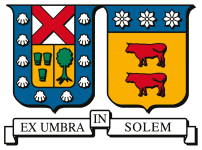

Universidad Técnica Federico Santa María
Departamento de Ingeniería Eléctrica
Santiago, Chile

Diseño de un sistema de control de energía para una planta de almacenamiento híbrida con tecnología PEM y control Grid Forming

Memoria de Grado presentada como requisito para optar
al título de Ingeniero Civil Electricista

Autor

Fabián Aguilera Otárola

Profesores

Johnny Rengifo Santana
Antonio Sánchez Squella

Octubre 2025

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: DISEÑO DE UN SISTEMA DE CONTROL DE ENERGÍA PARA UNA PLANTA DE ALMACENAMIENTO HÍBRIDA CON TECNOLOGÍA PEM Y CONTROL GRID FORMING

Nombre del candidato(a): FABIÁN HERNÁN AGUILERA OTÁROLA

Carrera / Grado: INGENIERÍA CIVIL ELÉCTRICA

Campus: SAN JOAQUÍN **Departamento:** INGENIERÍA ELÉCTRICA

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, JOHNNY WLADIMIR RENGIFO SANTANA, 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):

NO APLICA.

4.- FIRMAS

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

Fecha: 05/11/2025 **Firma:** _____

Estudiante o Candidato(a):

Fecha: 05.11.2025 **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.

Ancora imparo. —atr. a **Michelangelo Buonarroti**

Agradecimientos

El cierre de este trabajo, y con él el fin de mi etapa universitaria, no habría sido posible sin un grupo de personas que, de manera directa e indirecta, me acompañaron y aportaron en cada paso. En primer lugar, quiero agradecer a mis padres. Su crianza, su cariño y el esfuerzo cotidiano en su día a día me dieron la oportunidad real de llegar a la universidad. No me refiero exclusivamente a lo material, sino también al ejemplo que me enseñó constancia, rigor y dignidad para enfrentar cada día. Gracias por cada palabra de aliento cuando el ánimo flaqueaba y por recordarme que este trabajo es parte de un sueño que venimos construyendo hace años. Gracias también por la paciencia, por comprender mis ausencias y por estar siempre, incluso en los días difíciles. También quiero agradecer a mi hermana, pues sin su afecto y sentido del humor la elaboración de este trabajo hubiera sido un proceso mucho más tortuoso. Ahora te toca a ti seguir persiguiendo tu carrera universitaria; confío plenamente en que llegarás muy lejos y cuenta con que siempre tendrás mi apoyo en lo que te propongas. Quiero agradecer también a mis abuelos por su cariño y por animarme siempre a terminar mis estudios. Sé lo importante que es para ellos el haber cerrado esta etapa, y me honra profundamente poder compartirles esa alegría. En particular, quiero recordar a mi abuela paterna. Sé que le habría hecho muy feliz verme cerrar esta etapa; aunque la vida no nos dio esa oportunidad, sin lugar a dudas la llevo conmigo en este logro. También quiero agradecer a mis tíos. Cada vez que nos veíamos se interesaban por cómo me iba en la universidad, y ese cariño y preocupación me hizo sentir acompañado en una etapa que suele ser dura.

Me gustaría agradecer a mis amigos más entrañables por las risas, las conversaciones interminables y esos empujones anímicos que llegaron justo a tiempo. Su compañía hizo más livianos los días complicados y me ayudó a mantener el foco cuando las fuerzas flaqueaban. Gracias por estar, por celebrar cada pequeño avance y por recordarme, una y otra vez, por qué valía la pena terminar este trabajo. Finalmente, durante mis años en la universidad interactué con diversos profesores, funcionarias y funcionarios en clases, laboratorios y dependencias administrativas. Agradezco la disposición, la paciencia y la claridad en el ejercicio de la docencia, así como el tiempo dedicado a consultas y retroalimentación. Extiendo el reconocimiento al personal de apoyo —secretarías, biblioteca, laboratorios, aseo y mantención— por su trabajo constante, indispensable para el funcionamiento regular de la institución.

Índice general

Resumen

La creciente penetración de fuentes de energía renovable, inherentemente variables e intermitentes, plantea desafíos críticos para la estabilidad y flexibilidad de los sistemas eléctricos, principalmente debido a la consecuente reducción de la inercia rotacional efectiva y la disminución de las reservas de control primario. En este contexto, el presente trabajo de título aborda el diseño de un sistema de control de energía para una planta de almacenamiento híbrida. La planta está compuesta por un electrolizador, una celda de combustible y un banco de baterías, todos acoplados a la red mediante un convertidor formador de red. Esta configuración es especialmente atractiva, ya que aprovecha la complementariedad de las tecnologías: la batería proporciona una respuesta rápida frente a transitorios, la celda de combustible ofrece suministro continuo y el electrolizador permite la absorción de excedentes energéticos para la producción de hidrógeno. El convertidor formador de red, por su parte, otorga al sistema la capacidad de emular la inercia de una máquina sincrónica y participar activamente en el control de frecuencia.

Este trabajo propone un sistema de control de energía capaz de gestionar de forma coordinada los flujos de potencia de una planta de almacenamiento híbrida, con el doble propósito de contribuir al control de frecuencia de la red y maximizar la eficiencia energética del recurso primario. La metodología implementada se basó en el aprendizaje por refuerzo, utilizando el algoritmo gradiente de políticas determinista con redes profundas para entrenar un agente de control. La política de control se estructuró a través de una máquina de recompensas que guió el aprendizaje hacia acciones que optimizan la gestión del hidrógeno y respetan las restricciones operativas clave, tales como los límites de rampas de potencia y la disponibilidad constante del banco de baterías para amortiguar transitorios. Para acotar el costo computacional, el entrenamiento se llevó a cabo en un modelo promedio de la planta.

Los resultados demostraron que el entrenamiento del agente fue satisfactorio, alcanzando una convergencia estable de la política y validando la efectividad de la máquina de recompensas para imponer la lógica de control condicional. La contribución de la planta al control de frecuencia se materializó mediante la acción coordinada del convertidor formador de red y una estrategia adaptativa aplicada al control de la tensión del enlace de corriente directa. No obstante, se identificaron ciertos comportamientos del agente que sugieren oportunidades de mejora en la estabilidad de sus acciones y en la precisión del seguimiento de consignas. Finalmente, la validación en un modelo conmutado de mayor fidelidad sugirió que el uso del modelo promedio para el entrenamiento no constituyó una buena práctica, al observarse diferencias notables en el comportamiento del agente entre ambos entornos. A pesar de estas debilidades, el trabajo establece una integración armónica entre las estrategias de control y la gestión energética, clave para la incorporación efectiva de plantas de almacenamiento híbridas en sistemas con alta penetración renovable.

Palabras clave: convertidor formador de red, aprendizaje por refuerzo, DDPG, hidrógeno, electrolizador, celda de combustible, batería, control de frecuencia.

Capítulo 1

Introducción

La transición energética hacia fuentes renovables constituye una necesidad ineludible para reducir emisiones, mejorar la seguridad de suministro y permitir un crecimiento económico sostenido sin depender del consumo de combustibles fósiles. La rápida disminución de los costos asociados a la generación eólica y solar fotovoltaica ha impulsado su masificación, con factores de planta cada vez mayores y una expansión sostenida a gran escala. No obstante, estas tecnologías presentan un carácter inherentemente variable y parcialmente incierto, pues su producción depende de condiciones meteorológicas que sólo pueden predecirse dentro de un horizonte limitado y con márgenes de error apreciables. Como resultado, los sistemas eléctricos que avanzan hacia una alta participación renovable enfrentan nuevos desafíos de flexibilidad. Desde el punto de vista energético, se requiere capacidad de almacenamiento y de desplazamiento temporal de energía, mientras que desde el punto de vista del control se necesita mantener la frecuencia y la tensión dentro de márgenes aceptables frente a perturbaciones y variaciones rápidas en la inyección de potencia.

La creciente participación de generación basada en electrónica de potencia produce efectos sistémicos de gran relevancia. Desde el punto de vista energético, la naturaleza intermitente de las fuentes renovables, junto con la concentración de parques solares y eólicos en regiones con condiciones climáticas similares, provoca que extensas áreas del sistema presenten simultáneamente los mismos patrones de generación o de déficit, determinados por las variaciones meteorológicas dominantes. Esto da lugar a excedentes de energía en determinados momentos del día, como ocurre al mediodía solar, y a déficits en otros, especialmente durante las horas vespertinas. Esta variabilidad trae consigo importantes desafíos para la operación económica y aumenta la necesidad de contar con mecanismos que permitan desplazar energía en el tiempo. En lo referente al control, la sustitución progresiva de máquinas sincrónicas por convertidores conlleva una reducción de la inercia rotacional efectiva del sistema, una menor potencia de cortocircuito y una disminución de las reservas de control primario disponibles. Como consecuencia, el sistema eléctrico se vuelve más sensible a los desbalances entre potencia y frecuencia, presentando rampas netas más pronunciadas y eventos de frecuencia más frecuentes o de mayor magnitud si no se dispone de recursos capaces de proporcionar respuestas rápidas y servicios complementarios adecuados. Para preservar la estabilidad y la calidad del suministro eléctrico, resulta fundamental combinar estrategias de control avanzadas con tecnologías que puedan ofrecer regulación de frecuencia, amortiguamiento dinámico y soporte de tensión en intervalos de tiempo que van desde unos pocos milisegundos hasta varios minutos.

En este contexto, una planta de almacenamiento híbrida compuesta por un electrolizador, una celda de combustible y un sistema de baterías, todos acoplados a un enlace de corriente directa común e interconectados a la red mediante un convertidor formador de red, representa una alternativa especialmente atractiva para los sistemas eléctricos modernos. La principal razón radica en la complementariedad entre las tecnologías que la integran. El sistema de baterías ofrece una elevada eficiencia y una respuesta casi instantánea, lo que permite compensar desbalances transitorios de potencia y mantener estable la tensión del enlace de corriente directa. La celda de combustible, en cambio, entrega potencia de manera continua y controlada durante la operación estacionaria, siendo adecuada para cubrir requerimientos prolongados de suministro energético. Por su parte, el electrolizador posibilita la absorción de excedentes de generación renovable, transformándolos en hidrógeno y extendiendo así la autonomía energética del conjunto. El convertidor formador de red otorga al sistema la capacidad de regular la magnitud y la frecuencia de la tensión en sus terminales, emular

el comportamiento inercial de una máquina sincrónica y participar activamente en el control primario de frecuencia. De esta manera, se mitigan los efectos adversos asociados a la reducción de inercia del sistema eléctrico y se refuerza su resiliencia frente a perturbaciones. La sinergia entre la entrega rápida de potencia y la gestión coordinada de la energía, articulada mediante el control del convertidor formador de red, permite alcanzar una operación más robusta en escenarios de alta penetración renovable, en coherencia con las nuevas exigencias de estabilidad y flexibilidad del sistema.

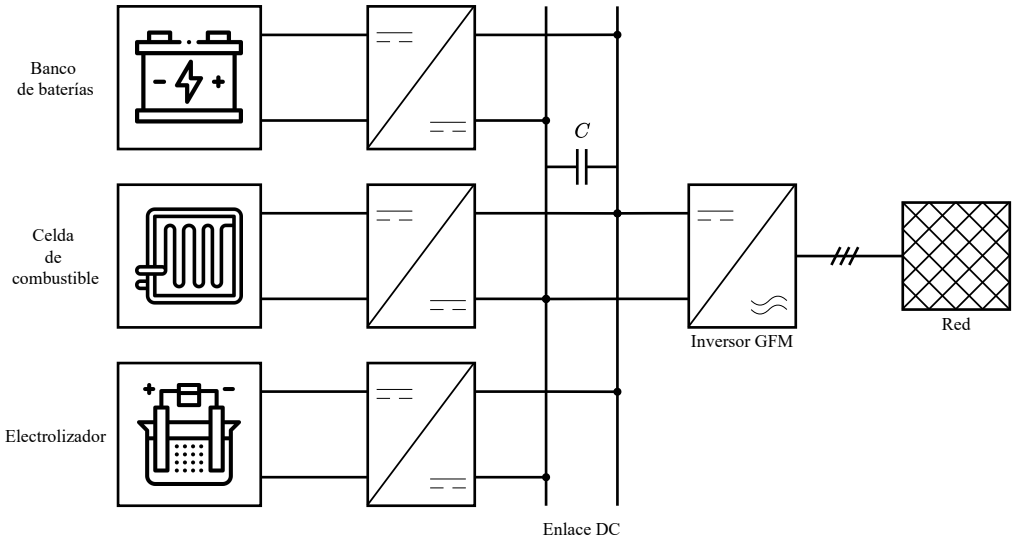

Figura 1.1: Esquema eléctrico simplificado de la planta de almacenamiento.

No obstante, el desempeño real de una planta de almacenamiento híbrida depende de manera crítica de la forma en que se gestionan sus intercambios energéticos. Cada uno de los subsistemas que la componen presenta características operativas particulares que condicionan la estrategia óptima de asignación de potencia a lo largo del tiempo. Las curvas de eficiencia del electrolizador y de la celda de combustible exhiben zonas de operación preferentes que no necesariamente coinciden con los perfiles de potencia instantánea requeridos por la red; mantener una operación prolongada fuera de dichas zonas puede traducirse en pérdidas energéticas acumuladas relevantes. A ello se suma que los tiempos de respuesta y las rampas de potencia admisibles difieren considerablemente entre tecnologías: el sistema de baterías puede modificar su potencia casi de forma instantánea, mientras que el electrolizador y la celda de combustible presentan dinámicas más lentas y restricciones más estrictas ante cambios bruscos. Forzar a estos últimos a seguir perturbaciones rápidas puede generar esfuerzos internos excesivos, aumentar el desgaste de los componentes y reducir su vida útil. Además, existen limitaciones de estado que deben considerarse, como el nivel de carga de la batería y la cantidad de hidrógeno almacenado, junto con los consumos asociados a los servicios auxiliares, que deben ser garantizados en todo momento. En conjunto, la coordinación adecuada del traspaso de potencia entre los distintos subsistemas, definiendo qué elemento entrega o absorbe energía, en qué magnitud y en qué momento, constituye no solo un problema de factibilidad eléctrica, sino también un desafío de gestión energética sujeto a restricciones físicas y dinámicas.

De lo expuesto se desprende la motivación central de este trabajo: diseñar un sistema de control de energía que, a través de un convertidor formador de red, gestione de manera coordinada los flujos de potencia entre los distintos subsistemas de la planta de almacenamiento híbrida, con el propósito de contribuir al control de frecuencia del sistema eléctrico y, al mismo tiempo, mejorar la gestión energética del recurso primario. El sistema de control propuesto debe ser capaz de aprovechar las ventajas propias de cada tecnología de almacenamiento según su escala temporal y su rango óptimo de operación, preservar los estados internos críticos, como el nivel de carga del sistema de baterías y las rampas de potencia, además de garantizar las condiciones de estabilidad de tensión tanto en el enlace de corriente directa como en la red. Esta integración armónica entre las estrategias de control y la gestión energética constituye un elemento clave para posibilitar la incorporación efectiva de plantas de almacenamiento híbridas en sistemas eléctricos con alta participación

de generación renovable, garantizando un desempeño estable y coherente con los requerimientos técnicos del sistema. En función de lo anterior, este trabajo plantea los siguientes objetivos que orientan su desarrollo y alcance.

Objetivo principal

- Diseñar un sistema de control de energía gestionando los intercambios energéticos de una planta de almacenamiento híbrida, contribuyendo en el control de frecuencia de la red maximizando la eficiencia energética.

Objetivos específicos

- Revisar el estado del arte relacionado con la planta de almacenamiento propuesta, abarcando los subsistemas que la componen y las estrategias de control asociadas.
- Determinar modelos que describan el comportamiento dinámico de la planta de almacenamiento, a partir de lo indicado por el estado del arte desarrollado a la fecha.
- Implementar un controlador de energía basado en aprendizaje por refuerzo, en base a métricas y recompensas alineadas con el objetivo de control.
- Analizar y evaluar cualitativamente los resultados obtenidos en las simulaciones, considerando las métricas utilizadas para la asignación de recompensa y el desempeño general del sistema de control desarrollado.

Con la finalidad de cumplir los objetivos propuestos, se revisó el estado del arte de la planta de almacenamiento y de sus estrategias de control; a partir de esta revisión se seleccionaron y parametrizaron modelos dinámicos pertinentes. Sobre dicha base se definió una política condicional orientada a optimizar la gestión del hidrógeno y a respetar las restricciones operativas, garantizando la disponibilidad del banco de baterías para amortiguar transitorios. Esta política se implementó en un agente de aprendizaje por refuerzo, entrenado con el algoritmo de gradiente de política determinista con redes profundas y guiado por una máquina de recompensas que encauza el aprendizaje hacia acciones coherentes con las consignas de operación. Para acotar el costo computacional, el entrenamiento se efectuó en un modelo promedio de la planta que preserva las dinámicas relevantes. Finalmente, se evaluó cualitativamente el desempeño del esquema tanto en el modelo promedio como en un modelo conmutado de mayor fidelidad, en términos del balance de potencia, seguimiento de consignas y respuesta transitoria. Una deinición importante del trabajo presentado es que, salvo indicación expresa, los parámetros y variables eléctricas se encuentran normalizados en por unidad.

El contenido que sigue a esta introducción se estructura en cinco capítulos principales. El Capítulo 2 presenta una revisión del estado del arte en la que se abordan los fundamentos teóricos y las estrategias de control asociadas al convertidor formador de red, junto con los modelos comúnmente empleados para representar el comportamiento dinámico del electrolizador, la celda de combustible y el sistema de baterías. En el Capítulo 3 se desarrolla el modelado de la planta de almacenamiento híbrida, describiendo los supuestos de representación, la arquitectura del control y las dinámicas características de cada componente. El Capítulo 4 introduce la metodología de control basada en aprendizaje por refuerzo, detallando la formulación del problema, la definición de la función de recompensa y la configuración del agente de control. En el Capítulo 5 se presentan y discuten los resultados obtenidos, analizando el desempeño del sistema propuesto bajo distintos escenarios de validación y contrastándolo con las métricas de referencia establecidas. Finalmente, en el Capítulo 6 se entregan las conclusiones generales del estudio y se proponen posibles líneas de desarrollo futuro.

Esta página se ha dejado intencionadamente en blanco.

Capítulo 2

Estado del arte

La transición energética ha impulsado el desarrollo de nuevas tecnologías y metodologías cuyo objetivo principal es minimizar el impacto ambiental de la generación eléctrica, manteniendo al mismo tiempo la operación segura y confiable de los sistemas eléctricos. A través de esta revisión del estado del arte, se establece el marco conceptual de las tecnologías y metodologías que permiten el desarrollo del presente trabajo, identificándose a su vez los avances y desafíos contemporáneos.

2.1. Convertidor formador de red

Los convertidores formadores de red (GFM, por sus siglas en inglés), concebidos originalmente para su uso en microrredes y sistemas aislados, recientemente se han identificado como una solución eficaz frente a los desafíos de estabilidad y resiliencia que plantea la alta penetración de generación basada en electrónica de potencia. Actualmente no existe una definición clara y universal de estos convertidores, pues a modo de ejemplo, no existe consenso sobre si la observación del ángulo de la tensión en el punto de conexión es crítica para decidir si un convertidor es GFM o no. Sin embargo, sus características y principios de operación los distinguen claramente de otras tecnologías descritas en la literatura académica. Estos convertidores tienen la capacidad de controlar la magnitud, frecuencia y ángulo de la tensión en sus terminales. Por ello, en su forma más básica, pueden representarse como fuentes de tensión controladas, asimilando su funcionamiento al de una máquina sincrónica convencional. Esta característica les permite participar activamente en el control de frecuencia del sistema eléctrico. Además, según la estrategia de control empleada, un convertidor GFM puede ofrecer capacidades de arranque en negro y suministrar inercia virtual al sistema, contribuyendo significativamente a mejorar su resiliencia y estabilidad [1].

En la práctica un convertidor formador de red es un convertidor estático convencional, cuya estrategia de control ha sido diseñada para proveer las características anteriormente mencionadas. Más aún, son variadas las estrategias de control existentes a la fecha, cada una con propiedades diferentes que serán atractivas en mayor o menor medida dependiendo de lo que se requiera. A continuación se describen las tres principales estrategias de control identificadas por la academia, desglosadas con mayor detalle en la Figura 2.1.

- *Droop Control*: Es una de las estrategias más utilizadas en convertidores GFM, con aplicaciones tanto en modo isla como en conexión a una red. La característica principal de este enfoque es la existencia de un compromiso lineal entre la frecuencia y tensión con la potencia activa y reactiva, respectivamente, dando lugar a un comportamiento similar al de los generadores sincrónicos en estado estacionario. Entre sus ventajas significativas se destacan su simplicidad inherente y la capacidad de aprovechar la rápida respuesta de los convertidores para estabilizar la red.

- *Virtual Machine*: Las estrategias de control basadas en máquina virtual buscan que los convertidores de potencia se comporten de manera similar a los generadores tradicionales, contribuyendo así a la estabilidad y regulación de la red eléctrica. Se fundamentan en modelar la dinámica mecánica y eléctrica de una máquina sincrónica o de inducción mediante algoritmos que simulan la presencia de inercia y reactancia interna, emulando el amortiguamiento de las máquinas tradicionales y el comportamiento

de las masas rotativas ante cambios de frecuencia, lo que ayuda a estabilizar la red y a suavizar la respuesta ante variaciones de carga o perturbaciones. El control basado en máquina virtual presenta desafíos, como la complejidad en la configuración de parámetros para lograr un comportamiento óptimo entre convertidores y la necesidad de alta capacidad de procesamiento en modelos complejos.

- *Virtual Oscillator Control*: Esta familia de controladores surge de los sistemas de osciladores acoplados, donde múltiples osciladores se sincronizan de forma natural mediante interacciones mutuas. Aplicado a los convertidores GFM, los controladores se diseñan para que cada convertidor actúe como un oscilador virtual autónomo, generando una señal de referencia sinusoidal y permitiendo que múltiples convertidores se sincronicen sin una referencia externa. Se ha probado que esta estrategia tiene el mismo comportamiento que *Droop Control* en estado estacionario, pese a que su estructura es completamente distinta. Una de las desventajas de este tipo de controladores es que adolecen de inercia virtual, lo que puede ser un desafío para mantener la estabilidad frente a significativos cambios de carga en redes grandes.

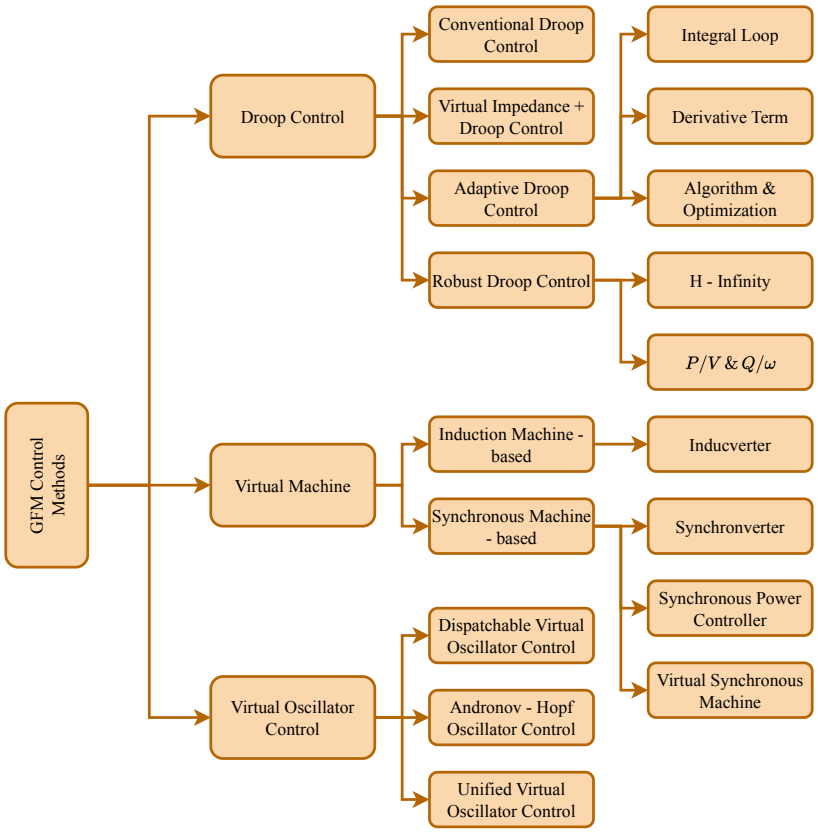

Figura 2.1: Estrategias de control GFM [1].

