 / +0,16
RMS en el arranque	A	0,36 / 0,42	0,20 / 0,13	0,32 / 0,47	0,43 / 0,39
RMS en la inversión	A	0,45 / 0,44	0,20 / 0,15	0,37 / 0,52	0,43 / 0,39
Máximo en la inversión	A	0,9 / 1,4	1,6 / 1,5	1,0 / 1,6	2,8 / 2,4

La Figura D.2 muestra la maniobra en PLECS. La compensación pone la mayor parte de la tensión del eje de cuadratura, 108,5 de 136 V a 0,40 pu, pero es de lazo abierto. El único que cierra el lazo de corriente es el agente, igual que el PI de corriente del control convencional con su prealimentación. El Agente 2 compensado no se energizó en el banco, y tampoco se simuló con los escalones de carga en este apéndice.



(a) Agente 2 compensado (b) Agente 2 refinado

Figura D.2: Maniobra simulada en PLECS: (a) Agente 2 con las perturbaciones compensadas; (b) Agente 2 refinado sin compensación. Corrientes con sus referencias, velocidad y tensión del eje de cuadratura, separada en la parte que pone la compensación y la que pone el agente. Las líneas punteadas marcan el comienzo de la magnetización, del arranque y de la inversión de marcha. Condiciones: las de la Tabla D.3, en PLECS; el origen de tiempos es el comienzo de la simulación. Escala común por fila.

D.3. Barrido de la ganancia adaptativa en el gemelo

El valor $k_{\text{adapt}} = 2$ del Capítulo 8 se eligió con el gemelo del banco. En el barrido, la ganancia del error vale k_{adapt} en los dos ejes y el paso se divide por k_{adapt}, de modo que $P = 1$. El gemelo usa el ruido de corriente calibrado contra el banco y, con $k_{\text{adapt}} = 1$, reproduce la desviación estándar del error que tuvo en el banco el Agente 2 refinado con la configuración del Capítulo 7, 0,30 A en el eje directo y 0,54 A en el de cuadratura.

La Tabla D.4 muestra el resultado. Con $k_{\text{adapt}} = 2$ el RMS del error de cuadratura

bajó de 0,498 a 0,401 A en el arranque y de 0,534 a 0,396 A en la inversión de marcha. Con 4, el arranque no mejoró y la inversión llegó a 0,363 A. Con 8, el arranque empeoró. El valor 2 es una elección equilibrada para este agente, no un óptimo general.

Tabla D.4: RMS del error de corriente en el gemelo del banco para cuatro valores de k_{adapt} con $P = 1$. Condiciones: ±0,30 pu, *DC link* de 284 V, sin carga, ruido de corriente de 0,055 A, ángulo al día, sin compensación del retardo.

k_{adapt}	1	2	4	8
RMS en el arranque, d / q [A]	0,262 / 0,498	0,258 / 0,401	0,250 / 0,404	0,299 / 0,423
RMS en la inversión, d / q [A]	0,351 / 0,534	0,247 / 0,396	0,268 / 0,363	0,264 / 0,401

D.4. Secuencia de cada período de control

El período de control T_s fija el intervalo entre dos decisiones consecutivas. El período de conmutación T_{sw} fija el ciclo de la portadora PWM. En este trabajo ambos valen 333,3 μs, pero corresponden a operaciones distintas. La unidad de control mide las corrientes en el mínimo de la portadora y ejecuta la rutina dentro del mismo ciclo. La modulante nueva llega a la PWM en el período siguiente. La Figura D.3 muestra esa secuencia y el lugar que ocupa el agente. Sus tiempos son los de la lectura tomada al terminar la campaña del Capítulo 7, que describe la Sección A.5. No son una cota de peor caso.

D.5. El umbral del primer paso grande

La política puede aplicar pasos finos desde el comienzo, pero no elige un paso grande hasta que el error observado alcanza un umbral e_{ag}. Si la referencia de cuadratura sube con pendiente S y el error inicial es e_0, la espera de ese paso es aproximadamente

$$t_{\text{espera}} \approx T_s + \frac{e_{\text{ag}}/g_q - e_0}{S} - t_{d,\text{ag}}, \qquad t_{\text{espera}} \geq 0, \tag{D.2}$$

donde g_q escala el error, $t_{d,\text{ag}}$ es la compensación y T_s es el período de atraso de la referencia. Con la configuración del Capítulo 7, la espera calculada es 6,1 ms y la medida, 6,67 ms.

Con $g_q = 2$, la espera calculada baja a unos 2,5 ms. En los registros finales hubo correcciones finas desde 0 o 0,67 ms, mientras el primer paso de nivel 5 o mayor apareció entre 1,33 y 2,00 ms. Este último instante no es la primera acción del agente ni se compara con el inicio de la orden del control convencional.