Pese a que existan diferentes familias de controladores para convertidores GFM, la estructura de control implementada es bastante similar entre estrategias. En la Figura 2.2 se presenta la estructura genérica de control para un convertidor GFM, identificando un lazo interno y otro externo. El lazo interno se encarga de regular la tensión a la salida del filtro, utilizando como actuación la tensión en terminales del convertidor. El diseño de este lazo de control está estrechamente ligado al tipo de filtro que se utilice a la salida del convertidor, cuya elección dependerá de las exigencias del operador de red en torno al contenido armónico permisible. Como norma general, los filtros LCL satisfacen de mejor manera estos requerimientos frente a otros tipos, como los filtros L y LC, logrando una buena atenuación a bajas frecuencias de conmutación [2].

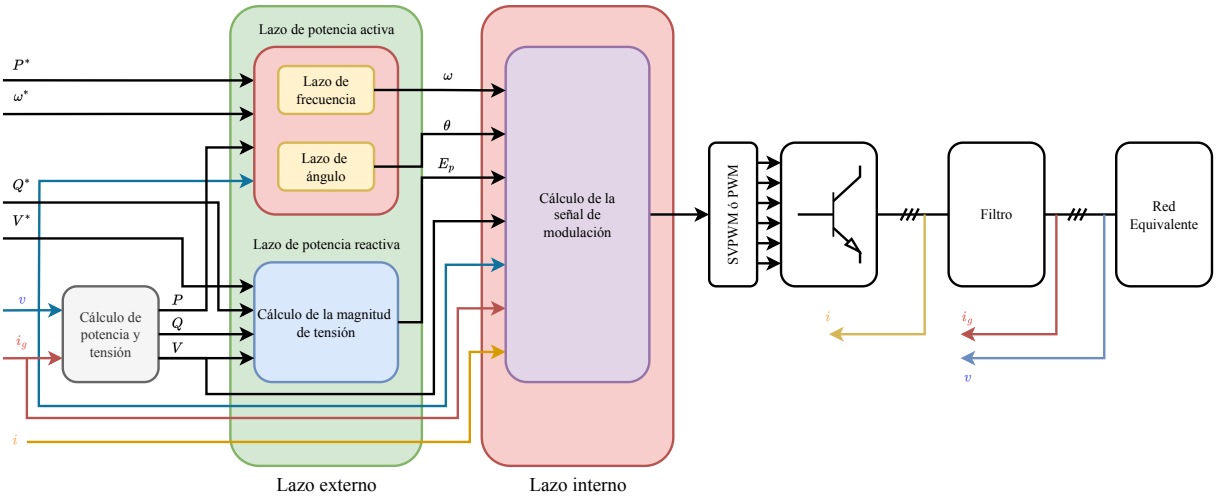

Figura 2.2: Estructura general de control GFM [3].

Por otra parte, el lazo externo de control se encarga de calcular la referencia de tensión a la salida del filtro, dadas sendas consignas de potencia activa y reactiva que se deseen intercambiar con la red equivalente. Es usual que el control de potencia activa se desacople del de potencia reactiva, puesto que, respectivamente, estas pueden ser modificadas con el ángulo y magnitud de tensión relativas entre la tensión a la salida del filtro y la impuesta por la red.

Generalmente los convertidores GFM se sincronizan con la red en función de su potencia activa a la salida, de manera similar a como lo hacen las máquinas sincrónicas convencionales. Sin perjuicio a lo anterior, algunas estrategias consideran el uso de un lazo de seguimiento de fase (PLL, por sus siglas en inglés) para lograr la sincronización con la red [4]. Es deseable evitar el uso de PLL en la estrategia de control, ya que de ser el caso los lazos internos de control se sincronizan con la tensión en el punto de conexión, siendo esta última altamente sensible a las variaciones de corriente a la salida del convertidor en redes con bajo niveles de cortocircuito (SCR, por sus siglas en inglés). Por otra parte, en redes rígidas con alto SCR los convertidores GFM tienden a perder el sincronismo con la red, ya que un leve cambio en la diferencia de fase entre el convertidor y los voltajes de la red puede llevar a grandes variaciones de potencia activa. De esta manera, se requiere un control con amortiguamiento robusto para operar en un amplio rango de condiciones de SCR [3].

Dado que los convertidores GFM operan de forma similar a las máquinas síncronas, un cortocircuito en la red puede exponerlos a una alta circulación de corriente, lo que compromete la integridad de sus semiconductores. Para enfrentar este desafío, la solución más simple consiste en cambiar de forma transitoria a un control por seguimiento de la red (GFL, por sus siglas en inglés), permitiendo limitar la corriente del convertidor sin perder el sincronismo con la red. Esta estrategia requiere observar instantáneamente el ángulo de la red a la cual está conectado el convertidor. En sistemas aislados, otro enfoque común es implementar saturadores en el lazo de control interno para las señales de referencia de corriente, lo cual permite limitar la corriente del convertidor. Sin embargo, esta estrategia puede fallar si no se toman medidas contra efectos indeseados como el "wind-up.º el "latch-up". Otra solución es introducir el concepto de impedancia virtual, el cual consiste en limitar la referencia de tensión del convertidor usando una impedancia ficticia variable en el lazo de control interno. Esto evita la generación de señales de referencia de corriente excesivamente altas a la entrada del lazo de control de corriente [3].

2.2. Electrolizador

Tal como se mencionó anteriormente, los efectos del cambio climático han forzado la búsqueda de alternativas energéticas que permitan continuar con el funcionamiento de los procesos industriales, minimizando a la vez su huella ecológica. Se ha identificado al hidrógeno como un potencial candidato a resolver esta problemática, siempre que este se produzca mediante técnicas de bajo impacto ambiental. En la actualidad, el hidrógeno se produce principalmente mediante el reformado de gas natural por vapor. Esta técnica implica la reacción de un hidrocarburo, particularmente metano, con vapor de agua a alta temperatura para producir hidrógeno, dióxido de carbono y monóxido de carbono. El proceso productivo descrito tiene la ventaja de ser económico, pero por otra parte, la liberación de dióxido de carbono es perjudicial para el medio ambiente. Sin embargo, existen alternativas sustentables que permiten hacer frente a la situación descrita como la electrólisis, la cual consiste en separar el agua en sus componentes aplicando una diferencia de potencial entre dos electrodos insertos en un electrolito.

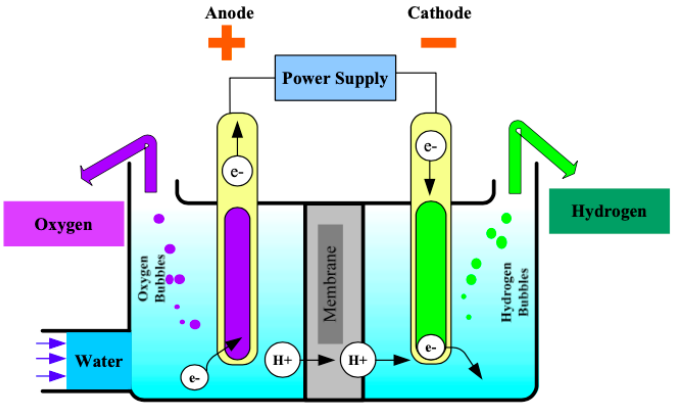

Figura 2.3: Electrólisis del agua [5].

Este proceso es endotérmico, por lo que requiere de energía externa para la activación de su reacción asociada. En la electrólisis, los electrones fluyen desde el ánodo hacia el cátodo a través del circuito externo. En el cátodo, los electrones participan en la reducción de los iones de hidrógeno para formar hidrógeno gaseoso, mientras que en el ánodo, el agua se oxida, liberando oxígeno y electrones. Las reacciones químicas involucradas en el proceso descrito son presentadas en (2.1). Dentro de las ventajas de esta técnica se encuentra la ausencia de gases invernaderos residuales y la alta pureza del hidrógeno obtenido [6, 7].

$$\begin{aligned} \text{Ánodo:} \quad & \mathrm{H_2O} \rightarrow \frac{1}{2}\mathrm{O_2} + 2\mathrm{H^+} + 2\mathrm{e^-} \\ \text{Cátodo:} \quad & 2\mathrm{H^+} + 2\mathrm{e^-} \rightarrow \mathrm{H_2} \\ \text{Reacción neta:} \quad & \mathrm{H_2O} \rightarrow \mathrm{H_2} + \frac{1}{2}\mathrm{O_2} \end{aligned} \tag{2.1}$$

El dispositivo físico en el que se lleva a cabo el proceso de electrólisis se denomina electrolizador, compuesto por una o varias celdas electrolíticas conectadas en serie o en paralelo, dependiendo de las restricciones de tensión y corriente específicas de cada celda. En términos generales, cada celda está compuesta por un ánodo, un cátodo, un medio electrolito y una fuente de electricidad externa que suministra la energía necesaria para realizar la reacción. La literatura clasifica los electrolizadores en tres tipos principales: alcalinos, de membrana de electrolito polimérico y de óxidos sólidos. Es importante señalar que, independientemente del tipo de celda, el principio de operación es equivalente para todos [7].

Los electrolizadores de membrana de electrolito polimérico (PEM, por sus siglas en inglés) son dispositivos que producen hidrógeno de alta pureza y presentan tiempos de respuesta rápidos, con un tiempo de arranque en frío inferior a 15 minutos. Estos electrolizadores utilizan una membrana de Nafion como electrolito sólido, que permite el transporte de iones de hidrógeno H^+ a través de la membrana. Los materiales comúnmente empleados en su construcción incluyen IrO_2 y RuO_2 en el ánodo, y platino Pt o paladio Pd en el cátodo,

todos ellos catalizadores preciosos. Entre sus principales ventajas se encuentran su compacidad, diseño simple y capacidad para producir hidrógeno de alta pureza. Sin embargo, estos electrolizadores también presentan desventajas, tales como el elevado costo debido al uso de catalizadores caros, menor durabilidad en comparación con otros tipos de electrolizadores y la naturaleza ácida del medio, lo que contribuye al desgaste de los componentes [6, 7].

Los electrolizadores de agua alcalinos (AEL, por sus siglas en inglés) son comúnmente utilizados para la producción de hidrógeno a gran escala debido a su alta estabilidad y bajo costo. Utilizan un electrolito de hidróxido de potasio en una concentración del 30–40 % y funcionan a temperaturas de entre 50 y 90 °C a presiones de 2 a 10 bar. En su construcción, se emplean aleaciones de níquel y óxidos de cobalto en los electrodos de ánodo y cátodo, respectivamente. Entre sus ventajas destacan la capacidad de operar a bajas temperaturas y el no uso de catalizadores costosos para la activación de las reacciones. Sin embargo, su tiempo de respuesta es inferior al de los electrolizadores de membrana de electrolito polimérico, lo que implica que los AEL son menos adecuados para aplicaciones que requieren una rápida activación. Además, presentan desventajas relacionadas con la permeación de gases y la corrosión de los electrodos debido a la naturaleza corrosiva del electrolito alcalino, lo cual representa un desafío importante en términos de durabilidad [6, 7].

Los electrolizadores de óxido sólido (SOEC, por sus siglas en inglés) operan a temperaturas considerablemente más altas que los electrolizadores alcalinos y los de membrana de electrolito polimérico, en un rango de 900 a 1000 °C. Utilizan electrodos de níquel y un electrolito cerámico sólido, lo que permite reducir la dependencia de la electricidad en el proceso de separación de hidrógeno, ya que gran parte de la energía necesaria se puede suministrar en forma de calor. Fuentes de calor adicionales, como el calor residual o energía nuclear, pueden ser aprovechadas para disminuir aún más el consumo eléctrico. A medida que aumenta la temperatura de operación, la eficiencia del SOEC también mejora; sin embargo, estas temperaturas elevadas aceleran la degradación del electrolito y reducen la vida útil del sistema. En comparación con los electrolizadores PEM, los SOEC presentan tiempos de respuesta más lentos y un tiempo de arranque en frío superior a los 60 minutos, lo cual los hace menos adecuados para aplicaciones que requieren rápida activación. Adicionalmente, las demostraciones actuales de los SOEC solo operan a nivel de kilovatios, limitando su uso en aplicaciones de mayor escala [6, 7]. De acuerdo con los antecedentes presentados, el tipo de electrolizador que mejor se ajusta a los requerimientos del caso de estudio es el de membrana de electrolito polimérico, considerando que es el que posee el tiempo de respuesta más rápido entre las alternativas.

El comportamiento de un electrolizador PEM puede ser modelado de diferentes formas, identificándose modelos electroquímicos, eléctricos, térmicos, de transferencia de masa, y de dinámica de fluidos [7]. Para el caso de estudio son de interés los modelos eléctricos y electroquímicos, particularmente los dinámicos, puesto que estos caracterizan instantáneamente el comportamiento eléctrico del electrolizador además de la producción de hidrógeno en función de la corriente circulante. Los fundamentos del modelo se indican a continuación: la tensión en los terminales de una celda electrolítica corresponde a la suma de las diversas caídas de tensión asociadas a los procesos reversibles e irreversibles que ocurren durante su funcionamiento. Idealmente, la electrólisis del agua es un proceso reversible, lo que implica que, con una cantidad adecuada de energía, las moléculas de agua pueden separarse en oxígeno e hidrógeno y, de manera inversa, al recombinarse oxígeno e hidrógeno, es posible generar agua y energía. En este contexto, la caída de tensión relacionada con la reacción de división del agua se denomina tensión reversible E_{rev}. Por otro lado, las caídas de tensión asociadas a los procesos irreversibles incluyen la iniciación de la reacción electroquímica, que implica el movimiento de electrones hacia y desde los colectores de corriente; la superación de las barreras resistivas que dificultan el movimiento de electrones a través de distintos caminos eléctricos, además del transporte de protones de hidrógeno cargados positivamente a través de la membrana; y la oposición del flujo de masa de los reactantes y productos a través de la membrana. Este último efecto es significativo bajo condiciones de alta densidad de corriente y presión elevada. Estos procesos irreversibles generan caídas de tensión conocidas como sobretensión de activación, sobretensión óhmica y sobretensión de difusión, respectivamente [8].

La tensión reversible es la caída de tensión asociada a la reacción reversible y es responsable de la producción de hidrógeno. Esta tensión depende de las variaciones en la temperatura y presión de operación de la celda. Sin embargo, en los modelos de circuitos equivalentes, se asume que es constante y se modela mediante una fuente de tensión continua. Este supuesto es válido pues el modelo térmico del electrolizador tiene una constante de tiempo en el rango de horas [9]. La cantidad de potencia eléctrica convertida en hidrógeno puede calcularse simplemente como el producto de la tensión reversible y la corriente eléctrica que pasa a través de los terminales de la celda [8].

La sobretensión óhmica representa la pérdida de energía causada por las barreras resistivas que dificultan el movimiento de los portadores de carga eléctrica a través de los diferentes caminos de la celda de electrólisis. Por lo tanto, la resistencia óhmica total es la suma de las resistencias óhmicas asociadas a los colectores de corriente, la membrana electrolítica y las resistencias de contacto entre estas placas. Para efectos prácticos, la componente eléctrica de la resistencia generalmente se descarta, y solo se considera la resistencia de la membrana electrolítica. De esta manera, la caída de tensión óhmica se emula mediante una resistencia en serie a la tensión reversible que modela la resistencia eléctrica de la membrana [8].

Desde una perspectiva eléctrica, la sobretensión de activación está destinada a emular dos acciones principales. La primera es la influencia de las barreras resistivas que deben superarse para liberar electrones de las superficies de los electrodos cargados, lo que se denomina resistencia de transferencia de carga. La segunda es para considerar el tiempo perdido debido a la acumulación de carga en el ánodo y el cátodo, que varía según la corriente de la celda. Por lo tanto, se necesita un elemento resistivo para emular la energía consumida por la reacción al liberar los electrones desde las superficies de los electrodos, así como un elemento capacitivo para emular la acumulación de carga. Para este propósito puede utilizarse una rama RC en paralelo, y esta puede repetirse tanto para el ánodo como el cátodo, pues ambos tienen diferentes velocidades de reacción. Sin embargo, la literatura indica que dependiendo de la densidad de corriente que atraviese a las celdas electrolíticas, el enfoque cambia de acuerdo a lo indicado a continuación [8].

- A bajas densidades de corriente, el efecto de la sobretensión de difusión es despreciado y pueden utilizarse dos ramas RC conectadas en serie para caracterizar la sobretensión de activación en el ánodo y el cátodo por separado, tal como se muestra en la Figura 2.4(a).

- A densidades de corriente moderadas, el efecto de la sobretensión de activación se combina con el de la sobretensión de difusión, ya sea mediante una fuente de tensión tal como se muestra en la Figura 2.4(b), o bien, como una única rama RC que caracteriza el comportamiento del ánodo y el cátodo en conjunto, como se muestra en la Figura 2.4(c).

- A altas densidades de corriente, la sobretensión de activación se modela mediante una resistencia conectada en serie al circuito de la sobretensión de difusión, para posteriormente conectarse ambos en paralelo con un capacitor que modela el efecto de doble capa en la celda electrolítica, como se muestra en la Figura 2.4(e) y 2.4(f).

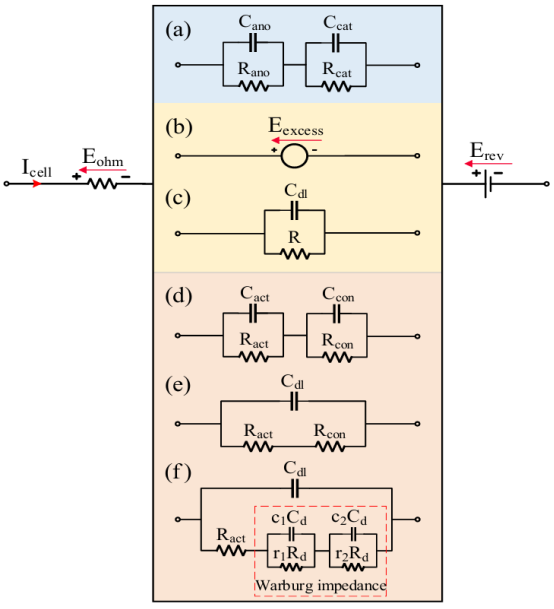

Figura 2.4: Enfoques para el modelo de un electrolizador PEM [8].

La capacitancia de doble capa es un fenómeno electroquímico que ocurre en la interfaz entre un electrodo y un electrolito, donde las cargas se acumulan formando una capa doble, una en la superficie del electrodo y otra en el electrolito cercano. Experimentalmente, se ha observado en muchos casos que los efectos de la sobretensión de activación y de difusión suelen estar relacionados entre sí, así como con el efecto de capacitancia de doble capa. Más aún, la sobretensión de activación y de difusión son derivadas de la misma ecuación general en el enfoque de modelado físico, haciendo referencia a la ecuación de Butler - Volmer. De forma similar a su contraparte de activación, la sobretensión de difusión se modela de diferentes formas según la densidad de corriente que atraviesa a las celdas electrolíticas [8].

- A bajas densidades de corriente, la sobretensión de difusión puede ignorarse pues su efecto es despreciable frente al resto de caídas de tensión. En este caso, sólo se toma en cuenta la sobretensión de activación tal como se muestra en la Figura 2.4(a).
- A densidades de corriente moderadas, el efecto de la sobretensión de difusión se combina con el de la sobretensión de activación, modelándose mediante un término de tensión adicional, como se muestra en la Figura 2.4(b).
- A altas densidades de corriente, la sobretensión de difusión puede modelarse mediante una resistencia en serie con la resistencia de la sobretensión de activación en una sola rama RC, tal como se muestra en la Figura 2.4(e). Una segunda opción es mediante una rama RC separada, conectada en serie con la rama RC de la sobretensión de activación de acuerdo con la Figura 2.4(d). El último enfoque supone utilizar dos ramas RC separadas para el ánodo y el cátodo, conectadas en serie entre sí y con la resistencia de activación en una rama RC más grande, según se muestra en la Figura 2.4(f).

Los parámetros de todos los modelos presentados se ajustan en función de la curva de polarización y de las formas de onda experimentales obtenidas de un electrolizador bajo inyecciones controladas de corriente. Por lo tanto, la elección del modelo dependerá directamente de qué tan bien este se ajuste a los resultados experimentales. Otro aspecto importante a considerar es que la mayoría de los modelos identificados en la literatura, así como los revisados previamente, están diseñados para caracterizar el rendimiento de una única celda. Es común que los autores asuman que el comportamiento del conjunto de celdas puede estimarse multiplicando los parámetros de una celda por el número de celdas conectadas en serie y en paralelo, sin embargo, esta no es una estimación particularmente precisa [8].

En relación con la producción de hidrógeno, aplicando balances de materia a una celda de electrólisis se puede demostrar que el hidrógeno producido es proporcional a la corriente que circula por ella. Formalmente esta relación es conocida como Ley de Faraday de la Electrólisis, en la que el flujo molar de hidrógeno a la salida de una celda depende de su corriente y de la constante de Faraday F [5].

$$\dot{n}_{\mathrm{H_2,out}} = \frac{I_{\mathrm{cell}}}{2F} \tag{2.2}$$

La eficiencia energética parcial de una celda de electrólisis, es decir, el cociente entre la potencia convertida en hidrógeno y la potencia eléctrica consumida en terminales, puede ser expresada en función de la tensión en terminales de la celda y su tensión reversible E_{rev} tal como se hace en (2.3).

$$\eta_{\mathrm{EL,p}} = \frac{E_{\mathrm{rev}}}{V_{\mathrm{cell}}} \tag{2.3}$$

La eficiencia energética global de una celda de electrólisis, es decir, el cociente entre la potencia convertida en hidrógeno y la potencia eléctrica consumida incorporando servicios auxiliares, queda dada por (2.4).

$$\eta_{\mathrm{EL,g}} = \frac{\eta_{\mathrm{EL,s}}\, P_{\mathrm{EL}}}{P_{\mathrm{EL}} + P_{\mathrm{Aux,EL}}} \tag{2.4}$$

2.3. Celda de combustible

Una celda de combustible es un dispositivo electroquímico capaz de convertir la energía química de un determinado combustible en energía eléctrica. Este tipo de tecnología ofrece ventajas significativas en comparación con los motores de combustión tradicionales, pues posee una eficiencia energética superior y sus emisiones son mucho menos contaminantes. El único subproducto del proceso en celdas de hidrógeno es agua, lo que elimina las emisiones de dióxido de carbono y otros contaminantes convencionales. Además, las celdas de combustible operan de manera silenciosa debido a que prácticamente no poseen partes móviles [10].

La composición de una celda de combustible basada en hidrógeno es esencialmente la misma que la de un electrolizador, contando con un ánodo, un electrolito y un cátodo. Sin embargo, su funcionamiento es inverso al del electrolizador: en el ánodo, el hidrógeno se oxida, generando cationes y electrones libres. Los cationes atraviesan el electrolito hacia el cátodo, mientras que los electrones fluyen a través de un circuito externo. En el cátodo, el oxígeno se reduce y reacciona con los iones de hidrógeno y los electrones para formar agua. Las reacciones químicas del proceso se muestran en (2.5), y adicionalmente en la Figura 2.5 se presenta gráficamente el principio de funcionamiento de una celda de combustible tipo membrana de electrolito polimérico.

$$\begin{aligned} \text{Ánodo:} \quad & \mathrm{H_2} \rightarrow 2\mathrm{H^+} + 2\mathrm{e^-} \\ \text{Cátodo:} \quad & \frac{1}{2}\mathrm{O_2} + 2\mathrm{H^+} + 2\mathrm{e^-} \rightarrow \mathrm{H_2O} \\ \text{Reacción neta:} \quad & 2\mathrm{H_2} + \mathrm{O_2} \rightarrow 2\mathrm{H_2O} \end{aligned} \tag{2.5}$$

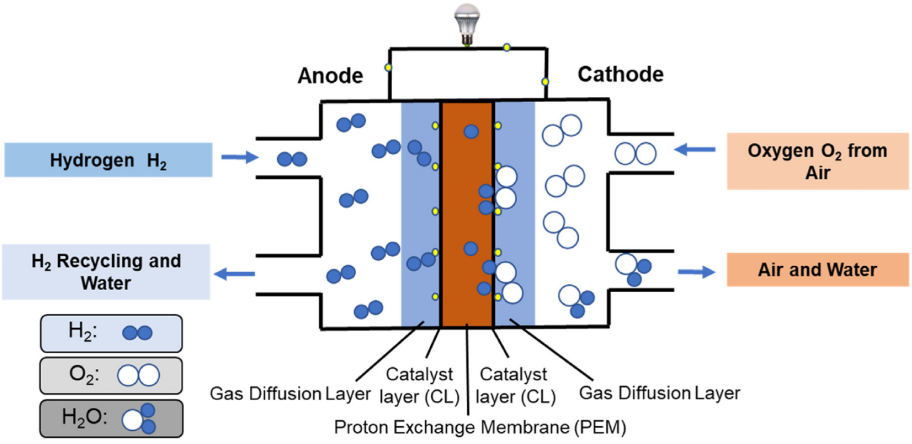

Figura 2.5: Funcionamiento de una celda de combustible tipo PEM [11].

Los diversos tipos de celdas de combustible presentan características y aplicaciones específicas según su diseño y temperatura de operación. Las celdas de combustible de membrana de electrolito polimérico operan a temperaturas inferiores a 120 °C y destacan por su rápida respuesta ante cambios de carga, lo que las hace ideales para aplicaciones como transporte, generación distribuida y respaldo de energía, aunque enfrentan desafíos relacionados con su alto costo y sensibilidad a impurezas en el combustible. Las celdas alcalinas funcionan a temperaturas menores de 100 °C y se caracterizan por su bajo costo de componentes, pero son sensibles al dióxido de carbono presente en el combustible y el aire. Las celdas de ácido fosfórico operan a temperaturas intermedias, entre 150 y 200 °C, y son adecuadas para cogeneración de calor y energía, aunque presentan tiempos largos de arranque y sensibilidad al azufre. Las celdas de carbonatos fundidos trabajan a temperaturas elevadas, entre 600 y 700 °C, y son eficientes y flexibles en el uso de combustibles, siendo ideales para servicios eléctricos y generación distribuida, pero enfrentan desafíos de corrosión y baja densidad de potencia. Finalmente, las celdas de óxidos sólidos operan en un rango de temperaturas que va desde los 500 hasta los 1000 °C, ofreciendo alta eficiencia y flexibilidad en el uso de combustibles, siendo útiles para aplicaciones auxiliares y de generación distribuida, aunque presentan limitaciones en el número de ciclos de arranque y parada debido a la corrosión a alta temperatura [12]. De acuerdo con los antecedentes

presentados, el tipo de celda de combustible que mejor se ajusta a los requerimientos del caso de estudio es el de membrana de electrolito polimérico, considerando que es el que posee la respuesta más rápida ante cambios de carga. Un antecedente interesante es que si se compara el tiempo de respuesta de un electrolizador y una celda de combustible, ambos de tipo PEM, el primero tiene una respuesta más lenta que el segundo frente a cambios en la carga [5].

En la literatura se identifican tres enfoques principales para modelar celdas de combustible: los modelos químicos, experimentales y eléctricos. Los modelos químicos se centran en describir fenómenos complejos dentro de la celda, como el transporte de masa, la transferencia de calor y la difusión de especies químicas, lo que los hace altamente detallados. Sin embargo, su implementación requiere una cantidad considerable de parámetros, lo que dificulta su integración en programas de simulación eléctrica. Por su parte, los modelos experimentales se basan en datos empíricos obtenidos directamente de pruebas realizadas sobre las celdas de combustible y representan su comportamiento mediante tablas de consulta o expresiones empíricas. A pesar de su simplicidad, no incorporan la termodinámica de las celdas ni los efectos de variables operativas, como la presión de entrada, el flujo de gases, su composición o la temperatura. Finalmente, los modelos eléctricos representan a la celda mediante circuitos eléctricos equivalentes, siendo especialmente útiles para la simulación de sistemas de generación energética. Aunque no consideran la termodinámica del sistema, resultan más prácticos para aplicaciones eléctricas. En cualquier enfoque, los parámetros del modelo son obtenidos mediante pruebas experimentales en celdas reales [13].

Para efectos del presente trabajo, se utilizará un modelo del tipo eléctrico considerando que la celda de combustible es parte de un sistema de almacenamiento híbrido, el cual se encuentra conectado a un sistema eléctrico de potencia. Este modelo se compone de una resistencia interna R_Ω y una fuente de tensión controlada E, que depende de la corriente de la celda de combustible tal como se muestra en la Figura 2.6.

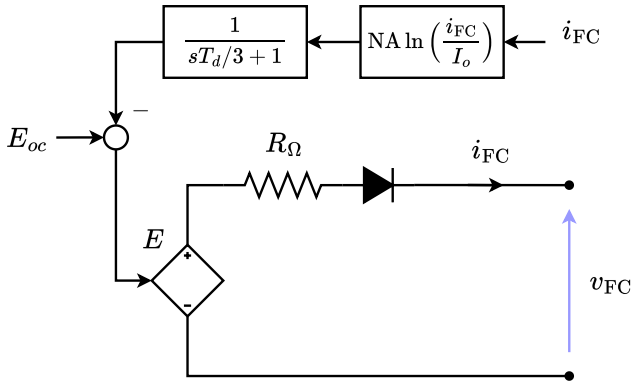

Figura 2.6: Circuito equivalente de celda de combustible [13].

El valor de la fuente de tensión controlada, de acuerdo con la Figura 2.6, puede obtenerse según

$$E = E_{oc} - \left(\frac{1}{\frac{T_d}{3}s + 1}\right) \mathrm{NA}\ \ln\left(\frac{i_{\mathrm{FC}}}{I_o}\right) \tag{2.6}$$

Por otra parte, la tensión a la salida de la celda de combustible v_{FC} es

$$v_{\mathrm{FC}} = E - R_\Omega\, i_{\mathrm{FC}} \tag{2.7}$$

A continuación se detallan los parámetros y variables del modelo.

- E_{oc}: Tensión de circuito abierto en V.
- N: Número de celdas en serie.
- A: Pendiente de Tafel en V.
- I_o: Corriente de intercambio en A.

- T_d: Tiempo de respuesta en s.
- R_Ω: Resistencia interna en Ω.
- v_{FC}: Tensión instantánea en terminales en V.
- i_{FC}: Corriente instantánea en terminales en A.

La Ecuación (2.6) representa la tensión del conjunto de celdas de combustible considerando solamente las pérdidas de activación, asociadas al tiempo que toma concretar las reacciones químicas en la superficie de los electrodos. Por otra parte, la Ecuación (2.7) considera las pérdidas óhmicas en los electrodos y el electrolito por medio de la resistencia R_Ω. Es importante mencionar que este modelo es válido siempre que las celdas operen en condiciones nominales de tensión y temperatura [13].

El flujo molar de hidrógeno consumido es calculado a partir de la Ley de Faraday, tal como se indica en (2.8) donde F corresponde a la constante de Faraday referida en la sección anterior y N_{FC} al número de celdas internas que componen la celda de combustible [14].

$$\dot{n}_{\mathrm{H_2,in}} = \frac{N_{\mathrm{FC}}\, i_{\mathrm{FC}}}{2F}, \tag{2.8}$$

La eficiencia energética parcial de una celda de combustible es calculada como el cociente entre la potencia eléctrica a la salida del equipo y la potencia calorífica del combustible consumido. Esta última puede ser estimada por medio del producto entre el Valor Calorífico Inferior del Hidrógeno (LHV, por sus siglas en inglés) y el flujo molar de hidrógeno consumido $\dot{n}_{\mathrm{H_2,in}}$ [15].

$$\eta_{\mathrm{FC,p}} = \frac{v_{\mathrm{FC}}\, i_{\mathrm{FC}}}{\dot{n}_{\mathrm{H_2,in}}\, \mathrm{LHV}} \tag{2.9}$$

La eficiencia energética global de una celda de combustible, es decir, el cociente entre la potencia eléctrica entregada y la potencia calorífica consumida incorporando servicios auxiliares, queda dada por (2.10).

$$\eta_{\mathrm{FC,g}} = \frac{P_{\mathrm{FC}}}{\frac{P_{\mathrm{FC}}}{\eta_{\mathrm{FC,p}}} + P_{\mathrm{Aux,FC}}} \tag{2.10}$$

2.4. Batería

Un sistema de almacenamiento de energía en baterías (BESS, por sus siglas en inglés) es una tecnología electroquímica diseñada para almacenar energía en forma química y liberarla como energía eléctrica. Estos sistemas, al integrarse en redes eléctricas, permiten almacenar el excedente de energía generado por fuentes renovables intermitentes, como la solar y la eólica, para su uso en momentos de alta demanda o baja generación. Además, contribuyen a mejorar la estabilidad de la red al proporcionar servicios auxiliares, tales como el control de tensión, control de frecuencia, compensación de potencia reactiva y la reducción de pérdidas en la transmisión. Los sistemas BESS están compuestos por baterías que abarcan diversas tecnologías, como las de ion-litio, níquel-cadmio, plomo-ácido y flujo *redox*, entre otras. Cada tecnología presenta características específicas en términos de capacidad, densidad energética, vida útil y costos, lo que determina su idoneidad para distintas aplicaciones dentro del sistema eléctrico.

Las baterías de plomo-ácido, con una vida útil de entre 300 y 3000 ciclos y una eficiencia del 70 % al 90 %, son económicas y fáciles de conseguir. Sin embargo, tienen baja densidad energética, entre 35 y 40 Wh/L, lo que limita su uso a aplicaciones como iluminación de emergencia, motores eléctricos y submarinos diésel-eléctricos. Además, su capacidad para soportar ciclos repetidos es limitada, y su impacto ambiental es significativo. Las baterías de níquel-cadmio ofrecen un ciclo de vida más largo, alcanzando los 3000 ciclos, con una eficiencia del 80 % y una densidad energética de 40 a 60 Wh/L. Su desempeño a bajas temperaturas es bueno y toleran condiciones exigentes, pero tienen una alta tasa de autodescarga, un impacto ambiental elevado y el conocido efecto memoria. Las baterías de níquel-hidruro metálico presentan una vida útil de aproximadamente 2000 ciclos, con una eficiencia de entre 66 % y 92 % y una densidad energética mayor, de 60 a 120 Wh/L. Comparadas con las de níquel-cadmio, tienen mejor desempeño en bajas temperaturas y son

más eficientes. No obstante, su alto costo y la posibilidad de dañarse si se descargan completamente limitan su uso a aplicaciones específicas, como sistemas recargables avanzados. Por último, las baterías de ion-litio son las más avanzadas, con un ciclo de vida de hasta 3000 ciclos, una eficiencia de entre 75 % y 90 % y una densidad energética que varía de 100 a 265 Wh/L. Destacan por su alta eficiencia, baja tasa de autodescarga y rápida respuesta, aunque su costo inicial es elevado y su seguridad depende del diseño. Estas baterías son esenciales en dispositivos portátiles, vehículos eléctricos y otras aplicaciones que requieren alto rendimiento energético [16].

De acuerdo con los antecedentes presentados, el tipo de sistema que mejor se ajusta a los requerimientos del caso de estudio es aquel que se compone de baterías de ion-litio, considerando su bajo tiempo de respuesta y alta eficiencia. Cada batería se conforma por cuatro componentes principales: el cátodo, el ánodo, el separador y el electrolito. El cátodo y el ánodo actúan como portadores de carga, siendo responsables del almacenamiento y liberación de energía en la batería. El separador es una barrera física que evita cortocircuitos internos al separar los electrodos, permitiendo al mismo tiempo el flujo de iones de litio a través de sus poros. Por último, el electrolito es el medio que transporta los iones, incluyendo los iones de litio, entre el cátodo y el ánodo durante los procesos de carga y descarga. El funcionamiento de estas baterías se basa en el movimiento de los iones de litio entre los electrodos. Durante el proceso de carga, los iones de litio son liberados desde el cátodo y se difunden a través del electrolito hasta el ánodo, donde se adosan. Al mismo tiempo, los electrones fluyen en dirección opuesta a través de un circuito externo para mantener la neutralidad electroquímica. Durante la descarga, este proceso se invierte: los iones de litio se trasladan desde el ánodo de regreso al cátodo a través del separador, mientras que los electrones fluyen nuevamente a través del circuito externo, suministrando energía al dispositivo conectado. Las reacciones químicas involucradas tanto en la carga como en la descarga se presentan en (2.11) [17].

$$\begin{aligned} \text{Cátodo:} \quad & \mathrm{LiCoO_2} \leftrightarrow \mathrm{Li_{(1-x)}CoO_2} + x\mathrm{Li^+} + xe^- \\ \text{Ánodo:} \quad & 6\mathrm{C} + x\mathrm{Li^+} + xe^- \leftrightarrow \mathrm{Li}_x\mathrm{C_6} \\ \text{Reacción global:} \quad & \mathrm{LiCoO_2} + 6\mathrm{C} \leftrightarrow \mathrm{Li_{(1-x)}CoO_2} + \mathrm{Li}_x\mathrm{C_6} \end{aligned} \tag{2.11}$$

En la literatura se identifican tres tipos de modelos de baterías: empíricos, electroquímicos y basados en circuitos eléctricos. Los modelos empíricos y electroquímicos no son adecuados para representar las dinámicas de las celdas con el propósito de estimar el estado de carga (SOC, por sus siglas en inglés). Sin embargo, los modelos basados en circuitos eléctricos pueden ser útiles para representar las características eléctricas de las baterías [18]. Considerando la naturaleza del caso de estudio, se hará uso de un modelo basado en circuitos eléctricos el cual está constituido por una fuente de tensión controlada y una resistencia, tal como se muestra en la Figura 2.7.

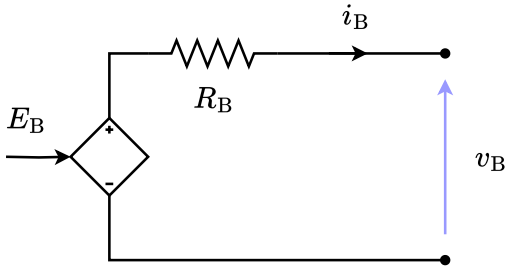

Figura 2.7: Circuito equivalente de batería [18].

La fuente de tensión controlada E_B depende de la corriente i_B y la capacidad extraída, de acuerdo con la expresión presentada en (2.12) para una batería de ion-litio.

$$E_\mathrm{B} = \begin{cases} E_0 - K\frac{Q}{Q-it}\cdot(it + {i_\mathrm{B}}^*) + A\exp(-B\cdot it), & \text{si } i_\mathrm{B} > 0 \quad \text{(Descarga)} \\ E_0 - K\frac{Q}{it-0{,}1\cdot Q}\cdot {i_\mathrm{B}}^* - K\frac{Q}{Q-it}\cdot it + A\exp(-B\cdot it), & \text{si } i_\mathrm{B} < 0 \quad \text{(Carga)} \end{cases} \tag{2.12}$$

Los parámetros y variables del modelo se detallan a continuación.

- E_0: Tensión constante en V.
- K: Constante de polarización en V/Ah, o resistencia de polarización en Ω.
- ${i_\mathrm{B}}^*$: Corriente de baja frecuencia en A.
- i_B: Corriente de la batería en A.
- it: Capacidad extraída en Ah.
- Q: Capacidad máxima de la batería en Ah.
- A: Tensión exponencial en V.
- B: Capacidad exponencial en Ah^{-1}.

La capacidad extraída it de la batería se calcula según lo indicado en (2.13). Por otro lado, la corriente de baja frecuencia se obtiene conforme a lo definido en (2.14), donde T_B representa el tiempo de respuesta de la batería al 95 % del valor final. Finalmente, el estado de carga se determina de acuerdo con lo establecido en (2.15).

$$it = \int_0^t i_\mathrm{B}(t)\,dt. \tag{2.13}$$

$$i_\mathrm{B}(s)^* = \frac{1}{\frac{T_B}{3}\,s + 1}\,i_\mathrm{B}(s) \tag{2.14}$$

$$\mathrm{SOC}_{\,\%} = 100\left(1 - \frac{1}{Q}\int_0^t i_\mathrm{B}(t)\,dt\right) \tag{2.15}$$

El modelo presenta ciertos supuestos que deben considerarse al interpretar sus resultados. Se asume que la resistencia interna de la batería permanece constante durante los ciclos de carga y descarga, independientemente de la amplitud de la corriente. Además, los parámetros del modelo se deducen únicamente a partir de las características de descarga y se consideran equivalentes para los procesos de carga. También se descarta el efecto Peukert, suponiendo que la capacidad de la batería no varía con la amplitud de la corriente. Asimismo, el modelo no tiene en cuenta los efectos de la temperatura ni representa el fenómeno de autodescarga de la batería. Por último, se asume que la batería carece de efecto memoria. Estos supuestos permiten simplificar el análisis, aunque pueden limitar la precisión del modelo en condiciones reales [18].

De acuerdo con la literatura, para modelar la eficiencia del sistema de baterías debe considerarse el ciclo de trabajo bajo el que se opera, es decir, si se encuentra en régimen de carga o descarga. Si ha lugar el régimen de carga, la eficiencia queda dada por η_cha que corresponde al cociente entre la potencia electroquímica P_B y aquella absorbida desde la red. Por otra parte, si ha lugar el régimen de descarga, la eficiencia queda dada por η_dis que corresponde al cociente entre la potencia vertida hacia la red y la potencia electroquímica [19].

$$\eta_\mathrm{cha} = \frac{P_\mathrm{B}}{v_\mathrm{B}\,i_\mathrm{B}} \qquad\qquad \eta_\mathrm{dis} = \frac{v_\mathrm{B}\,i_\mathrm{B}}{P_\mathrm{B}} \tag{2.16}$$

La energía libre de Gibbs es un concepto termodinámico que representa la energía química interna almacenada en baterías, dependiente del potencial reversible del sistema. Sin embargo, su cálculo experimental es complejo debido a que es sumamente difícil medir con precisión dicho potencial bajo condiciones dinámicas de operación. Por esta razón, se hace uso de la potencia electroquímica, que utiliza la tensión en circuito abierto E_B como una aproximación práctica del potencial reversible, permitiendo un enfoque más accesible y adaptado a las capacidades experimentales y de simulación. De esta forma, la potencia electroquímica del sistema se calcula según lo indicado por (2.17) [19].

$$P_\mathrm{B} = E_\mathrm{B}(\mathrm{SOC}_{\,\%})\,i_\mathrm{B} \tag{2.17}$$

2.5. Aprendizaje por refuerzo

El aprendizaje por refuerzo (RL, por sus siglas en inglés) constituye un enfoque del aprendizaje automático centrado en la interacción de un agente con su entorno con el propósito de alcanzar un objetivo a través de la maximización de una señal de recompensa acumulada. A diferencia de otros paradigmas tradicionales, como el aprendizaje supervisado o no supervisado, el aprendizaje por refuerzo se caracteriza por un proceso de descubrimiento autónomo en el que el agente no recibe instrucciones explícitas sobre qué acciones tomar. En su lugar, debe aprender a actuar de manera óptima mediante la experiencia directa, evaluando las consecuencias de sus decisiones a lo largo del tiempo por medio de una señal de recompensa. De esta manera, es claro advertir que esta forma de aprendizaje está profundamente inspirada en la manera en que los seres vivos adquieren conocimiento a través de la exploración activa de su entorno. Dos características fundamentales distinguen al aprendizaje por refuerzo de otros enfoques. La primera es el aprendizaje por prueba y error, es decir, la necesidad de explorar distintas acciones para evaluar sus efectos. La segunda es la presencia de recompensas diferidas, ya que las consecuencias de una acción no siempre se reflejan inmediatamente, sino que pueden influir en el curso futuro de la interacción. Estas propiedades introducen una complejidad significativa, ya que el agente debe aprender a considerar no solo los beneficios inmediatos, sino también los efectos a largo plazo de sus decisiones [20]. Un problema típico de RL se modela como un proceso de decisión de Markov, definido por la tupla $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$ donde

- $\mathcal{S}$ es el conjunto de estados posibles del entorno.
- $\mathcal{A}$ es el conjunto de acciones que el agente puede ejecutar.
- $\mathcal{P}(s', r \mid s, a)$ es la función de transición de estados, que define la probabilidad de pasar el del estado s' al ejecutar la acción a en el estado s.
- $\mathcal{R}$ es el conjunto de recompensas inmediatas alcanzadas por el agente a causa de tomar la acción a en el estado s.

El objetivo del agente es aprender una política óptima que maximice las recompensas futuras. Una política debe entenderse como la forma en que el agente interactúa con el entorno dado un estado determinado. Formalmente son definidas por una función μ que toma una acción a dado el estado s. Por otra parte, el concepto de recompensa futura debe comprenderse como la suma de recompensas que el agente recibirá en los próximos pasos luego del instante t. Formalmente queda definida por la expresión presentada en (2.18).

$$G_t = r_{t+1} + r_{t+2} + r_{t+3} + \cdots + r_T \tag{2.18}$$

Este enfoque no distingue entre recompensas a corto y largo plazo, lo cual resulta problemático por dos razones: por un lado, la expresión diverge cuando $T \to \infty$; por otro, podría favorecer políticas que no presentan urgencia por obtener recompensas [21]. Para abordar esta situación, se introduce el concepto de recompensa futura descontada, el cual incorpora una penalización progresiva sobre las recompensas a medida que se alejan en el tiempo. Este concepto se expresa formalmente mediante (2.19), donde la penalización está determinada por la tasa de descuento $\gamma \in [0, 1)$.

$$G_t = \sum_{k=0}^{\infty} \gamma^k \, r_{t+k+1} \tag{2.19}$$

Tal como se mencionó anteriormente, el aprendizaje por refuerzo es esencialmente un proceso de prueba y error. En este contexto, el agente enfrenta constantemente el dilema entre explotar el conocimiento actual para maximizar recompensas inmediatas y explorar nuevas acciones que podrían conducir a mejores resultados en el futuro. Un exceso de explotación puede llevar a soluciones subóptimas al ignorar alternativas potencialmente superiores, mientras que una exploración excesiva puede ralentizar el aprendizaje e incluso llevarlo a no ser exitoso. Lograr un equilibrio adecuado entre ambos comportamientos es fundamental para el éxito del proceso de aprendizaje [21].

Previamente se mencionó que uno de los objetivos fundamentales del aprendizaje por refuerzo es desarrollar agentes capaces de maximizar recompensas acumuladas a lo largo del tiempo. Para ello, se han

propuesto diferentes estrategias de aprendizaje que pueden clasificarse en tres grandes categorías: métodos basados en políticas, métodos basados en funciones de valor y métodos actor - crítico, que constituyen un enfoque híbrido.

Los métodos basados en políticas consisten en aprender directamente una política π que asigna una acción futura, de forma estocástica o determinista, a cada posible par estado–acción. El aprendizaje se enfoca en ajustar los parámetros de esta política para maximizar el retorno esperado, sin necesidad de evaluar explícitamente el valor de cada acción. Este enfoque resulta especialmente adecuado para problemas con espacios de acción continuos o cuando se requieren políticas estocásticas. Por ejemplo, si un agente está aprendiendo a andar en bicicleta, un método basado en política implicaría aprender directamente qué acción debe ejecutar al detectar cierta inclinación, con el objetivo de mantenerse en equilibrio [21].

Por otra parte, los métodos basados en funciones de valor buscan estimar el valor esperado de cada acción en cada estado, mediante funciones como $V(s)$ o $Q(s, a)$, que representan la utilidad esperada de un estado o de una acción tomada en ese estado, respectivamente. A partir de estas estimaciones, se puede derivar una política óptima seleccionando la acción con mayor valor estimado. Esta aproximación es eficiente en entornos con espacios de acción discretos y ha sido la base de algoritmos fundamentales como Q-learning o DQN. Siguiendo el mismo ejemplo, un agente que aprende a andar en bicicleta mediante un enfoque basado en valores asignaría un puntaje a cada combinación de inclinación y acción, y seleccionaría aquella que tenga mejor evaluación [21].

Finalmente, los métodos actor - crítico combinan las ventajas de los dos enfoques anteriores. En este esquema, el actor representa la política y decide las acciones a ejecutar, mientras que el crítico estima una función de valor que evalúa el desempeño del actor. Al aprovechar tanto la eficiencia del aprendizaje basado en valores como la flexibilidad de las políticas parametrizadas, los métodos actor-crítico se han consolidado como una de las estrategias más efectivas para resolver problemas con espacios de acción continuos, dinámicas complejas o entornos parcialmente observables. Ejemplos representativos de este enfoque son los algoritmos DDPG, TD3, SAC, A2C y PPO.