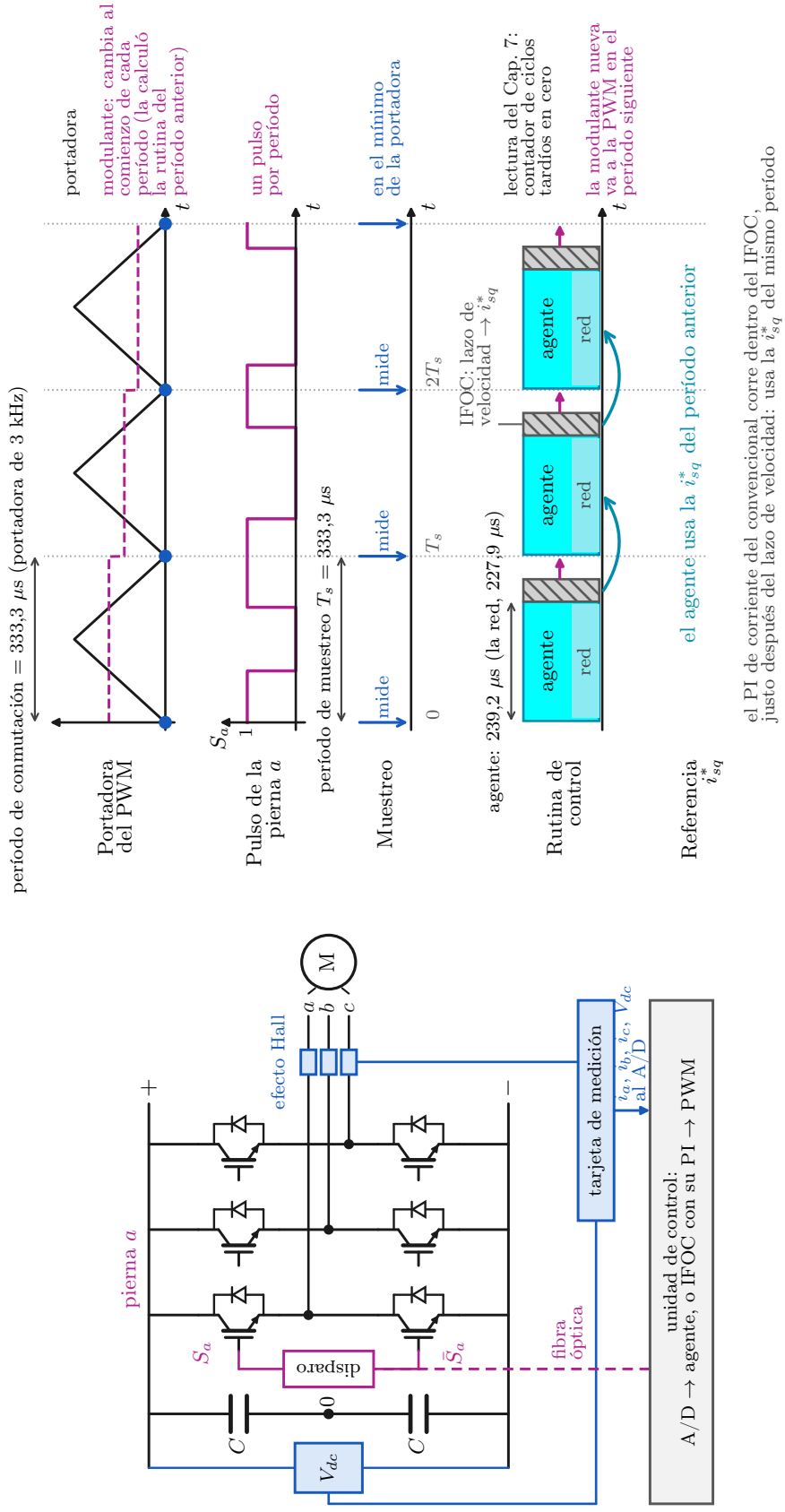


Figura D.3: Esquema del banco y de la secuencia de cada período de control: a la izquierda, el inversor con el *DC link*, los sensores de corriente y la unidad de control, con los interruptores S_a y $\bar{S}_a$ de la pierna a; a la derecha, la portadora triangular de 3 kHz, el pulso de esa pierna, el muestreo en el mínimo de la portadora y la rutina de control, con el agente antes que el IFOC. La flecha al final de cada rutina marca que la modulante nueva llega a la PWM en el período siguiente. Los tiempos del agente son los de la lectura tomada al terminar la campaña del Capítulo 7.

Bibliografía

[1] D. J. Barrios Bustamante, "Control digital de velocidad para un banco de pruebas," Memoria de Titulación para optar al título de Ingeniero Civil Electricista, Departamento de Electricidad, Universidad Técnica Federico Santa María, Valparaíso, Chile, abril 2007.