De forma transversal a las categorías previamente descritas, los algoritmos de aprendizaje también pueden clasificarse según el uso que hacen de un modelo del entorno. Los métodos que no incorporan un modelo aprenden exclusivamente a partir de la experiencia directa, sin representar explícitamente las dinámicas del sistema. Aunque suelen requerir un mayor volumen de datos, presentan una mayor robustez frente a errores de modelado. En contraste, los métodos que sí hacen uso de un modelo del entorno – o que aprenden una aproximación de su comportamiento – lo emplean para simular trayectorias y planificar acciones antes de ejecutarlas. Esta capacidad puede reducir significativamente la cantidad de interacciones necesarias, aunque su eficacia depende críticamente de la fidelidad del modelo utilizado.

Tabla 2.1: Clasificación de algoritmos de aprendizaje por refuerzo.

Algoritmo	**Entorno**	**Modelo**	**Enfoque**	**Referencia**
Q-Learning	Discreto	No requiere	Función de valor	[22]
DQN	Discreto	No requiere	Función de valor	[23]
REINFORCE	Ambos	No requiere	Basado en política	[24]
DDPG	Continuo	No requiere	Actor - crítico	[25]
TD3	Continuo	No requiere	Actor - crítico	[26]
SAC	Continuo	No requiere	Actor - crítico	[27]
A2C	Ambos	No requiere	Actor - crítico	[28]
PPO	Ambos	No requiere	Actor - crítico	[29]
Dyna-Q	Discreto	Requiere	Función de valor	[30]
MBPO	Continuo	Requiere	Actor - crítico	[31]
PETs	Continuo	Requiere	Actor - crítico	[32]
PlaNet	Continuo	Requiere	Basado en política	[33]
Dreamer	Continuo	Requiere	Actor - crítico	[34]

Existen numerosos algoritmos de aprendizaje por refuerzo desarrollados en la literatura, cada uno con características particulares orientadas a distintos tipos de entornos, estructuras de política, requerimientos de modelo o estrategias de exploración. En la Tabla 2.1 se presenta una selección representativa de los algoritmos más influyentes y utilizados en la actualidad, agrupados según criterios como el tipo de entorno, la necesidad de un modelo del sistema y el enfoque de aprendizaje adoptado. Con el objetivo de acotar el alcance del presente trabajo, se ha decidido utilizar el algoritmo gradiente de políticas determinista con redes profundas (DDPG, por sus siglas en inglés) para el aprendizaje del controlador de energía. Esta elección se fundamenta en que DDPG es un método actor - crítico, libre de modelo, especialmente adecuado para espacios de acción continuos, lo que lo hace compatible con la naturaleza del sistema a controlar y permite explorar estrategias de control que no dependen de una modelación explícita de la dinámica del entorno. La evaluación y comparación de otros algoritmos, como TD3 o SAC, se propone como línea de investigación futura que podría complementar y ampliar los resultados de este trabajo.

Como se indicó previamente, el algoritmo DDPG se implementa siguiendo la arquitectura actor-crítico, en la cual se entrenan dos redes neuronales profundas con roles diferenciados. La red actor representa la política del agente y tiene como función generar acciones continuas a partir del estado observado del entorno. La red crítica, en cambio, estima la calidad de las acciones tomadas, calculando su valor esperado en términos de la recompensa futura. Esta estimación corresponde a la llamada función acción-valor, denotada $Q(s, a)$, que representa la recompensa acumulada esperada al ejecutar una acción a en un estado s, y seguir la política actual en los pasos siguientes. Para entrenar estas redes el agente interactúa con el entorno recogiendo datos de sus propias decisiones. En cada instante de tiempo t, observa el estado actual del entorno, representado como $s_t \in \mathbb{R}^n$, y genera una acción $a_t \in \mathbb{R}^m$ utilizando el actor. Luego, ejecuta dicha acción en el entorno, recibe una recompensa inmediata $r_t \in \mathbb{R}$, y accede a un nuevo estado s_{t+1}. Esta experiencia se almacena como una tupla (s_t, a_t, r_t, s_{t+1}) en una memoria de experiencias que permite desacoplar el aprendizaje de las correlaciones temporales, facilitando el entrenamiento mediante minilotes aleatorios [25].

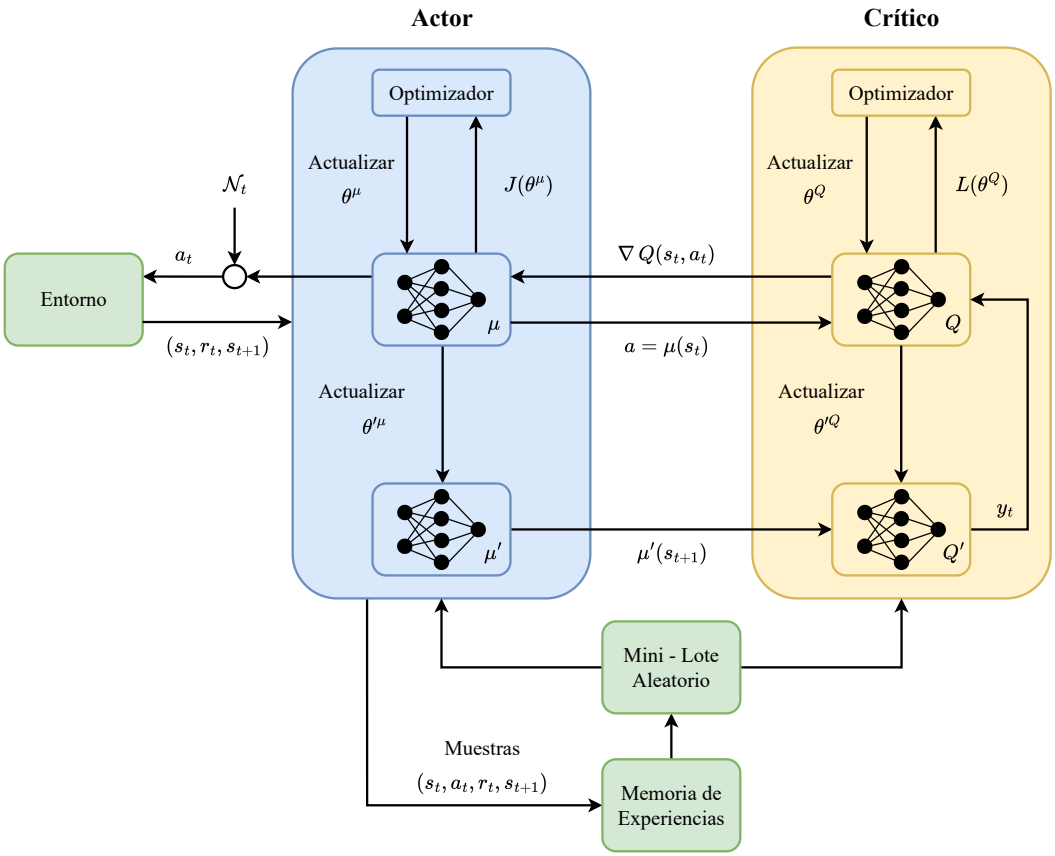

Figura 2.8: Algoritmo DDPG [35].

Durante el entrenamiento del crítico se utiliza una versión desacoplada y más estable de las redes principales, llamadas redes objetivo. Estas redes, denotadas μ' y Q', son copias del actor y el crítico, respectivamente, pero se actualizan de forma lenta para mejorar la estabilidad del aprendizaje. A partir de una muestra (s_t, a_t, r_t, s_{t+1}), se calcula un valor objetivo y_t, que representa una estimación de la recompensa acumulada esperada

$$y_t = r_t + \gamma\, Q'\left(s_{t+1}, \mu'(s_{t+1})\right) \tag{2.20}$$

donde γ es el factor de descuento, que determina la importancia relativa de las recompensas futuras con respecto a las inmediatas. El término $\mu'(s_{t+1})$ representa la acción que generaría la política objetivo al observar el siguiente estado, y $Q'(s_{t+1}, \mu'(s_{t+1}))$ estima el valor de esa acción en ese estado . La red crítica principal se entrena para aproximar este valor objetivo, minimizando el error cuadrático

$$L(\theta^Q) = \left(Q(s_t, a_t|\theta^Q) - y_t\right)^2 \tag{2.21}$$

donde θ^Q representa los parámetros de la red crítica. Este entrenamiento se realiza utilizando retropropagación y un optimizador de gradiente, como Adam.

El actor, por su parte, se entrena para generar acciones que maximicen el valor estimado por el crítico. Como no se dispone de una señal directa de error como en el caso del crítico, se utiliza el gradiente del valor $Q(s, a)$ con respecto a la acción a, evaluado en la acción actual $a = \mu(s)$, y se aplica la regla de la cadena para retropropagar ese gradiente a través del actor

$$\nabla_{\theta^\mu} J(\mu) \approx \mathbb{E}_{s\sim\mathcal{D}}\left[\nabla_a Q(s, a|\theta^Q)\big|_{a=\mu(s)} \cdot \nabla_{\theta^\mu}\mu(s|\theta^\mu)\right] \tag{2.22}$$

donde θ^μ son los parámetros del actor, y $\mathcal{D}$ representa el conjunto de muestras obtenidas desde la memoria de experiencias. Esta operación ajusta los pesos del actor para que produzca acciones que el crítico valore como más beneficiosas. Las redes objetivo se actualizan suavemente mediante un promedio móvil exponencial entre los parámetros actuales de las redes principales y los valores anteriores de la red objetivo [35]. La fórmula de actualización es

$$\theta' \leftarrow \tau\theta + (1 - \tau)\theta' \tag{2.23}$$

donde θ representa los parámetros actuales (ya sea del actor o del crítico), θ' los parámetros de la red objetivo correspondiente, y τ es una constante pequeña que controla la velocidad de actualización. Esta técnica evita fluctuaciones abruptas en el valor objetivo y contribuye a la estabilidad del entrenamiento. Dado que la política utilizada en DDPG es determinista, el agente repite las mismas acciones para los mismos estados, lo que limita la exploración del espacio de soluciones. Para mitigar este efecto, durante la recolección de experiencias se añade una señal de ruido $\mathcal{N}_t$ a las acciones generadas por el actor

$$a_t = \mu(s_t) + \mathcal{N}_t \tag{2.24}$$

donde $\mathcal{N}_t$ puede ser un ruido gaussiano o generado mediante un proceso de Ornstein–Uhlenbeck, que introduce correlación temporal entre pasos consecutivos. Este mecanismo incentiva al agente a explorar regiones del espacio de acciones que de otro modo no serían visitadas, lo cual es fundamental para descubrir estrategias de mayor rendimiento.

En resumen, DDPG entrena una política determinista a partir de la retroalimentación que le proporciona un estimador crítico del valor de las acciones. Este entrenamiento se realiza utilizando experiencias almacenadas, redes desacopladas para generar objetivos estables, y estrategias explícitas de exploración. Gracias a esta combinación de técnicas, el algoritmo logra aprender políticas eficaces en entornos continuos, con una buena relación entre estabilidad y eficiencia computacional [25].

Capítulo 3

Modelado de sistemas

En este capítulo se presenta cómo fueron modelados los diferentes sistemas que componen la planta de almacenamiento híbrida, detallando los supuestos y estrategias adoptadas. En primer lugar, se abordan las consideraciones relativas al convertidor formador de red, que incluyen el diseño del filtro LCL, el modelo de la red equivalente y la estrategia de control aplicada al convertidor. Posteriormente son descritos los modelos dinámicos de los subsistemas de almacenamiento, así como sus controladores asociados. Finalmente, se introduce el control de la tensión del enlace de corriente directa del convertidor.

3.1. Convertidor formador de red

Filtro LCL

Antes de presentar la estrategia de control asociada al convertidor, son descritos los criterios empleados para dimensionar el filtro LCL en función de las directrices estipuladas en [36]. Los parámetros de entrada requeridos para el diseño del filtro son la tensión nominal del convertidor V_n, la potencia aparente nominal del convertidor S_n, la frecuencia nominal de la red f_n, la frecuencia nominal de conmutación del convertidor f_{sw} y la tensión nominal del enlace de corriente directa v_{dc}.

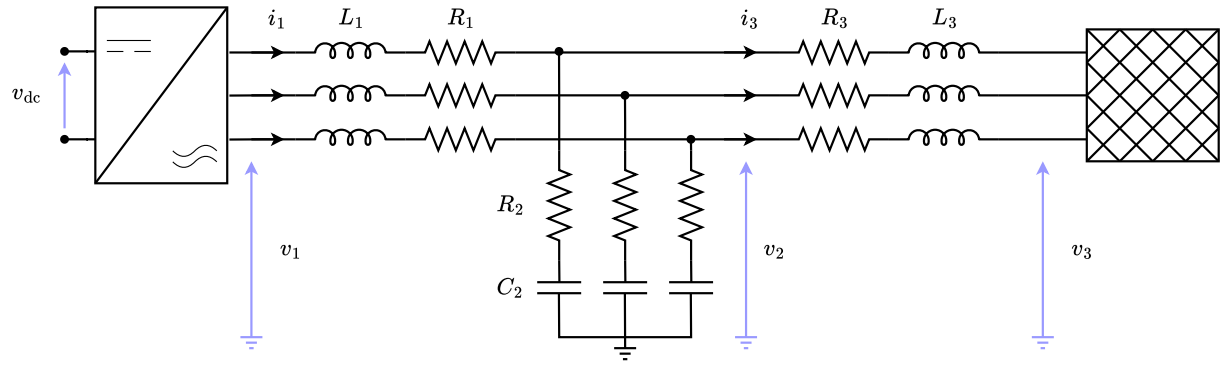

Figura 3.1: Sistema trifásico.

En primer lugar son calculadas la impedancia base, capacitancia base y corriente máxima de acuerdo con las expresiones presentadas a continuación.

$$Z_b = \frac{{V_n}^2}{S_n} \qquad C_b = \frac{1}{2\,\pi\,f_n\,Z_b} \qquad I_{\text{máx}} = \frac{S_n}{V_n}\sqrt{\frac{2}{3}} \tag{3.1}$$

La referencia indica que la capacitancia del filtro debe ser igual a un 5 % de la capacitancia base calculada. Por otra parte, la inductancia del lado del convertidor es dimensionada según lo indicado en (3.2).

$$L_1 = \frac{v_{\text{dc}}}{0{,}6\,f_{sw}\,I_{\text{máx}}} \tag{3.2}$$

La inductancia del lado de la red es dimensionada conforme a lo indicado por (3.3), donde k_a corresponde al coeficiente de atenuación del filtro.

$$L_3 = \frac{1+\sqrt{\frac{1}{{k_a}^2}}}{4\pi^2\,C_2\,{f_{sw}}^2} \tag{3.3}$$

La frecuencia de resonancia del filtro se calcula de acuerdo con (3.4), y debe verificarse que esta sea a lo menos 10 veces más grande que la frecuencia de la red, pero no más grande que la mitad de la frecuencia de conmutación del convertidor.

$$f_{res} = \frac{1}{2\,\pi}\sqrt{\frac{L_1+L_3}{L_1\,L_3\,C_2}} \tag{3.4}$$

Satisfecha la restricción anterior, se calcula la resistencia R_2 que tiene por objetivo atenuar el pico de resonancia característico de los filtros LCL. El dimensionamiento de este componente se lleva a cabo según lo indicado por la siguiente expresión.

$$R_2 = \frac{1}{6\,\pi\,f_{res}\,C_2} \tag{3.5}$$

Los parámetros de entrada considerados para el diseño del filtro LCL son los que se detallan en la Tabla 3.1, mientras que los parámetros dimensionados son presentados en la Tabla 3.2. Es importante indicar que las resistencias conectadas en serie a los inductores son estimadas considerando una razón X/R igual a 10, permitiendo capturar la dinámica resistiva sin recurrir a parámetros exactos.

Tabla 3.1: Parámetros de entrada.

Parámetro	**Unidad**	**Valor**
V_n	V	400
S_n	MVA	1
f_n	Hz	50
f_{sw}	kHz	10
v_{dc}	V	850
k_a	-	0.20

Tabla 3.2: Parámetros dimensionados.

Parámetro	**Unidad**	**Valor**
R_1	mΩ	2.50
R_2	mΩ	13.00
R_3	μΩ	48.00
L_1	μH	70.00
C_2	μF	995.00
L_3	μH	1.50

Al emplear modulación por ancho de pulso de vector espacial (SVPWM, por sus siglas en inglés), la tensión RMS línea - línea máxima que el convertidor puede suministrar sin entrar en sobremodulación se determina mediante la siguiente expresión.

$$V_{\text{máx}} = \frac{v_{\text{dc}}}{\sqrt{2}} \tag{3.6}$$

En la Figura 3.2 se presenta la carta de operación del convertidor, donde el círculo azul corresponde al límite por sobremodulación y el círculo naranja corresponde al círculo de potencia aparente nominal. Es importante indicar que la curva azul fue obtenida considerando tensión nominal en el punto de conexión, y que la base de potencia considerada es propia. Como el círculo azul envuelve al círculo naranja, es posible concluir que el dimensionamiento de los parámetros es correcto desde una perspectiva operacional.

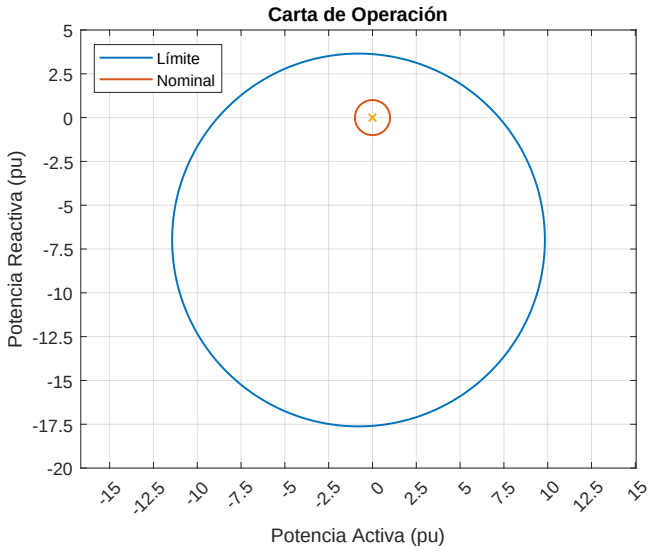

Figura 3.2: Carta de operación del convertidor.

Con el objetivo de verificar la adecuada capacidad del filtro para atenuar armónicos, se conectó a sus terminales una carga puramente resistiva, con un valor igual a la impedancia base definida durante el dimensionamiento de parámetros, para posteriormente imponer tensión nominal en terminales del convertidor. La Figura 3.3 muestra las formas de onda obtenidas bajo esta condición de ensayo.

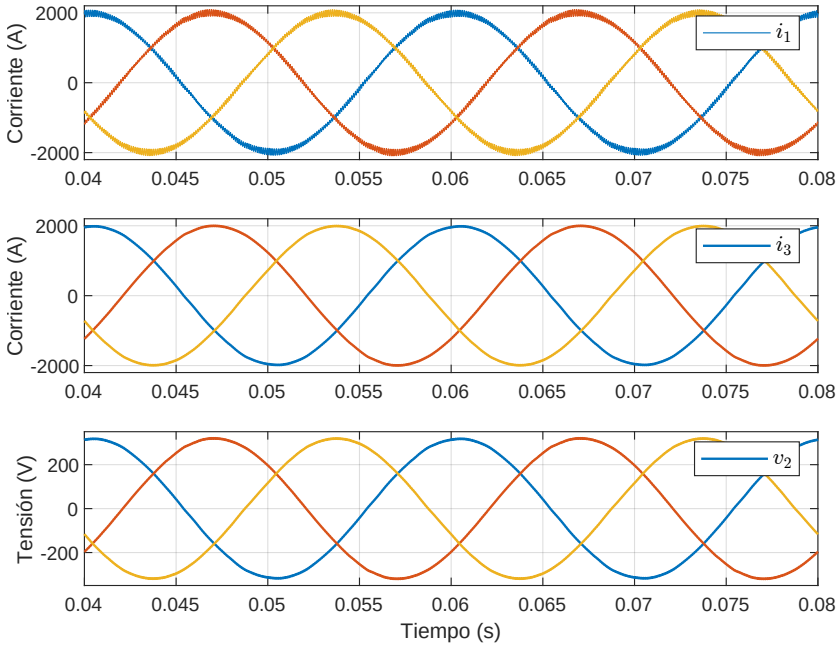

Figura 3.3: Formas de onda obtenidas en el filtro.

Cualitativamente se comprueba la efectividad de los parámetros dimensionados, pues el contenido armónico en la carga resistiva es prácticamente imperceptible a simple vista. Adicionalmente, con la finalidad de proveer un análisis más preciso, se calculó la distorsión armónica total de cada señal asegurando el cumplimiento de los requisitos establecidos por la norma IEEE 519 - 2022 para el sistema estudiado [37].

$$\text{THD}\,(i_1) = 2{,}81\,\% \qquad \text{THD}\,(i_3) = 0{,}61\,\% \qquad \text{THD}\,(v_2) = 0{,}64\,\%$$

Estrategia de control

La estrategia de control implementada en el convertidor trifásico se basa en [38] y sigue en gran medida la filosofía expuesta en la Figura 2.2. El esquema comprende dos niveles: un lazo interno que regula la tensión en la rama capacitiva del filtro y un lazo externo encargado de gestionar el intercambio de potencia activa y reactiva. Un rasgo distintivo de la estrategia adoptada es que prescinde del lazo de seguimiento de fase para sincronizarse con la red, pues el regulador de potencia activa alcanza el sincronismo de forma autónoma al basar su ley de control en la ecuación de equilibrio mecánico de una máquina sincrónica.

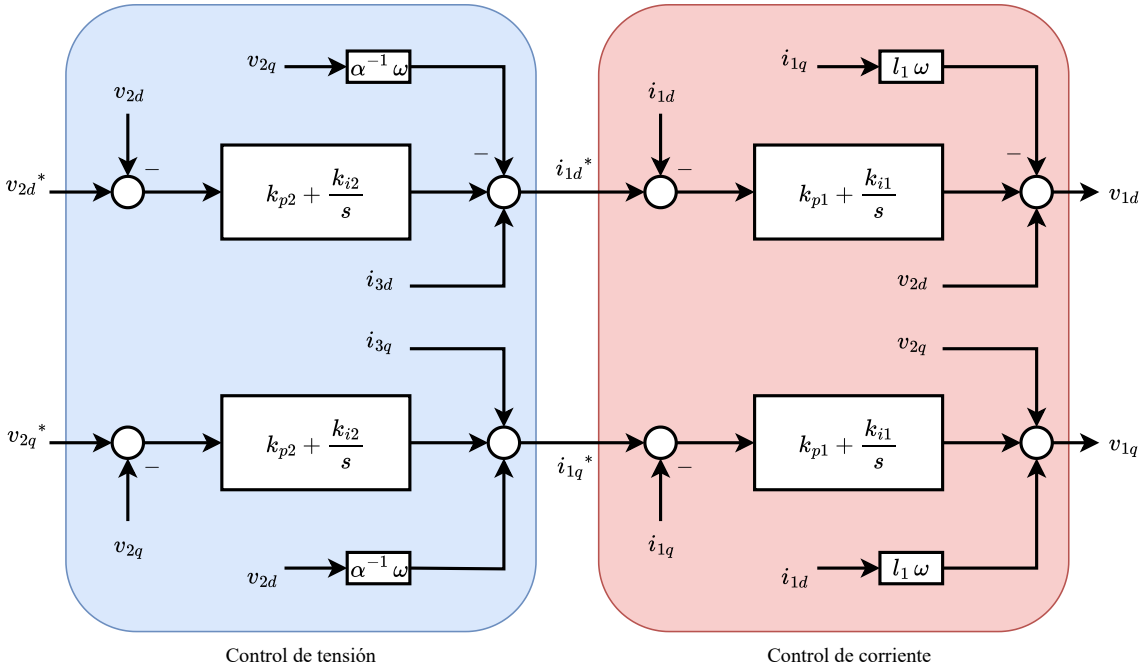

Figura 3.4: Lazo interno de control [38].

En la Figura 3.4 se presenta el lazo interno de control del convertidor, el cual adopta una estructura en cascada. La implementación del controlador incorpora esquemas anti-windup basados en recálculo, cuya finalidad es evitar la sobremodulación del convertidor y limitar la corriente en sus terminales. Sin embargo, dichos esquemas no se describen en el cuerpo del documento con el propósito de mantener la claridad y el orden en la presentación de los contenidos.

La sintonización convencional de los controladores suele basarse en la separación temporal entre los lazos de control, asumiendo que el lazo de tensión debe ser aproximadamente diez veces más lento que el lazo de corriente. Este enfoque busca desacoplar la dinámica entre ambos lazos y simplificar el diseño del sistema de control. Sin embargo, si bien esta estrategia ofrece un desempeño adecuado en condiciones de operación aislada, presenta limitaciones importantes cuando el sistema se conecta a la red. En particular, se observan respuestas lentas ($T > 3s$) y, en ciertos casos, la aparición de modos inestables asociados a la interacción entre los lazos de control de tensión y potencia activa. Aunque en la literatura se han propuesto técnicas como el control deslizante o el uso de controladores mixtos H_∞/H_2, estas soluciones suelen ser complejas y conllevan una alta carga computacional. Otras metodologías, como la obtención de ganancias PI

a partir de la ubicación de polos basada en el análisis modal, logran mejorar la estabilidad pero no resuelven completamente la lentitud de la respuesta de tensión. Esto pone de manifiesto la necesidad de revisar el enfoque de sintonización tradicional del lazo de control interno, especialmente cuando se desea operar el convertidor bajo la modalidad GFM, donde una rápida respuesta del lazo de tensión es esencial para la resiliencia del sistema frente a perturbaciones [39].

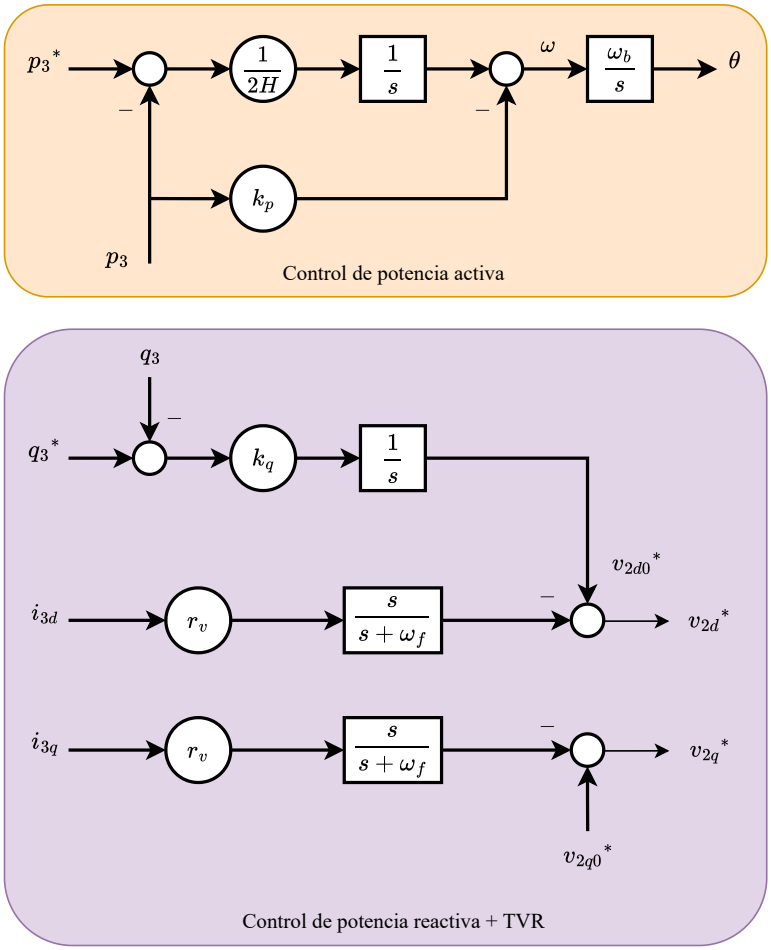

Figura 3.5: Lazo externo de control [38].

En la Figura 3.5 se muestra el lazo externo de control del convertidor, en el que se identifica un controlador orientado al manejo de la potencia activa y otro encargado de la regulación de la potencia reactiva. Con el objetivo de presentar de forma cualitativa el principio de funcionamiento de este lazo, a continuación se introducen las expresiones que describen la inyección de potencia activa y reactiva en estado estacionario. Cabe señalar que ambas expresiones se encuentran normalizadas conforme al sistema por unidad definido en el Apéndice A.

$$p_3 = \frac{|\mathbf{v}_3|\,|\mathbf{v}_2|}{|z_3|}\cos(\theta_3 - \theta_2 + \theta_{z_3}) - \frac{|\mathbf{v}_3|^2}{|z_3|}\cos(\theta_{z_3}) \tag{3.7}$$

$$q_3 = \frac{|\mathbf{v}_3|\,|\mathbf{v}_2|}{|z_3|}\sin(\theta_3 - \theta_2 + \theta_{z_3}) - \frac{|\mathbf{v}_3|^2}{|z_3|}\sin(\theta_{z_3}) \tag{3.8}$$

La diferencia angular $\theta_3 - \theta_2$ suele ser pequeña, y si se cumple que $\theta_{z_3} \approx \pi/2$, es posible establecer las siguientes expresiones simplificadas.

$$p_3 = \frac{|\mathbf{v}_3|\,|\mathbf{v}_2|}{|z_3|}\,(\theta_2 - \theta_3) \tag{3.9}$$

$$q_3 = \frac{|\mathbf{v}_3|}{|z_3|}\,(|\mathbf{v}_2| - |\mathbf{v}_3|) \tag{3.10}$$

Es claro que la potencia activa depende de la diferencia angular entre los fasores espaciales $\mathbf{v}_2$ y $\mathbf{v}_3$, mientras que la potencia reactiva está determinada por la diferencia entre sus respectivas magnitudes de tensión. La estrategia de control opera sobre un marco de referencia sincrónico dq, el cual se presenta en la Figura 3.6.

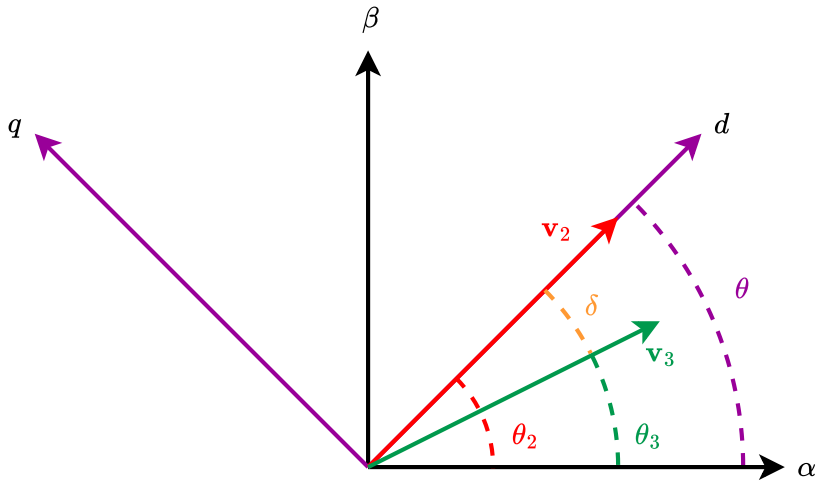

Figura 3.6: Marco de referencia dq.

La filosofía de control sigue según el siguiente razonamiento: el fasor espacial de tensión $\mathbf{v}_3$ puede encontrarse en cualquier lugar del marco de referencia estacionario $\alpha\beta$, y prescindir de un lazo de seguimiento de fase implica que el controlador desconoce su posición. Sin embargo, si convenientemente el fasor espacial de tensión $\mathbf{v}_2$ se orienta con el eje directo del marco de referencia sincrónico, el controlador de potencia acelerará o frenará dicho marco de referencia hasta alcanzar una diferencia angular constante γ, implicando la sincronización del convertidor con la red y un intercambio controlado de potencia activa. La condición de orientación es alcanzada imponiendo

$$v_{2d0}{}^{*} = |\mathbf{v}_2| \qquad\qquad v_{2q0}{}^{*} = 0 \tag{3.11}$$

Bajo la condición de orientación adoptada, la magnitud del fasor espacial de tensión $\mathbf{v}_2$ está determinada únicamente por su componente de eje directo, lo cual es aprovechado por el controlador de potencia reactiva para regular la inyección de reactivos, conforme a lo establecido en (3.10). Adicionalmente, se incorpora una resistencia virtual transitoria (TVR, por sus siglas en inglés) con el propósito de mejorar la respuesta dinámica del enlace, ya que la predominancia de la componente inductiva implica un bajo amortiguamiento ante perturbaciones.

Sintonización de controladores

Tal como se señaló previamente, la sintonización convencional del lazo interno presenta limitaciones que pueden derivar en un desempeño ineficiente e incluso contraproducente al momento de establecer la conexión con la red. Para superar dichos inconvenientes se adoptó una estrategia de sintonización multivariable, fundamentada en el algoritmo de optimización por enjambre de partículas (PSO, por sus siglas en inglés)

donde la ubicación de los polos se establece como criterio de diseño. Cuando el modelo se representa en el dominio de variables de estado, los polos se encuentran determinados por los valores propios de la matriz $\mathbf{A}$. En este marco, el proceso de optimización busca ajustar los parámetros de los controladores del lazo interno de modo que los polos se ubiquen en las posiciones deseadas, lo cual se consigue iterando sobre la matriz $\mathbf{A}$ mediante el algoritmo PSO. La dinámica conjunta del filtro LCL y del lazo de control interno puede describirse a través del siguiente modelo de estados

$$\frac{1}{\omega_b}\,\mathbf{F_o}\,\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{x} = \mathbf{A_o}\,\mathbf{x} + \mathbf{B_o}\,\mathbf{u} \tag{3.12}$$

Los vectores de estado y entrada del sistema son

$$\mathbf{x} = \begin{bmatrix} i_{1d} & i_{1q} & i_{3d} & i_{3q} & v_{2d} & v_{2q} & x_{1d} & x_{1q} & x_{2d} & x_{2q} \end{bmatrix}^{\mathrm{T}} \qquad \mathbf{u} = \begin{bmatrix} {v_{2d}}^* & {v_{2q}}^* & v_{3d} & v_{3q} \end{bmatrix}^{\mathrm{T}}$$

La matriz $\mathbf{A_o}$ del sistema es

$$\mathbf{A_o} = \begin{bmatrix} -k_{p1} - r_1 & 0 & 0 & 0 & -k_{p1}\,k_{p2} - 1 & -\frac{k_{p1}\,\omega}{\alpha} & k_{i1} & 0 & k_{i2}\,k_{p1} & 0 \\ 0 & -k_{p1} - r_1 & 0 & 0 & \frac{k_{p1}\,\omega}{\alpha} & -k_{p1}\,k_{p2} - 1 & 0 & k_{i1} & 0 & k_{i2}\,k_{p1} \\ 0 & 0 & -r_3 & l_3\,\omega & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -l_3\,\omega & -r_3 & 0 & 1 & 0 & 0 & 0 & 0 \\ \sigma_3 & -\omega\,r_2 & \beta & 0 & \sigma_2 & \omega - \sigma_1 & \frac{k_{i1}\,r_2}{l_1} & 0 & \sigma_4 & 0 \\ \omega\,r_2 & \sigma_3 & 0 & \beta & \sigma_1 - \omega & \sigma_2 & 0 & \frac{k_{i1}\,r_2}{l_1} & 0 & \sigma_4 \\ -1 & 0 & 0 & 0 & -k_{p2} & -\frac{\omega}{\alpha} & 0 & 0 & k_{i2} & 0 \\ 0 & -1 & 0 & 0 & \frac{\omega}{\alpha} & -k_{p2} & 0 & 0 & 0 & k_{i2} \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Las matrices $\mathbf{B_o}$ y $\mathbf{F_o}$ del sistema son respectivamente

$$\mathbf{B_o} = \begin{bmatrix} k_{p1}\,k_{p2} & 0 & 0 & 0 \\ 0 & k_{p1}\,k_{p2} & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ \frac{k_{p1}\,k_{p2}\,r_2}{l_1} & 0 & \frac{r_2}{l_3} & 0 \\ 0 & \frac{k_{p1}\,k_{p2}\,r_2}{l_1} & 0 & \frac{r_2}{l_3} \\ k_{p2} & 0 & 0 & 0 \\ 0 & k_{p2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad \mathbf{F_o} = \begin{bmatrix} l_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & l_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & l_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & l_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Es relevante señalar que el modelo se encuentra normalizado conforme al sistema por unidad definido en el Anexo A. En consecuencia, todas las constantes se expresan en por unidad y aquellas que no han sido referidas previamente se detallan a continuación.

$$\alpha = \frac{1}{c_2} - \frac{r_1\, r_2}{l_1} \qquad \beta = \frac{r_2\, r_3}{l_3} - \frac{1}{c_2} \qquad \gamma = r_2 \left(\frac{1}{l_1} + \frac{1}{l_3}\right)$$

$$\sigma_1 = \frac{k_{p1}\, \omega\, r_2}{\alpha\, l_1} \qquad \sigma_2 = -\gamma - \frac{k_{p1}\, k_{p2}\, r_2}{l_1} \qquad \sigma_3 = \alpha - \frac{k_{p1}\, r_2}{l_1} \qquad \sigma_4 = \frac{k_{i2}\, k_{p1}\, r_2}{l_1}$$

La matriz $\mathbf{A}$ se calcula según

$$\mathbf{A} = \omega_b\, {\mathbf{F_o}}^{-1}\, \mathbf{A_o} \tag{3.13}$$

Todos los polos del sistema tendrán la forma

$$p = -\zeta\, \omega_n \pm j\, \omega_n\, \sqrt{1-\zeta^2} = -\rho \pm j\, \lambda \tag{3.14}$$

La condición que debe satisfacerse para que el amortiguamiento de p sea igual o superior a un amortiguamiento mínimo $\zeta_{\min}$ es aquella que se presenta a continuación.

$$\frac{\lambda}{\rho} \leq \frac{\sqrt{1-{\zeta_{\min}}^2}}{\zeta_{\min}} \tag{3.15}$$

Adicionalmente, el tiempo de asentamiento del polo en cuestión puede ser estimado según la siguiente expresión.

$$\tau_p \sim \frac{4}{\rho} \tag{3.16}$$

El tiempo de asentamiento global queda dado por el polo dominante del sistema, por lo que para un tiempo de asentamiento objetivo deberá exigirse la siguiente condición.

$$\rho \geq \frac{4}{\tau_{p,o}} = \rho_{\min} \tag{3.17}$$

De la misma manera, ningún polo puede ser más rápido que el retardo equivalente del convertidor dada la estructura en cascada, imponiéndose así la última condición sobre el problema [39].

$$\rho \leq 2 f_{sw} = \rho_{\max} \tag{3.18}$$

El objetivo del algoritmo PSO es ajustar las ganancias de los controladores de manera que los polos del sistema se ubiquen dentro de la región definida por las condiciones previamente establecidas, cuya geometría característica corresponde a un trapecio. Sea $\mathcal{P}$ el conjunto de los polos del sistema, entonces, siempre que no exista un polo inestable en el conjunto, la función de costo del algoritmo es

$$J(\mathcal{P}) = \sum_{p\, \in\, \mathcal{P}} w_1\, [\text{máx}(0, \rho_{\min} - \rho)^2 + \text{máx}(0, \rho - \rho_{\max})^2] + w_2\, \text{máx}(0, \zeta_{\min} - \zeta)^2 \tag{3.19}$$

donde w_1 y w_2 son respectivamente los pesos asociados a la penalización por abandonar la banda de tiempo de asentamiento, y por poseer un amortiguamiento inferior al mínimo. En caso de que alguno de los polos de $\mathcal{P}$ sea inestable se penaliza directamente sin evaluar con 1×10^9. Como criterios de sintonización se definen

$$\rho_{\min} = 8{,}00 \text{ rad/s} \qquad \rho_{\max} = 20 \times 10^3 \text{ rad/s} \qquad \zeta_{\min} = 0{,}707$$

En la Tabla 3.3 se presentan los parámetros definidos para la ejecución del algoritmo de partículas, y en la Figura 3.7 es posible visualizar la evolución de la función de costo durante la optimización.

Tabla 3.3: Configuraciones de PSO.

Parámetro	**Valor**
Número de partículas	30
Iteraciones	1000
Rango de inercia	[0,1 1,1]
Pesos $[w_1,\ w_2]$	$[10^5\ \ 10^2]$

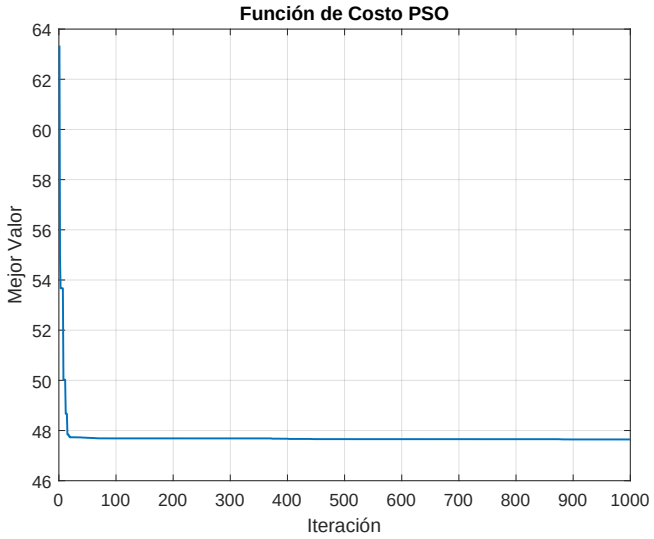

Figura 3.7: Evolución de la función de costo PSO.

En la Figura 3.8 se presenta el mapa de polos resultante de la optimización, mientras que en la Tabla 3.4 se informan los parámetros sintonizados. Se observa que no todas las condiciones establecidas fueron estrictamente satisfechas, dado que algunos polos se ubican fuera de la región trapezoidal definida como permisible. No obstante, dichos polos presentan un rápido decaimiento por lo que es altamente probable que su influencia sobre el desempeño global del sistema resulte despreciable.

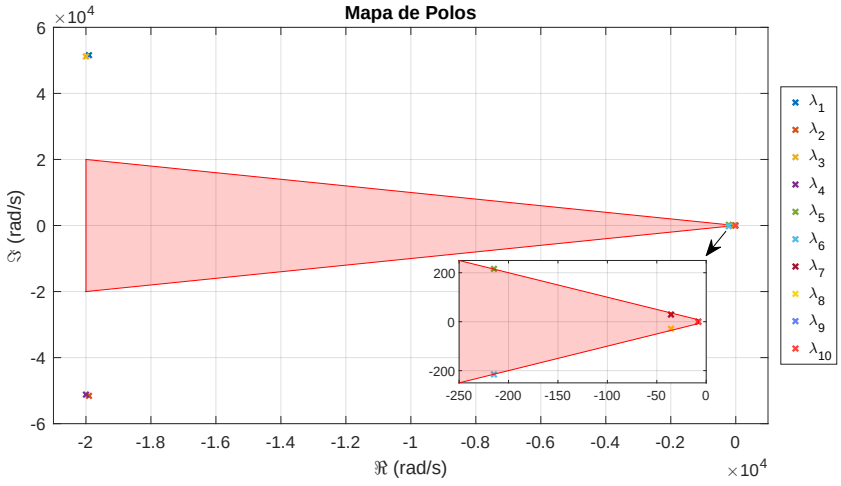

Figura 3.8: Mapa de polos del lazo interno.

Tabla 3.4: Parámetros sintonizados en lazo interno.

Parámetro	**Valor**
k_{p1}	0.35
k_{p2}	472.42
k_{i1}	0.0088
k_{i2}	86.23

En la Figura 3.9 se presenta la respuesta del lazo interno frente a un escalón aplicado a la referencia de la tensión $\mathbf{v_2}$. En el instante $t = 1$ se introduce un escalón en la componente de eje directo, mientras que en $t = 2$ se aplica un escalón en la componente de eje en cuadratura. Se observa que el controlador exhibe un desempeño satisfactorio en ambos canales, dado que asegura un adecuado seguimiento de referencia y un correcto rechazo a perturbaciones. Asimismo, se confirma que los polos que no cumplían los criterios de diseño no ejercen una influencia significativa en la dinámica.

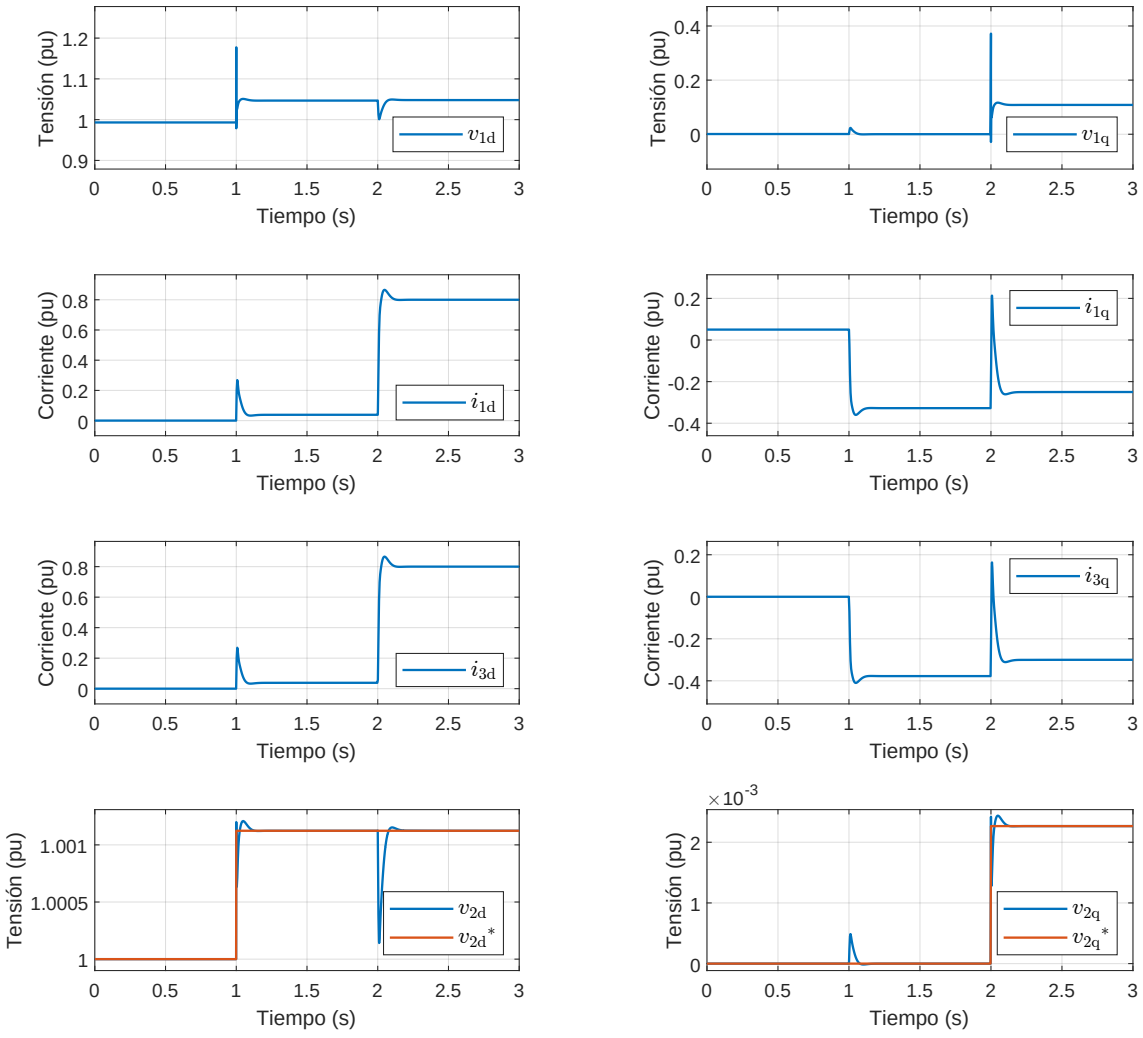

Figura 3.9: Desempeño del lazo interno.

Cabe destacar que los resultados presentados se obtuvieron imponiendo tensión nominal en el punto de conexión, considerando además que sus componentes de eje directo y en cuadratura permanecen constantes. Bajo estas consideraciones, en la Figura 3.9 el convertidor se encuentra operando en régimen GFL.