[2] A. Mora Castro, "Control vectorial de máquinas de inducción," Apuntes para la asignatura Accionamientos Eléctricos, Universidad Técnica Federico Santa María, Santiago, Chile, octubre 2021.

[3] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[5] D. G. Holmes and T. A. Lipo, *Pulse Width Modulation for Power Converters: Principles and Practice*, ser. IEEE Press Series on Power Engineering. Piscataway, NJ, USA: IEEE Press / Wiley-Interscience, 2003.

[6] J. Holtz, "Pulsewidth modulation for electronic power conversion," *Proceedings of the IEEE*, vol. 82, no. 8, pp. 1194–1214, 1994.

[7] A. M. Hava, R. J. Kerkman, and T. A. Lipo, "Simple analytical and graphical methods for carrier-based PWM-VSI drives," *IEEE Transactions on Power Electronics*, vol. 14, no. 1, pp. 49–61, 1999.

[8] R. P. Aguilera, P. Lezana, and D. E. Quevedo, "Finite-control-set model predictive control with improved steady-state performance," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 2, pp. 658–667, May 2013.

[9] M. Norambuena, P. Lezana, and J. Rodriguez, "A method to eliminate steady-state error of model predictive control in power electronics," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 7, no. 4, pp. 2525–2530, Dec 2019.

[10] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, 1992.

[11] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[12] B. K. Bose, "Neural network applications in power electronics and motor drives—an introduction and perspective," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 1, pp. 14–33, 2007.

[13] C. I. Cortés Barraza, "Aplicación de red neuronal artificial al control de un convertidor trifásico de condensadores flotantes," Memoria de titulación, Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Valparaíso, Chile, octubre 2024.

[14] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 15, 2011, pp. 315–323. [Online]. Available: https://proceedings.mlr.press/v15/glorot11a.html

[15] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 2094–2100.

[16] W. Ma, J. Li, H. Sheng, and Q. Song, "Optimizing multi-user dynamic spectrum access with double DQN and bivariate control," in *2025 8th International Conference on Information Communication and Signal Processing (ICICSP)*, 2025, pp. 896–902.

[17] M. Schenke, W. Kirchgässner, and O. Wallscheid, "Controller design for electrical drives by deep reinforcement learning: A proof of concept," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4650–4658, 2020.

[18] A. Traue, G. Book, W. Kirchgässner, and O. Wallscheid, "Toward a reinforcement learning environment toolbox for intelligent electric motor control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 3, pp. 919–928, 2022.

[19] M. Schenke and O. Wallscheid, "A deep Q-learning direct torque controller for permanent magnet synchronous motors," *IEEE Open Journal of the Industrial Electronics Society*, vol. 2, pp. 388–400, 2021.

[20] G. Book, A. Traue, P. Balakrishna, A. Brosch, M. Schenke, S. Hanke, W. Kirchgässner, and O. Wallscheid, "Transferring online reinforcement learning for electric motor control from simulation to real-world experiments," *IEEE Open Journal of Power Electronics*, vol. 2, pp. 187–201, 2021.

[21] M. J. Kaemling Arévalo, "Análisis y diseño de control basado en inteligencia artificial para aplicaciones de convertidores de potencia monofásicos y trifásicos," Memoria de Titulación, Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Valparaíso, Chile, marzo 2026.

[22] D. Gálvez, "Compensación de Rs usando mediciones de corriente de estator para una máquina de inducción controlada por PTC," Tesis de Magíster, Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Valparaíso, Chile, junio 2026.

[23] D. M. Reed, "Direct field-oriented control of an induction machine using an adaptive rotor resistance estimator," Master's thesis, The Pennsylvania State University, 2009.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 2015, arXiv:1412.6980.

[25] P. A. Lezana Illesca, "Manejo de la plataforma digital BRAIn para el control de accionamientos," Manual de la plataforma, Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Valparaíso, Chile, agosto 2022.

[26] Plexim GmbH, *PLECS User Manual*, Plexim GmbH, Zürich, Switzerland, 2024.

[27] F. Fröhr and F. Orttenburger, *Introducción al control electrónico.* Barcelona, España; Berlín y Múnich, Alemania: Marcombo / Siemens Aktiengesellschaft, 1986.

[28] J. Juliet Avilés, "Nociones sobre sistemas de tiempo discreto y aplicación en accionamientos eléctricos," Material de apoyo de Accionamientos Eléctricos I, ELI328, Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, 2022.

[29] P. A. Lezana Illesca, "Diseño de controladores digitales," Apéndice A, material de cátedra, Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Valparaíso, Chile, 2017.

[30] K. Ogata, *Modern Control Engineering*, 5th ed. Upper Saddle River, NJ, USA: Pearson, 2010.