Una vez verificado el adecuado funcionamiento del lazo interno de control, se presenta la metodología adoptada para la sintonización del lazo externo. El primer paso consiste en integrar al modelo en espacio de estados la resistencia virtual transitoria, concebida como un esquema de control destinado a proporcionar amortiguamiento adicional en el punto de conexión. Se redefine entonces

$$\frac{1}{\omega_b}\,\mathbf{F_o}\,\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{x} = \mathbf{A_o}\,\mathbf{x} + \mathbf{B_o}\,\mathbf{u} \tag{3.20}$$

Los vectores de estado y entrada del modelo redefinido son

$$\mathbf{x} = \begin{bmatrix} i_{1d} & i_{1q} & i_{3d} & i_{3q} & v_{2d} & v_{2q} & x_{1d} & x_{1q} & x_{2d} & x_{2q} & \xi_d & \xi_q \end{bmatrix}^{\mathrm{T}} \qquad \mathbf{u} = \begin{bmatrix} {v_{2d0}}^* & {v_{2q0}}^* & v_{3d} & v_{3q} \end{bmatrix}^{\mathrm{T}}$$

La matriz $\mathbf{A_o}$ redefinida es

$$\mathbf{A_o} = \begin{bmatrix}
-k_{p1}-r_1 & 0 & -k_{p1}\,k_{p2}\,r_v & 0 & -k_{p1}\,k_{p2}-1 & -\frac{k_{p1}\,\omega}{\alpha} & k_{i1} & 0 & k_{i2}\,k_{p1} & 0 & \sigma_6 & 0 \\
0 & -k_{p1}-r_1 & 0 & -k_{p1}\,k_{p2}\,r_v & \frac{k_{p1}\,\omega}{\alpha} & -k_{p1}\,k_{p2}-1 & 0 & k_{i1} & 0 & k_{i2}\,k_{p1} & 0 & \sigma_6 \\
0 & 0 & -r_3 & l_3\,\omega & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -l_3\,\omega & -r_3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\sigma_5 & -\omega\,r_2 & \sigma_2 & 0 & \sigma_4 & \omega-\sigma_3 & \frac{k_{i1}\,r_2}{l_1} & 0 & \sigma_8 & 0 & \sigma_1 & 0 \\
\omega\,r_2 & \sigma_5 & 0 & \sigma_2 & \sigma_3-\omega & \sigma_4 & 0 & \frac{k_{i1}\,r_2}{l_1} & 0 & \sigma_8 & 0 & \sigma_1 \\
-1 & 0 & -k_{p2}\,r_v & 0 & -k_{p2} & -\frac{\omega}{\alpha} & 0 & 0 & k_{i2} & 0 & \sigma_7 & 0 \\
0 & -1 & 0 & -k_{p2}\,r_v & \frac{\omega}{\alpha} & -k_{p2} & 0 & 0 & 0 & k_{i2} & 0 & \sigma_7 \\
0 & 0 & -r_v & 0 & -1 & 0 & 0 & 0 & 0 & 0 & \frac{\omega_f\,r_v}{\omega_b} & 0 \\
0 & 0 & 0 & -r_v & 0 & -1 & 0 & 0 & 0 & 0 & 0 & \frac{\omega_f\,r_v}{\omega_b} \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\omega_f}{\omega_b} & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{\omega_f}{\omega_b}
\end{bmatrix}$$

Las matrices $\mathbf{B_o}$ y $\mathbf{F_o}$ redefinidas son respectivamente

$$\mathbf{B_o} = \begin{bmatrix}
k_{p1}\,k_{p2} & 0 & 0 & 0 \\
0 & k_{p1}\,k_{p2} & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1 \\
\frac{k_{p1}\,k_{p2}\,r_2}{l_1} & 0 & \frac{r_2}{l_3} & 0 \\
0 & \frac{k_{p1}\,k_{p2}\,r_2}{l_1} & 0 & \frac{r_2}{l_3} \\
k_{p2} & 0 & 0 & 0 \\
0 & k_{p2} & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix} \qquad \mathbf{F_o} = \begin{bmatrix}
l_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & l_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & l_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & l_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

De la misma forma en que son redefinidas las matrices, deben ser redefinidas las constantes σ_i. Las constantes α, β y γ no se ven alteradas pues estas son inherentes al filtro LCL.

$$\sigma_1 = \frac{k_{p1}\, k_{p2}\, \omega_f\, r_2\, r_v}{l_1\, \omega_b} \qquad \sigma_2 = \beta - \frac{k_{p1}\, k_{p2}\, r_2\, r_v}{l_1} \qquad \sigma_3 = \frac{k_{p1}\, \omega\, r_2}{\alpha\, l_1} \qquad \sigma_4 = -\gamma - \frac{k_{p1}\, k_{p2}\, r_2}{l_1}$$

$$\sigma_5 = \alpha - \frac{k_{p1}\, r_2}{l_1} \qquad \sigma_6 = \frac{k_{p1}\, k_{p2}\, \omega_f\, r_v}{\omega_b} \qquad \sigma_7 = \frac{k_{p2}\, \omega_f\, r_v}{\omega_b} \qquad \sigma_8 = \frac{k_{i2}\, k_{p1}\, r_2}{l_1}$$

Las potencias activa y reactiva en el punto de conexión quedan dadas por las siguientes expresiones en por unidad.

$$p_3 = v_{3d}\, i_{3d} + v_{3q}\, i_{3q} \qquad\qquad q_3 = v_{3q}\, i_{3d} - v_{3d}\, i_{3q} \tag{3.21}$$

Imponiendo la condición de orientación de la Figura 3.6 y linealizando en torno a un punto de operación se obtiene

$$\begin{aligned} \Delta p_3 &= \hat{v}_3 \cos\delta_0\, \Delta i_{3d} - \hat{v}_3 \sin\delta_0\, \Delta i_{3q} - (\hat{v}_3 \sin\delta_0\, i_{3d0} + \hat{v}_3 \cos\delta_0\, i_{3q0})\, \Delta\delta \\ \Delta q_3 &= -\hat{v}_3 \sin\delta_0\, \Delta i_{3d} - \hat{v}_3 \cos\delta_0\, \Delta i_{3q} + (\hat{v}_3 \sin\delta_0\, i_{3q0} - \hat{v}_3 \cos\delta_0\, i_{3d0})\, \Delta\delta \end{aligned} \tag{3.22}$$

Se define convenientemente

$$\Delta\mathbf{y} = \begin{bmatrix} \Delta p_3 & \Delta q_3 \end{bmatrix}^{\mathrm{T}}$$

Si la condición de orientación es sustituida en el modelo de estados redefinido y este es linealizado en torno a un punto de operación, es posible obtener

$$\begin{gathered} \frac{1}{\omega_b}\, \mathbf{F_o}\, \frac{\mathrm{d}}{\mathrm{d}t}\, \Delta\mathbf{x} = \mathbf{A_o}\, \Delta\mathbf{x} + \mathbf{B_o}'\, \Delta\mathbf{u} \\ \Delta\mathbf{y} = \mathbf{C}\, \Delta\mathbf{x} + \mathbf{D}\, \Delta\mathbf{u} \end{gathered} \tag{3.23}$$

donde

$$\mathbf{B_o}' = \begin{bmatrix} k_{p1}\, k_{p2} & 0 \\ 0 & 0 \\ 0 & \hat{v}_3 \sin\delta_0 \\ 0 & \hat{v}_3 \cos\delta_0 \\ \frac{k_{p1}\, k_{p2}\, r_2}{l_1} & -\frac{r_2}{l_3}\, \hat{v}_3 \sin\delta_0 \\ 0 & -\frac{r_2}{l_3}\, \hat{v}_3 \cos\delta_0 \\ k_{p2} & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \hat{v}_3 \cos\delta_0 & -\hat{v}_3 \sin\delta_0 \\ -\hat{v}_3 \sin\delta_0 & -\hat{v}_3 \cos\delta_0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^{T}$$

$$\mathbf{D} = \begin{bmatrix} 0 & -\hat{v}_3 \left(\sin\delta_0\, i_{3d0} + \cos\delta_0\, i_{3q0}\right) \\ 0 & -\hat{v}_3 \left(\cos\delta_0\, i_{3d0} - \sin\delta_0\, i_{3q0}\right) \end{bmatrix}$$

Los vectores de estado y entrada del modelo linealizado son

$$\Delta\mathbf{x} = \begin{bmatrix} \Delta i_{1d} & \Delta i_{1q} & \Delta i_{3d} & \Delta i_{3q} & \Delta v_{2d} & \Delta v_{2q} & \Delta x_{1d} & \Delta x_{1q} & \Delta x_{2d} & \Delta x_{2q} & \Delta\xi_d & \Delta\xi_q \end{bmatrix}^{\mathrm{T}}$$

$$\Delta\mathbf{u} = \begin{bmatrix} \Delta v_{2d0}{}^* & \Delta\delta \end{bmatrix}^{\mathrm{T}}$$

Los estados del modelo no se ven modificados, por lo que las matrices $\mathbf{A_o}$ y $\mathbf{F_o}$ no se ven alteradas. La construcción del modelo linealizado supone que la tensión en el punto de conexión es esencialmente constante, al igual que la frecuencia ω.

$$\hat{v}_3 = 1 \text{ pu} \qquad\qquad \omega = 1 \text{ pu}$$

Aplicando la transformada de Laplace sobre (3.23) y reordenando términos, es posible obtener

$$\begin{aligned} \Delta\mathbf{y} &= \left[\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{B} + \mathbf{D}\right]\Delta\mathbf{u} \\ &= \mathbf{G}(s)\,\Delta\mathbf{u} \end{aligned} \tag{3.24}$$

En su forma extendida

$$\begin{bmatrix} \Delta p_3 \\ \Delta q_3 \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} \Delta v_{2d0}{}^* \\ \Delta\delta \end{bmatrix} \tag{3.25}$$

De esta manera, es posible obtener una función de transferencia tanto para la sintonización del controlador de potencia activa como reactiva. Más aún, a partir del análisis de pequeña señal es posible identificar la razón por la cual el controlador puede operar sin un lazo de seguimiento de fase: el ángulo de la red supone una perturbación que es rechazada por el controlador.

$$\Delta\delta = \Delta\theta - \Delta\theta_3 \tag{3.26}$$

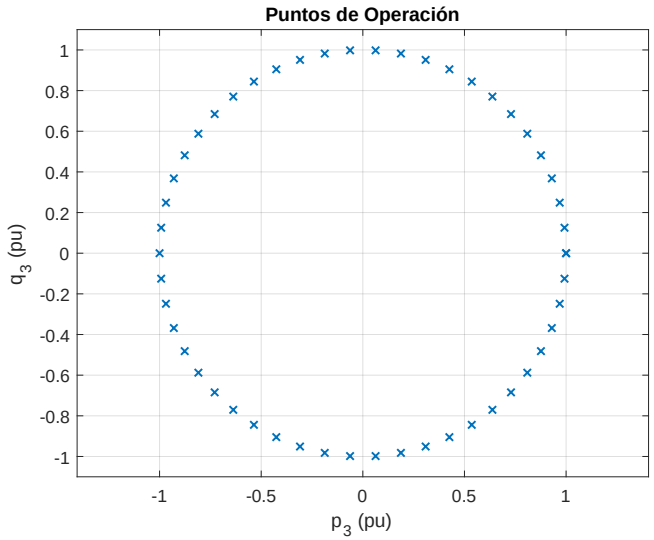

Figura 3.10: Puntos de operación evaluados.

En la Figura 3.10 se muestran los puntos de operación considerados para el estudio de las funciones de transferencia $G_{12}(s)$ y $G_{21}(s)$. Una revisión detallada de la operación del convertidor no es parte del alcance del presente trabajo, por lo que el análisis del impacto de las componentes $G_{11}(s)$ y $G_{22}(s)$ se reserva para investigaciones futuras.

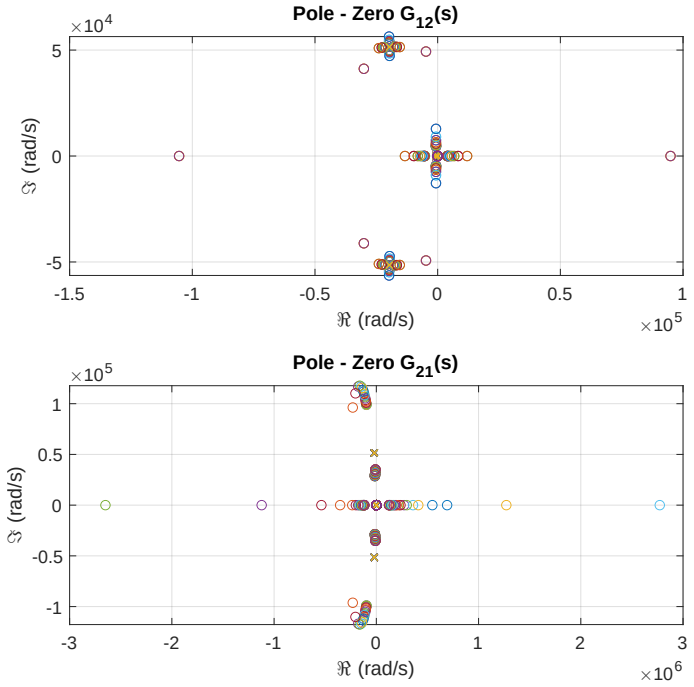

Figura 3.11: Mapa de polos y ceros para $G_{12}(s)$ y $G_{21}(s)$.

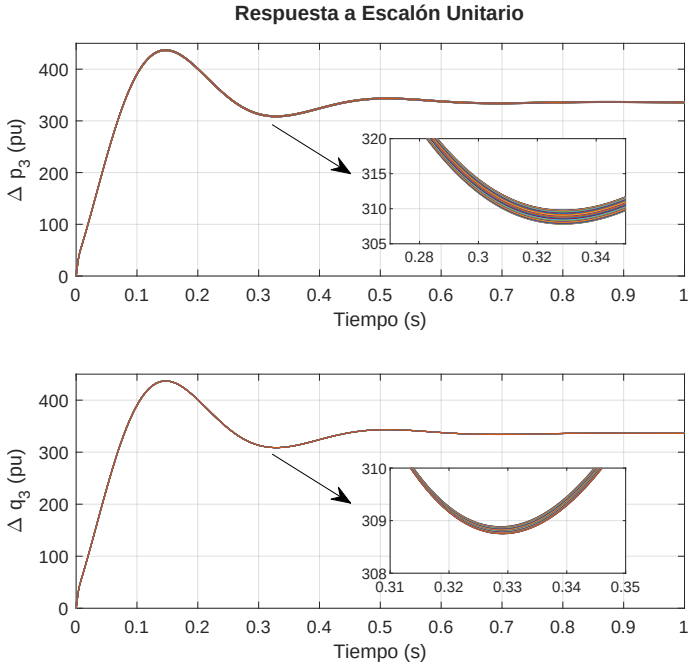

Figura 3.12: Respuesta a escalón unitario para $G_{12}(s)$ y $G_{21}(s)$.

En la Figura 3.11 se presenta el mapa de polos y ceros de las funciones de transferencia analizadas, donde se observa que la ubicación de los ceros depende del punto de operación mientras que la posición de los polos permanece invariante. Por su parte, la Figura 3.12 muestra la respuesta a un escalón unitario de los sistemas estudiados, evidenciando que, si bien la respuesta temporal varía según el punto de operación, dichas diferencias no resultan sustanciales. En síntesis, puede establecerse que el punto de operación no ejerce una influencia significativa sobre la dinámica global y que la variación en la localización de los ceros se manifiesta principalmente en las pequeñas diferencias entre los sobreimpulsos de cada respuesta. La independencia del punto de operación en el modelo de pequeña señal permite concluir que pequeñas variaciones tanto del ángulo δ como de la referencia de tensión ${v_{2d0}}^*$ generan un efecto que no depende del grado de carga. En consecuencia, resulta razonable suponer que una sintonización con ganancias constantes en los controladores de potencia es suficiente, sin que sea estrictamente necesaria la implementación de estrategias adaptativas. Es importante destacar que para estos resultados se consideraron los siguientes valores de resistencia virtual y frecuencia de corte.

$$r_v = 20\, r_3 \qquad\qquad \omega_f = 60 \text{ rad/s}$$

En el caso de los convertidores GFM el valor de la constante de inercia virtual no está determinado por factores mecánicos, como ocurre en las máquinas síncronas. Su magnitud se define en función de los requerimientos de la red eléctrica y de la energía disponible en el bus de corriente directa. El alcance del presente trabajo se restringe al desarrollo de un modelo funcional de convertidor GFM, sin abordar en profundidad los requerimientos específicos de la red. Para simplificar la elección de la constante de inercia virtual, se adopta un valor de 3,5 s correspondiente a la Unidad 15 de la Central Térmica Maule, la cual cuenta con una potencia nominal de 0.95 MVA. Esta definición resulta suficiente ya que el valor escogido es comparable al de una unidad generadora real, considerando que los convertidores GFM están concebidos para emular y reemplazar este tipo de generadores.

En la Figura 3.13 se presenta el lugar geométrico de las raíces para el lazo cerrado de potencia activa, considerando la ganancia k_p como grado de libertad en el rango [0,00 0,10]. Se observan polos en el semiplano derecho para ganancias relativamente pequeñas, los que paulatinamente se desplazan al semiplano izquierdo conforme la ganancia aumenta. A partir de simulaciones, se ha definido que un valor de 0.02 para k_p es suficiente para una operación estable y razonable del lazo de potencia activa.

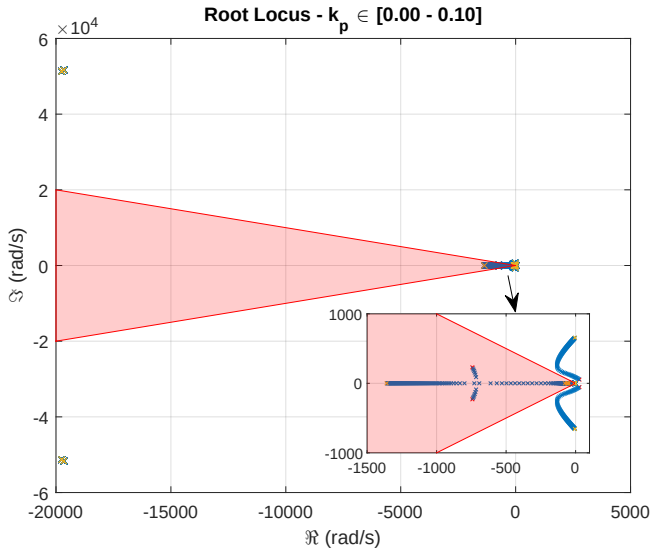

Figura 3.13: Lugar geométrico de las raíces $k_p \in$ [0,00 0,10].

En la Figura 3.14 se presenta el lugar geométrico de las raíces para el lazo cerrado de potencia reactiva, considerando la ganancia k_q como grado de libertad en el rango [0,01 1,00]. Se observa que los polos se desplazan al semiplano izquierdo conforme la ganancia se hace más grande o más pequeña, existiendo un rango intermedio en que el lazo estudiado es inestable. A partir de simulaciones, se ha definido que un valor de 0.02 para k_q es suficiente para una operación estable y razonable del lazo de potencia reactiva.

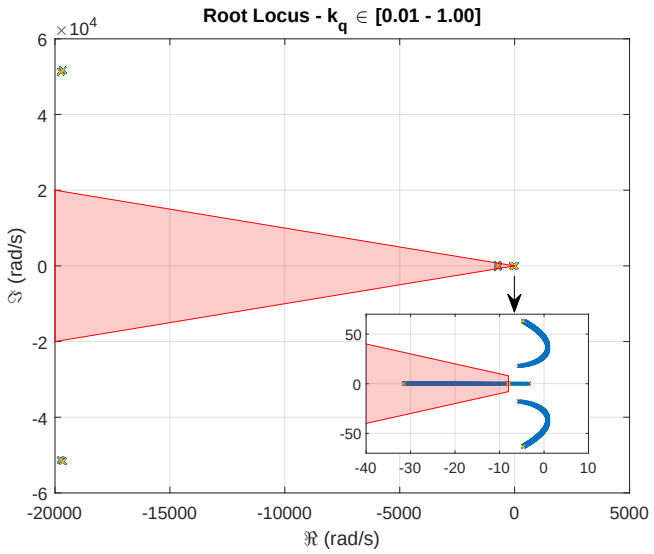

Figura 3.14: Lugar geométrico de las raíces $k_q \in$ [0,01 1,00].

Definidas las constantes del lazo externo de control, se procede a verificar el desempeño del convertidor en una red equivalente de frecuencia variable, cuya dinámica queda dada por el diagrama de bloques de la Figura 3.15. Cabe destacar que este modelo simplificado corresponde al utilizado en [38] para la validación de sus resultados.

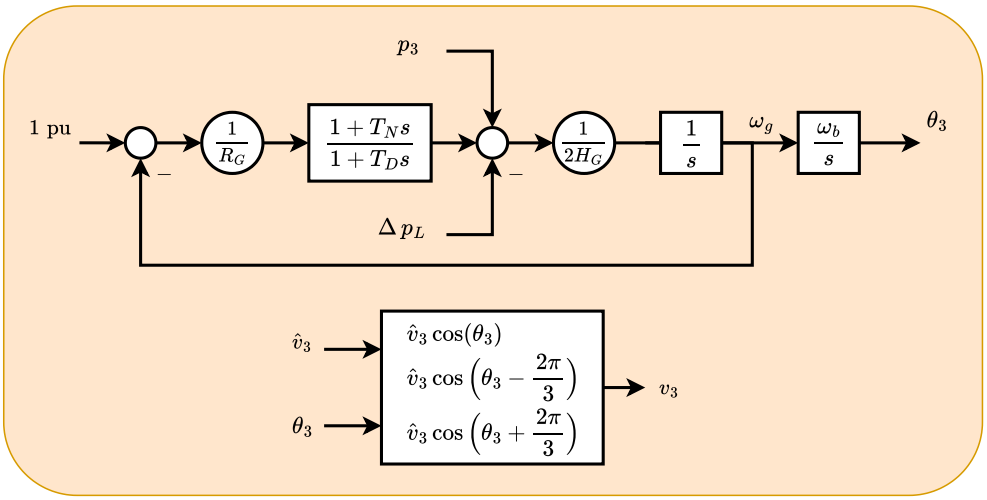

Figura 3.15: Modelo simplificado de red equivalente.

Para los valores de las constantes de tiempo de adelanto y retardo, respectivamente T_N y T_D, se consideran las cantidades definidas en el artículo referido. En relación con la constante de inercia equivalente del sistema H_G, esta es estimada a partir de (3.27) donde H_i y S_i corresponden respectivamente a la constante de inercia y potencia nominal aparente de la i-ésima central presente en la red equivalente.

$$H_G \approx \sum_i \frac{H_i\, S_i}{S_b} \tag{3.27}$$

El caso de estudio considera que el convertidor se conecta a una red de pequeña escala compuesta por dos unidades térmicas de distinta potencia y constante de inercia, cuyos valores se presentan en la Tabla 3.5. Asimismo, se implementa un control por estatismo en el lazo externo de potencia del convertidor, con el propósito de permitir el reparto de los desbalances de potencia. El estatismo R del convertidor se ajusta en 5 %, mientras que el estatismo de la red R_G se ajusta en 4 %.

Tabla 3.5: Parámetros de las unidades generadoras de la red equivalente.

Unidad	**Potencia nominal**	**Constante de inercia**
1	1.20 MVA	3.0 s
2	0.96 MVA	2.5 s

En la Figura 3.16 se muestra la respuesta del lazo de control interno ante un incremento de carga en $t = 5$ s. Se aprecia un transitorio propio de la perturbación, el cual decae progresivamente hasta alcanzarse un nuevo punto de operación en régimen estacionario. Asimismo, se confirma el correcto funcionamiento del lazo interno, pues la condición de orientación para v_{2q} se mantiene prácticamente satisfecha durante todo el evento.

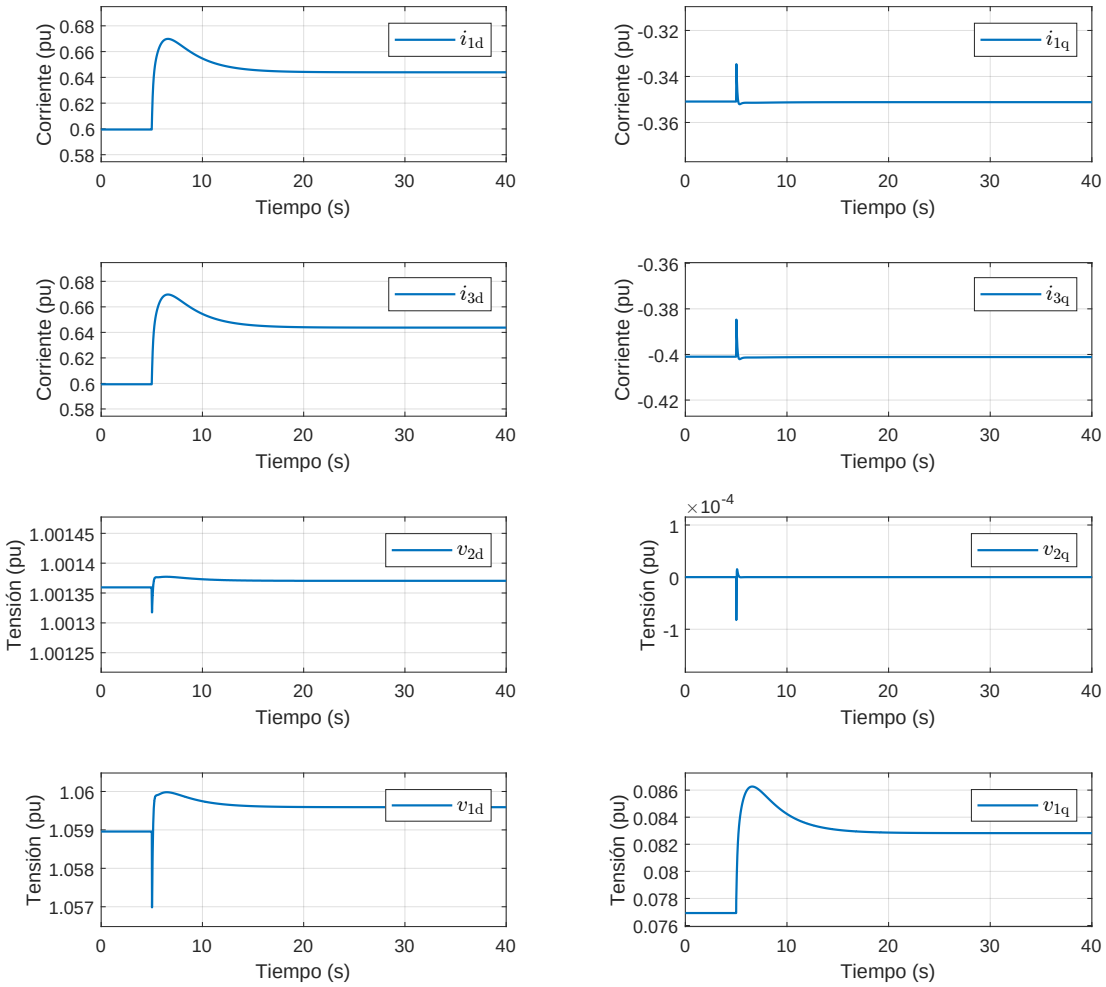

Figura 3.16: Respuesta del lazo de control interno ante una perturbación en la red.

En las Figuras 3.17 y 3.18 se presenta la respuesta del lazo externo de control frente a la perturbación previamente descrita. El desempeño observado resulta consistente con lo deseado, evidenciándose un intercambio regulado de potencia activa con la red como consecuencia de la perturbación. Asimismo, el controlador de potencia reactiva actúa correctamente al rechazar la perturbación inducida por la variación de la potencia activa, manteniendo el seguimiento de su referencia. Finalmente, la evolución de la frecuencia del convertidor se ajusta al comportamiento previsto por la teoría del control por estatismo, pues frente a un déficit de potencia en la red la frecuencia disminuye y la potencia activa aportada por el convertidor se incrementa.

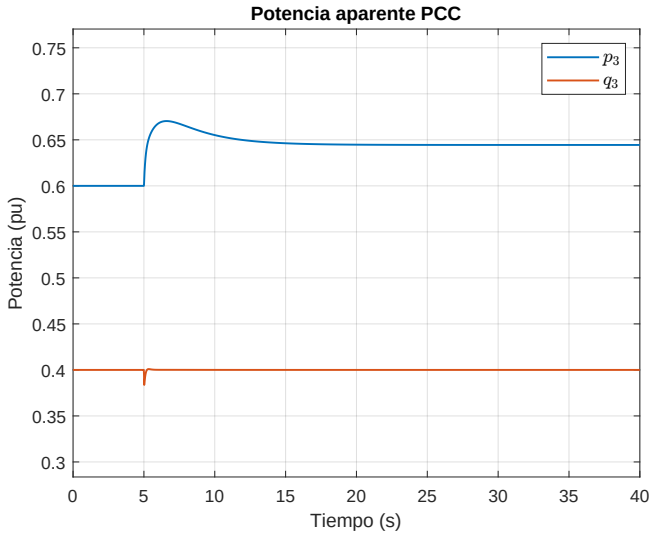

Figura 3.17: Respuesta del lazo de control externo ante una perturbación en la red.

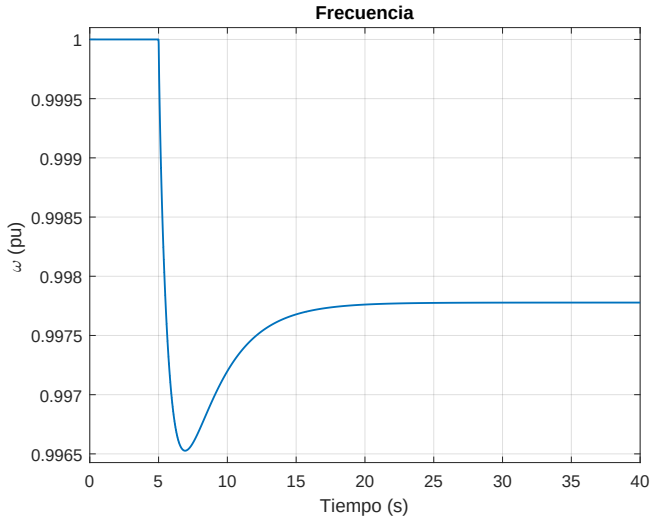

Figura 3.18: Frecuencia del convertidor ante una perturbación en la red.

Las simulaciones presentadas se realizaron utilizando modelos promedio, centrados en el comportamiento general del sistema sin considerar los efectos de alta frecuencia, como la conmutación. Este enfoque resulta adecuado, ya que para el desarrollo de la estrategia de control de energía es la dinámica de baja frecuencia la que determina el intercambio de potencia entre el convertidor y la red.

3.2. Electrolizador

El caso de estudio considera un electrolizador de membrana de electrolito polimérico, con una potencia preliminar de 1 MW para la producción de hidrógeno. Como se mencionó anteriormente, esta definición se sustenta en que dicha tecnología ofrece la respuesta más rápida dentro de todas las alternativas revisadas. El comportamiento del dispositivo será descrito utilizando el modelo para bajas densidades de corriente, ilustrado en la Figura 2.4(a). Dicho modelo ha sido validado experimentalmente para un electrolizador de tres celdas con una potencia y tensión nominal de 65 W y 8 V respectivamente, tal como se detalla en [40]. En la Tabla 3.6 se presentan los parámetros obtenidos para el equipo referido.

Tabla 3.6: Parámetros del electrolizador estudiado en [40].

Parámetro	**Unidad**	**Valor**
$R_{\text{ano,Uni}}$	mΩ	318.00
$R_{\text{cat,Uni}}$	mΩ	35.00
$C_{\text{ano,Uni}}$	F	37.26
$C_{\text{cat,Uni}}$	F	37.26
$R_{\text{ohm,Uni}}$	mΩ	88.00
$E_{\text{rev,Uni}}$	V	4.38

Tal como se mencionó previamente, modelar el comportamiento de un electrolizador mediante la conexión en serie y/o paralelo de unidades menores no constituye una aproximación completamente precisa. Una metodología rigurosa implicaría obtener los parámetros del electrolizador de forma experimental, analizando directamente su respuesta en los terminales, tal como se llevó a cabo en [40]. No obstante, en el presente trabajo se opta por representar el electrolizador a partir de la conexión en serie y/o paralelo de unidades menores, dado que no es factible obtener parámetros experimentales para un equipo de la potencia nominal definida. Si bien esta aproximación no es exacta, se considera adecuada para los fines de este estudio. La configuración adoptada para el electrolizador principal contempla m ramas en paralelo, donde cada una incluye n electrolizadores menores conectados en serie. A partir de los parámetros característicos de estas unidades, es posible calcular los parámetros equivalentes del equipo principal, conforme a lo indicado en (3.28). Los desarrollos matemáticos que respaldan esta formulación se detallan en el Apéndice B.

$$\begin{aligned} R_{\text{ano,EL}} &= \frac{n}{m} R_{\text{ano,Uni}} \qquad C_{\text{ano,EL}} = \frac{m}{n} C_{\text{ano,Uni}} \qquad R_{\text{ohm,EL}} = \frac{n}{m} R_{\text{ohm,Uni}} \\ R_{\text{cat,EL}} &= \frac{n}{m} R_{\text{cat,Uni}} \qquad C_{\text{cat,EL}} = \frac{m}{n} C_{\text{cat,Uni}} \qquad E_{\text{rev,EL}} = n\, E_{\text{rev,Uni}} \end{aligned} \tag{3.28}$$

En la Figura 3.19 se presenta el sistema completo de electrólisis, en el cual el electrolizador se conecta al enlace de corriente directa mediante un convertidor tipo Buck y un filtro L. La estrategia de control se basa en regular la corriente a través del filtro, ya que de este modo se controla indirectamente la corriente que circula por el electrolizador, tal como se implementa en [41].

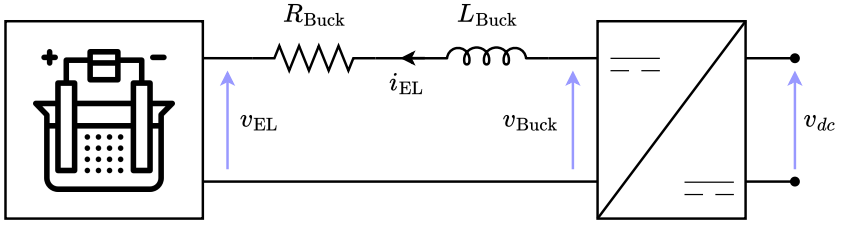

Figura 3.19: Conexión del electrolizador al enlace de corriente directa.

La potencia nominal del electrolizador principal depende directamente del arreglo configurado a partir de sus unidades menores. Al establecer una potencia nominal mínima $P_{\mathrm{n,mín}}$ para el equipo, las combinaciones de celdas que resultan válidas son aquellas que satisfacen la condición establecida en (3.29).

$$m \geq \frac{P_{\mathrm{n,mín}}}{n\, v_{\mathrm{n,Uni}}\, i_{\mathrm{n,Uni}}} \tag{3.29}$$

La capacidad del controlador para alcanzar el punto de operación nominal puede verse comprometida por una configuración $n \times m$ inadecuada. Con el fin de asegurar que el controlador pueda operar en el punto nominal, el ciclo de trabajo nominal D_n debe ser superior a cero e inferior a la unidad, incluyendo los límites, imponiendo así dos restricciones adicionales al problema. Se considera una banda de tolerancia del $\pm 5\,\%$ para la tensión del enlace de corriente directa, siendo la condición más desfavorable el límite inferior.

$$\begin{aligned} -\frac{n\, v_{\mathrm{n,Uni}}}{R_{\mathrm{Buck}}\, i_{\mathrm{n,Uni}}} &\leq m \\ \frac{v_{\mathrm{dc,mín}} - n\, v_{\mathrm{n,Uni}}}{R_{\mathrm{Buck}}\, i_{\mathrm{n,Uni}}} &\geq m \end{aligned} \tag{3.30}$$

Con el objetivo de limitar la potencia disipada en la componente resistiva del filtro, se debe establecer una cota de pérdidas asociada al punto de operación nominal. Dicha cota se define en función de un porcentaje de pérdidas objetivo $\%_{\mathrm{Loss}}$ conforme a lo indicado en (3.31).

$$\frac{\%_{\mathrm{Loss}}\, n\, v_{\mathrm{n,Uni}}}{R_{\mathrm{Buck}}\, i_{\mathrm{n,Uni}}} \geq m \tag{3.31}$$

Presentadas las condiciones anteriores, se define que los grados de libertad de diseño serán la potencia nominal mínima $P_{\mathrm{n,mín}}$, la resistencia del filtro R_{Buck} y el porcentaje de pérdidas objetivo $\%_{\mathrm{Loss}}$. En la Figura 3.20 se presenta el mapa de combinaciones que permite visualizar los diferentes arreglos que satisfacen las condiciones anteriormente presentadas, considerando una potencia nominal mínima de 1 MW, pérdidas del $2\,\%$ y una resistencia del inductor igual a 0.5 mΩ.

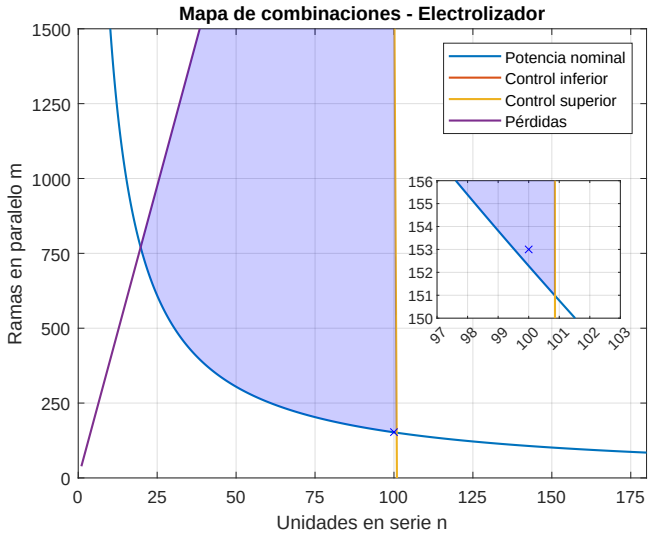

Figura 3.20: Mapa de combinaciones del sistema de electrólisis.

Se escoge el arreglo que considera 153 ramas en paralelo constituidas por 100 unidades en serie. Esta combinación cumple con todas las restricciones impuestas, encontrándose dentro del área factible denotada por color azul en la Figura 3.20. De esta manera, el electrolizador se adjudica una potencia nominal de 1 MW con una tensión y corriente nominal de 800 V y 1256 A, mientras que sus parámetros equivalentes son los detallados en la Tabla 3.7.

Tabla 3.7: Parámetros del electrolizador diseñado.

Parámetro	**Unidad**	**Valor**
$R_{\text{ano,EL}}$	mΩ	207.84
$R_{\text{cat,EL}}$	mΩ	22.88
$R_{\text{ohm,EL}}$	mΩ	57.52
$C_{\text{ano,EL}}$	F	57.01
$C_{\text{cat,EL}}$	F	57.01
$E_{\text{rev,EL}}$	V	438.00

La Ecuación (3.32) describe la dinámica de la corriente que circula por el filtro L, que es la misma que circula por el electrolizador. Es importante destacar que esta expresión se encuentra normalizada de acuerdo con el sistema por unidad definido en el Apéndice A.

$$\frac{l_{\text{Buck}}}{\omega_b}\frac{\mathrm{d}}{\mathrm{d}t}i_{\text{EL}} + r_{\text{Buck}}\, i_{\text{EL}} = v_{\text{Buck}} - v_{\text{EL}} \tag{3.32}$$

La arquitectura de control resultante es relativamente simple, ya que la variable de actuación del lazo corresponde a la tensión de salida del convertidor Buck, mientras que la tensión en los terminales del electrolizador supone una perturbación. El comportamiento dinámico de esta tensión se describe mediante (3.33), en la cual se advierte la presencia de dos retardos característicos, siendo el del ánodo más significativo que el del cátodo, adjudicándose un tiempo característico de 11.85 s. Si el controlador de corriente se sintoniza para responder con mayor rapidez que la perturbación, esta última tiende a comportarse como una señal cuasiestática desde la perspectiva del lazo de control, lo que favorece su supresión efectiva.

$$V_{\text{EL}}(s) = \left(\frac{R_{\text{ano,EL}}}{R_{\text{ano,EL}}\, C_{\text{ano,EL}} s + 1} + \frac{R_{\text{cat,EL}}}{R_{\text{cat,EL}}\, C_{\text{cat,EL}} s + 1} + R_{\text{ohm,EL}}\right) I_{\text{EL}}(s) + E_{\text{rev,EL}}(s) \tag{3.33}$$

Se ha identificado que el rizado de corriente característico de los convertidores de potencia acelera la degradación de electrolizadores tipo PEM [42]. En consecuencia, el criterio adoptado para dimensionar el inductor del filtro se basa en el enfoque clásico utilizado en convertidores Buck, aunque con un margen de diseño conservador. El dimensionamiento se lleva a cabo de acuerdo con (3.34), donde v_{out} corresponde a la tensión de salida del convertidor, D al ciclo de trabajo, f_{sw} a la frecuencia de conmutación y ΔI al rizado de corriente admitido.

$$L_{\text{Buck}} = \frac{v_{\text{out}}\,(1-D)}{f_{sw}\,\Delta I} \tag{3.34}$$

El ciclo de trabajo es calculado como el cociente entre la tensión nominal del equipo y la tensión nominal del enlace de corriente directa. Para la tensión de salida se considera la tensión nominal del electrolizador, y se admitirá un rizado máximo del 1 % de la corriente nominal del equipo, operando a una frecuencia de conmutación de 10 kHz. Haciendo uso de (3.34) el dimensionamiento resulta en un inductor de 375 μH.

La planta asociada al control de corriente es la presentada en (3.35), en la que se ha incluido el retardo asociado a la acción del convertidor mediante la constante de tiempo τ_{sw}, estimada como el inverso de la frecuencia de conmutación.

$$G_{\text{EL}}(s) = \frac{1}{r_{\text{Buck}}}\left(\frac{1}{\frac{l_{\text{Buck}}}{r_{\text{Buck}}\,\omega_b}\, s + 1}\right)\left(\frac{1}{\tau_{sw}\, s + 1}\right) \tag{3.35}$$

Se selecciona un controlador tipo PI para la regulación de corriente, cuya función de transferencia se presenta en (3.36). Las ganancias del controlador fueron ajustadas conforme a las directrices del criterio óptimo simétrico. Cabe destacar que el controlador cuenta con un esquema anti-windup, con el fin de mitigar los efectos derivados de la imposibilidad de aplicar una tensión superior a la disponible en el enlace de corriente directa.

$$C_{\mathrm{EL}}(s) = k_{c,\mathrm{EL}}\left(\frac{T_{n,\mathrm{EL}}\,s+1}{T_{n,\mathrm{EL}}\,s}\right) \qquad k_{c,\mathrm{EL}} = \frac{l_{\mathrm{Buck}}}{2\,\omega_b\,\tau_{sw}} \qquad T_{n,\mathrm{EL}} = 4\,\tau_{sw} \tag{3.36}$$

En la Figura 3.21 se muestra el desempeño del controlador de corriente del electrolizador, donde la corriente del equipo se incrementa desde el 50 % hasta el 100 % de su valor nominal. La referencia de control se limita a variaciones máximas del 10 % de la corriente nominal por segundo, con el fin de resguardar la integridad del electrolizador. Se observa que el desempeño del controlador es adecuado, pues existe seguimiento de referencia a lo largo de toda la maniobra. Adicionalmente, el seguimiento de referencia no se ve comprometido por las variaciones de la tensión en los terminales del electrolizador, lo que permite concluir que el rechazo a perturbaciones resulta efectivo.

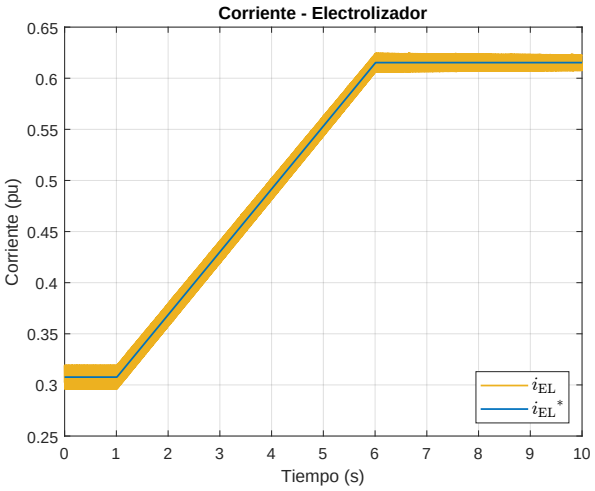

Figura 3.21: Seguimiento de referencia.

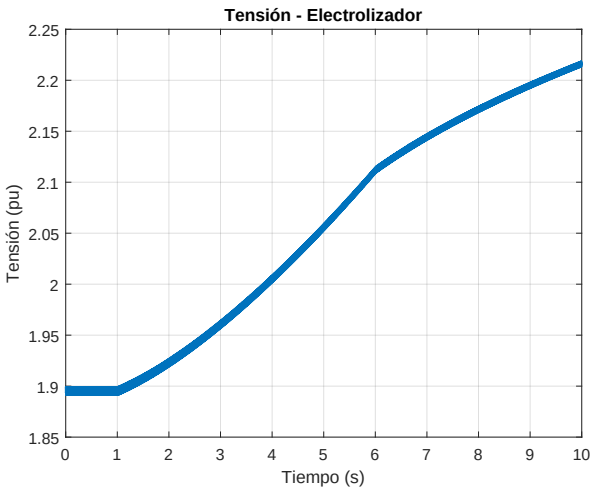

Figura 3.22: Tensión en terminales.

3.3. Celda de combustible

El caso de estudio considera una celda de combustible de membrana de electrolito polimérico, con una potencia preliminar de 1.25 MW destinada a la generación de energía eléctrica. Tal como se indicó previamente, esta elección se fundamenta en que dicha tecnología ofrece la respuesta más rápida entre las alternativas evaluadas. Cabe señalar que el comportamiento del dispositivo será descrito mediante el modelo presentado en la Figura 2.6.

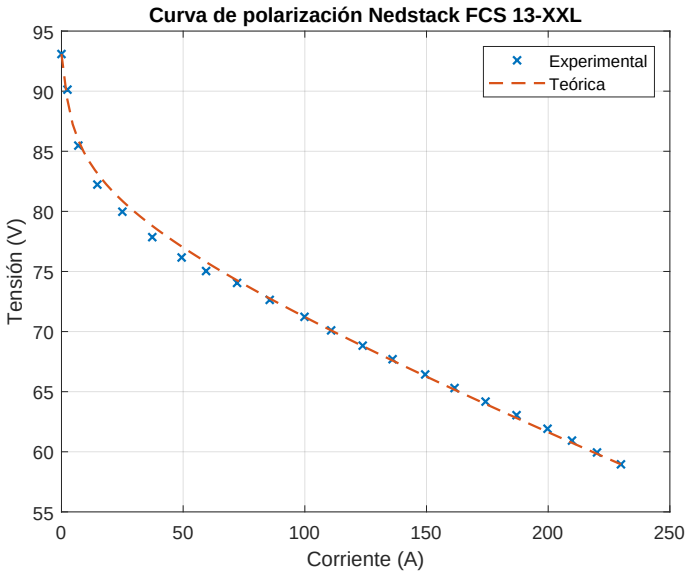

Figura 3.23: Curva de polarización de la celda de combustible Nedstack FCS 13-XXL.

En la Figura 3.23 se presentan las curvas de polarización experimental y teórica correspondientes a la celda de combustible Nedstack FCS 13-XXL, con una tensión y corriente nominal igual a 60 V y 230 A respectivamente. La curva teórica fue construida utilizando los parámetros indicados en la Tabla 3.8, determinados conforme al procedimiento descrito en [13].

Tabla 3.8: Parámetros teóricos Nedstack FCS 13-XXL.

Parámetro	**Unidad**	**Valor**
$E_{oc,\mathrm{Uni}}$	V	93.09
$\mathrm{NA}_{\mathrm{Uni}}$	V	2.85
$I_{o,\mathrm{Uni}}$	A	0.67
$R_{\Omega,\mathrm{Uni}}$	mΩ	76.08
T_d	s	5.00

La conexión en serie y/o en paralelo de múltiples unidades FCS 13-XXL permite la implementación de sistemas industriales de generación, como el Nedstack CHP-FCPS-600 con una potencia nominal de 600 kW. En este contexto, el modelo presentado en [13] es escalado considerando un conjunto de m ramas en paralelo, cada una compuesta por n celdas menores Nedstack FCS 13-XXL conectadas en serie. Los supuestos y desarrollos que permiten escalar el modelo a partir de los parámetros de la celda individual se detallan en el Apéndice B. La constante de tiempo T_d, obtenida directamente de la hoja de datos de la celda menor, no se ve modificada en el proceso de escalamiento. En la Ecuación (3.37) se presentan las expresiones que permiten ajustar los parámetros del modelo en función de la celda base y del arreglo considerado.

$$E_{oc,\mathrm{FC}} = n\,E_{oc,\mathrm{Uni}} \qquad I_{o,\mathrm{FC}} = m\,I_{o,\mathrm{Uni}} \qquad \mathrm{NA}_{\mathrm{FC}} = n\,\mathrm{NA}_{\mathrm{Uni}} \qquad R_{\Omega,\mathrm{FC}} = \frac{n}{m}\,R_{\Omega,\mathrm{Uni}} \tag{3.37}$$

En la Figura 3.24 se ilustra el sistema completo asociado a la celda de combustible, donde se observa que esta se conecta al enlace de corriente directa mediante un convertidor Boost y un filtro tipo L. La estrategia de control se basa en regular la corriente que circula a través del filtro, ya que de este modo se controla indirectamente la corriente suministrada por la celda, tal como se implementa en [41].

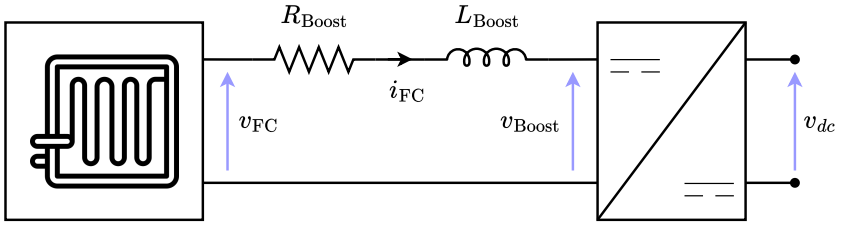

Figura 3.24: Conexión de la celda de combustible al enlace de corriente directa.

La potencia nominal de la celda de combustible mayor dependerá del arreglo escogido para el conjunto de celdas menores que la componen. Si se define una potencia nominal mínima $P_{\text{n,mín}}$ para el equipo en cuestión, las combinaciones de celdas que satisfacen lo requerido son aquellas que verifican (3.38).

$$m \geq \frac{P_{\text{n,mín}}}{n\, v_{\text{n,Uni}}\, i_{\text{n,Uni}}} \tag{3.38}$$

La capacidad del controlador para alcanzar el punto de operación nominal puede verse comprometida por una configuración $n \times m$ inadecuada. Con el fin de asegurar que el controlador pueda operar en el punto nominal, el ciclo de trabajo nominal D_n debe ser superior a cero e inferior a la unidad, excluyendo la unidad, imponiendo así dos restricciones adicionales al problema. Se considera una banda de tolerancia del $\pm\,5\,\%$ para la tensión del enlace de corriente directa, siendo la condición más desfavorable el límite inferior.

$$\begin{aligned} \frac{n\, v_{\text{n,Uni}} - v_{\text{dc,mín}}}{R_{\text{Boost}}\, i_{\text{n,Uni}}} &\leq m \\ \frac{n\, v_{\text{n,Uni}}}{R_{\text{Boost}}\, i_{\text{n,Uni}}} &> m \end{aligned} \tag{3.39}$$

Con el objetivo de limitar la potencia disipada en la componente resistiva del filtro, se debe establecer una cota de pérdidas asociada al punto de operación nominal. Dicha cota se define en función de un porcentaje de pérdidas objetivo $\%_{\text{Loss}}$ conforme a lo indicado en (3.40).

$$\frac{\%_{\text{Loss}}\, n\, v_{\text{n,Uni}}}{R_{\text{Boost}}\, i_{\text{n,Uni}}} \geq m \tag{3.40}$$

La potencia mínima de operación se alcanza cuando el ciclo de trabajo es nulo, ya que en dicha condición la tensión a la salida del convertidor Boost se maximiza, igualándose a la tensión del enlace de corriente directa. Con el fin de asegurar que el controlador sea capaz de anular la corriente entregada por la celda de combustible, la tensión en vacío de esta última no debe superar el valor mínimo esperado para la tensión del enlace, imponiendo una última restricción al problema.

$$n \leq \text{floor}\left(\frac{v_{\text{dc,mín}}}{E_{oc,\text{Uni}}}\right) \tag{3.41}$$

Presentadas las condiciones anteriores, se define que los grados de libertad de diseño serán la potencia nominal mínima $P_{\text{n,mín}}$, la resistencia del filtro R_{Boost} y el porcentaje de pérdidas objetivo $\%_{\text{Loss}}$. En la Figura 3.25 se presenta el mapa de combinaciones que permite visualizar los diferentes arreglos que satisfacen las condiciones anteriormente presentadas, considerando una potencia nominal mínima de 1.25 MW, pérdidas del $2\,\%$ y una resistencia del inductor igual a 0.5 mΩ.

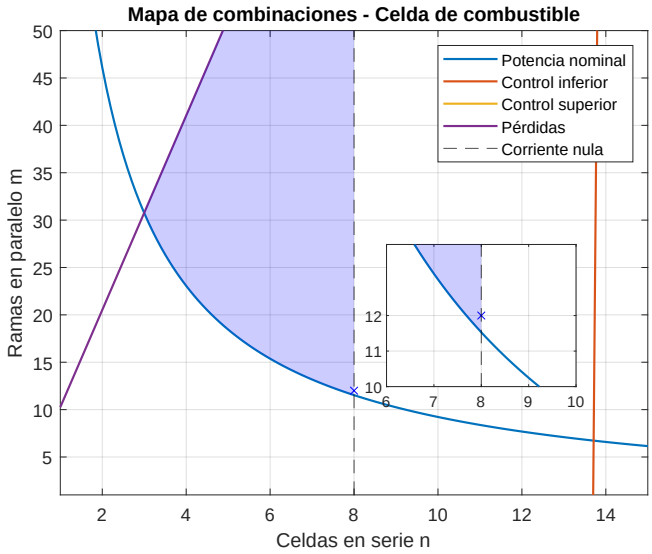

Figura 3.25: Mapa de combinaciones de la celda de combustible.

Se escoge el arreglo que considera 12 ramas en paralelo constituidas por 8 unidades en serie. Esta combinación cumple con todas las restricciones impuestas, encontrándose dentro del área factible denotada por color azul en la Figura 3.25. De esta manera, la celda de combustible se adjudica una potencia nominal de 1.3 MW con una tensión y corriente nominal de 470 V y 2760 A, mientras que sus parámetros equivalentes son los detallados en la Tabla 3.9.

Tabla 3.9: Parámetros de la celda de combustible diseñada.

Parámetro	**Unidad**	**Valor**
$E_{oc,\mathrm{FC}}$	V	744.71
$I_{o,\mathrm{FC}}$	A	8.08
$R_{\Omega,\mathrm{FC}}$	mΩ	50.72
$\mathrm{NA}_{\mathrm{FC}}$	V	22.82
$T_{d,\mathrm{FC}}$	s	5

La Ecuación (3.42) describe la dinámica de la corriente que circula por el filtro L, que es la misma que circula por la celda de combustible. Es importante destacar que esta expresión se encuentra normalizada de acuerdo con el sistema por unidad definido en el Apéndice A.

$$\frac{l_{\mathrm{Boost}}}{\omega_b}\frac{\mathrm{d}}{\mathrm{d}t}i_{\mathrm{FC}} + r_{\mathrm{Boost}}\, i_{\mathrm{FC}} = v_{\mathrm{FC}} - v_{\mathrm{Boost}} \tag{3.42}$$

La arquitectura del controlador es prácticamente equivalente a la implementada en el electrolizador, dado que la variable de actuación del lazo corresponde a la tensión de salida del convertidor Boost, mientras que la tensión en los terminales de la celda supone una perturbación. La evolución temporal de esta perturbación está determinada por la constante de tiempo T_d, por lo que, si el controlador de corriente se sintoniza con una respuesta significativamente más rápida, dicha perturbación tiende a comportarse como una señal cuasiestática desde la perspectiva del controlador, facilitando su supresión efectiva.

Se ha identificado que el rizado de corriente característico de los convertidores de potencia acelera la degradación de celdas de combustible tipo PEM, tal como ocurre en electrolizadores del mismo tipo [43]. En consecuencia, el criterio adoptado para dimensionar el inductor del filtro se basa en el enfoque clásico utilizado en convertidores Boost, aunque con un margen de diseño conservador. El dimensionamiento se lleva a cabo de acuerdo con (3.43), donde v_{in} corresponde a la tensión de entrada del convertidor, D al ciclo de trabajo, f_{sw} a la frecuencia de conmutación, ΔI al rizado de corriente admitido y v_{out} a la tensión de salida del convertidor.

$$L_{\text{Boost}} = \frac{v_{\text{in}}\, D}{f_{sw}\, \Delta I} \qquad\qquad D = 1 - \frac{v_{\text{in}}}{v_{\text{out}}} \tag{3.43}$$

El ciclo de trabajo es calculado considerando que la tensión de entrada es igual a la tensión nominal de la celda de combustible, mientras que la tensión a la salida es igual a la tensión del enlace de corriente directa. Se admitirá un rizado máximo del 1 % de la corriente nominal del equipo, operando a una frecuencia de conmutación de 10 kHz. Haciendo uso de (3.43) el dimensionamiento resulta en un inductor de 761 μH.

La planta asociada al control de corriente es la presentada en (3.44), en la que se ha incluido el retardo asociado a la acción del convertidor mediante la constante de tiempo τ_{sw}, estimado como el inverso de la frecuencia de conmutación.

$$G_{\text{FC}}(s) = -\frac{1}{r_{\text{Boost}}}\left(\frac{1}{\frac{l_{\text{Boost}}}{r_{\text{Boost}}\,\omega_b}\,s + 1}\right)\left(\frac{1}{\tau_{sw}\,s + 1}\right) \tag{3.44}$$

Se selecciona un controlador tipo PI para la regulación de corriente, cuya función de transferencia se presenta en (3.45). Las ganancias fueron ajustadas conforme a las directrices del criterio óptimo simétrico. Es importante señalar que el controlador incorpora un esquema anti-windup, con el propósito de mitigar los efectos asociados a la imposibilidad de aplicar una tensión superior a la disponible en el enlace de corriente directa.

$$C_{\text{FC}}(s) = k_{c,\text{FC}}\left(\frac{T_{n,\text{FC}}\,s + 1}{T_{n,\text{FC}}\,s}\right) \qquad k_{c,\text{FC}} = -\frac{l_{\text{Boost}}}{2\,\omega_b\,\tau_{sw}} \qquad T_{n,\text{FC}} = 4\,\tau_{sw} \tag{3.45}$$

En la Figura 3.26 se muestra el desempeño del controlador de corriente de la celda de combustible, donde la corriente del equipo se incrementa desde el 50 % hasta el 100 % de su valor nominal. La referencia de control se limita a variaciones máximas del 10 % de la corriente nominal por segundo, con el fin de resguardar la integridad de la celda. Se observa que el desempeño del controlador es adecuado, pues existe seguimiento de referencia a lo largo de toda la maniobra. Adicionalmente, el seguimiento de referencia no se ve comprometido por las variaciones de la tensión en los terminales de la celda de combustible, lo que permite concluir que el rechazo a perturbaciones resulta efectivo.

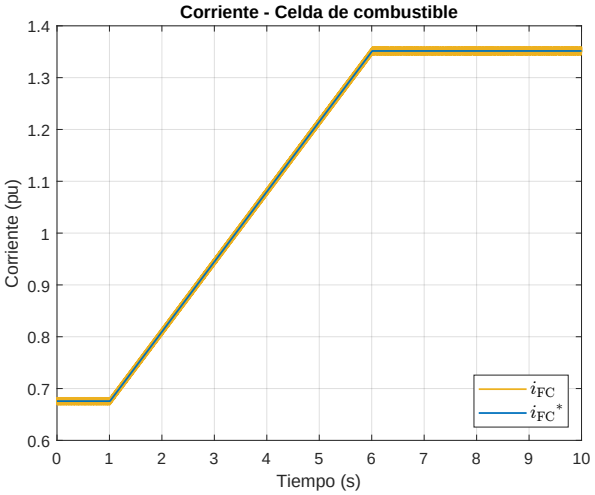

Figura 3.26: Seguimiento de referencia.

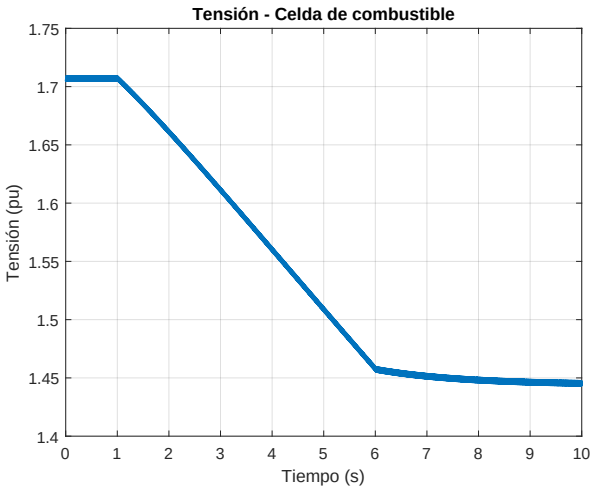

Figura 3.27: Tensión en terminales.

3.4. Eficiencia y servicios auxiliares

Los servicios auxiliares de una celda de combustible y un electrolizador cumplen un papel clave en el funcionamiento seguro y estable de estos sistemas. Incluyen equipos como bombas, compresores, sistemas de enfriamiento, controladores y dispositivos de purga, cuya función es mantener las condiciones adecuadas de presión, temperatura y pureza de los gases. De esta manera, permiten que los procesos electroquímicos se desarrollen de forma estable y aseguran una respuesta confiable ante variaciones de carga o condiciones operativas. En el contexto del presente trabajo, estos servicios se modelarán como una carga constante siempre que la potencia de operación del equipo sea inferior al 15 % de su potencia nominal; en caso contrario, su consumo aumentará linealmente desde un 5 % hasta un 10 % de la potencia nominal [44].

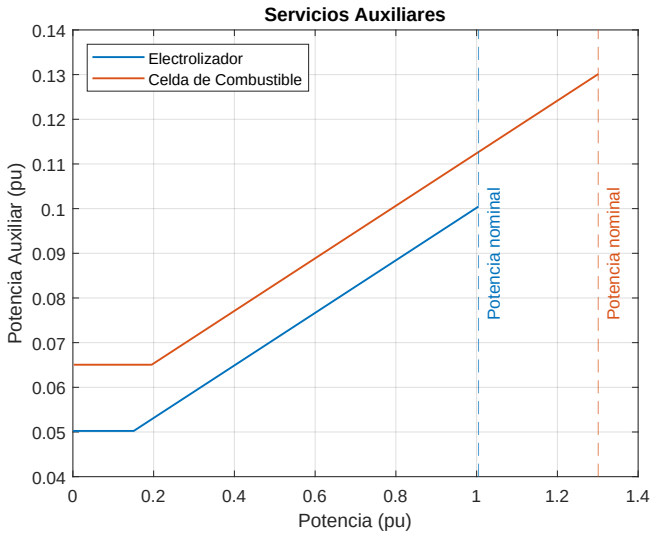

Figura 3.28: Servicios auxiliares.

Con los servicios auxiliares previamente definidos, la eficiencia global de cada sistema se evalúa considerando su contribución al consumo total de energía. En particular, la eficiencia del electrolizador se estima de acuerdo con (2.4), mientras que la de la celda de combustible se determina conforme a (2.10). De este modo se obtiene una representación más realista del desempeño energético de cada equipo, al incorporar las pérdidas adicionales asociadas al funcionamiento de sus servicios auxiliares.

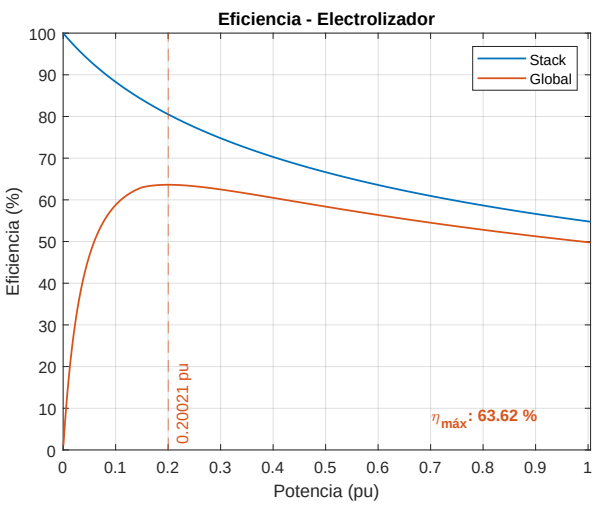

Figura 3.29: Eficiencia del electrolizador.

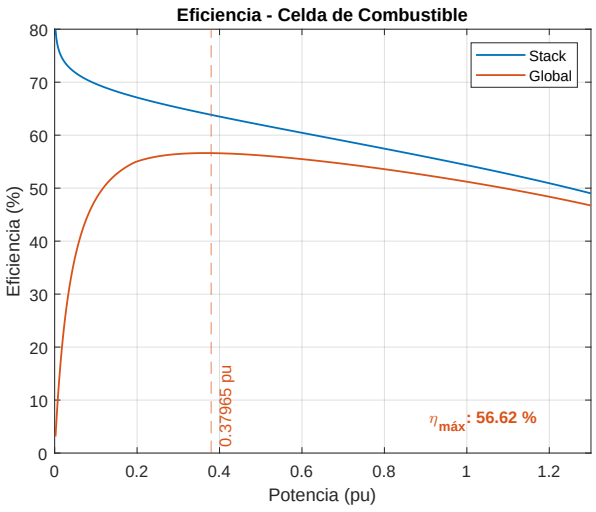

Figura 3.30: Eficiencia de la celda de combustible.

3.5. Batería

El caso de estudio considera un sistema de almacenamiento basado en baterías con celdas del tipo ion-litio, con una potencia preliminar de 1 MW para la generación y almacenamiento de energía eléctrica. Como se mencionó anteriormente, esta definición se sustenta en que esta tecnología ofrece una respuesta rápida y de alta eficiencia. El comportamiento del dispositivo será descrito utilizando el modelo presentado en la Figura 2.7, y sus parámetros serán obtenidos a partir de una celda real de acuerdo con el procedimiento definido en [18].

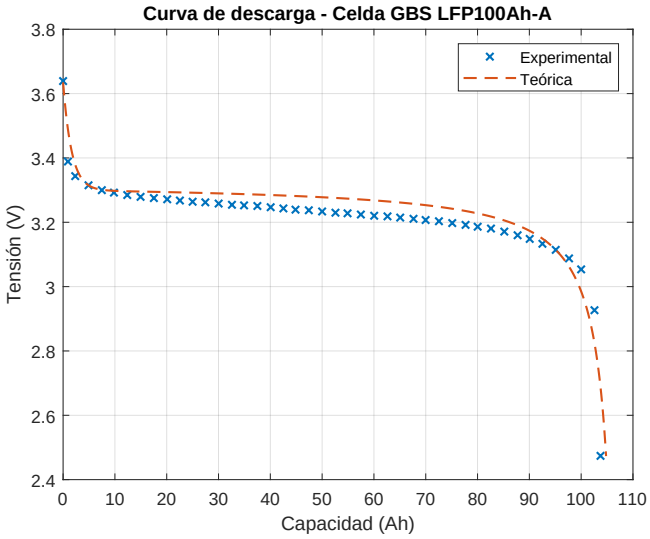

Figura 3.31: Curva de descarga de la celda GBS LFP100Ah-A.

En la Figura 3.31 se presentan las curvas de descarga experimental y teórica correspondientes a la celda ion-litio GBS LFP100Ah-A, con una tensión y corriente nominal igual a 3.2 V y 100 A respectivamente. La curva teórica fue construida utilizando los parámetros indicados en la Tabla 3.10, determinados conforme al procedimiento descrito en [18].

Tabla 3.10: Parámetros teóricos GBS LFP100Ah-A.

Parámetro	**Unidad**	**Valor**
E_0	V	3.311
K	mV/Ah	0.197
Q	Ah	108.00
A	V	0.336
B	Ah^{-1}	0.611
R_B	mΩ	0.40

Incorporar un sistema de baterías en la planta de almacenamiento tiene como propósito compensar la lenta respuesta de los sistemas basados en hidrógeno, por lo que inicialmente no se espera que este asuma funciones de generación o carga de largo plazo. Más aún, cuando la capacidad de la batería es alta, el estado de carga apenas varía en intervalos temporales pequeños, representando también un desafío computacional para el entrenamiento del agente de control de energía. Para observar fluctuaciones significativas en el estado de carga sería necesario realizar simulaciones de gran duración, que eventualmente superarán la capacidad de cómputo disponible si se busca capturar también los transitorios de la planta. La celda unitaria considerada puede suministrar su corriente nominal de manera continua durante 60 minutos, lo que implica que cualquier variación sustancial en el estado de carga requeriría horizontes de simulación demasiado extensos. Por ambas razones, se opta por reducir artificialmente la capacidad de la celda modificando los parámetros K y Q,

de tal forma que la pendiente relativa de la curva de descarga permanezca inalterada. Así, para una nueva capacidad máxima Q', la constante de polarización redefinida K' se determina mediante (3.46).

$$K' = K\,\frac{Q}{Q'} \tag{3.46}$$

La reducción artificial de la capacidad de la celda unitaria se establece bajo ciertos supuestos. En primer lugar, únicamente se modifica la capacidad máxima de la celda, de modo que sus propiedades químicas no se ven alteradas, ya que no se modifica ni la tensión ni la capacidad exponencial. En segundo lugar, la resistencia interna y los valores nominales de la celda se mantienen constantes. Estos supuestos son adecuados para los objetivos del presente trabajo, pues en el contexto del desarrollo del controlador de energía, lo crucial es capturar la relación entre el estado de carga y la tensión en los terminales de la batería, así como su comportamiento como almacenador de energía. De este modo, el nivel de detalle alcanzado en esta etapa se considera suficiente, mientras que un diseño más preciso se reserva para estudios futuros. En la Tabla 3.11 se informan los parámetros considerados para la celda unitaria redefinida, cuya capacidad constituye un 16.7 % de la capacidad máxima original. Adicionalmente, en la Figura 3.32 se presenta su curva de descarga.

Tabla 3.11: Parámetros teóricos de la celda redefinida.

Parámetro	**Unidad**	**Valor**
E_0	V	3.311
K'	mV/Ah	1.179
Q'	Ah	18.00
A	V	0.336
B	Ah^{-1}	0.611
R_B	mΩ	0.40

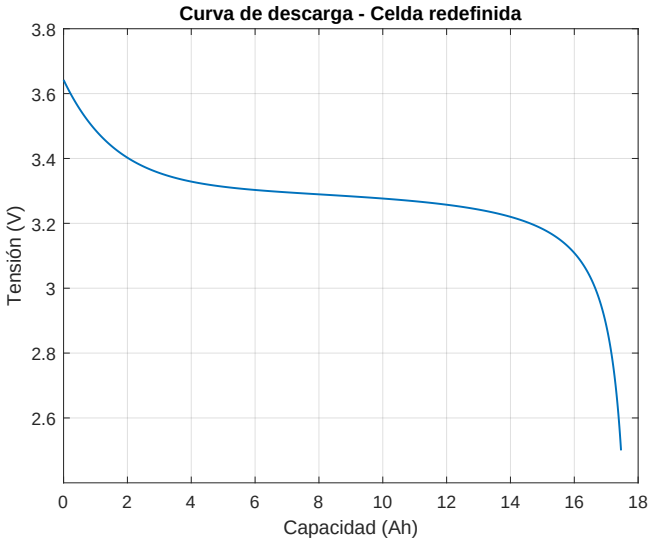

Figura 3.32: Curva de descarga de la celda redefinida.

Debido al carácter teórico del caso de estudio, no es posible estimar experimentalmente la tensión de circuito abierto de la celda estudiada. Como alternativa, se propone inferir dicha tensión a partir del modelo presentado en [18] imponiendo corriente nula. Esto permite expresar la tensión de circuito abierto en función del estado de carga de la batería, tal como se indica en (3.47).

$$E_{\mathrm{B}}(\mathrm{SOC}_{\,\%}) = E_0 - K'\,Q'\,\frac{100 - \mathrm{SOC}_{\,\%}}{\mathrm{SOC}_{\,\%}} + A\exp\left(-B\,Q'\,\frac{100 - \mathrm{SOC}_{\,\%}}{100}\right) \tag{3.47}$$

En la Figura 3.33 se muestra la relación entre la tensión de circuito abierto y el estado de carga de la celda. Es importante señalar que esta aproximación queda indefinida cuando el estado de carga es nulo, lo que no refleja el comportamiento real de las baterías. Por lo tanto, la aproximación propuesta no es válida para estados de carga bajos. No obstante, la literatura indica que el rango óptimo de operación para sistemas de almacenamiento basados en celdas de ion-litio se encuentra entre el 20 % y el 80 % del estado de carga, ya que dentro de este intervalo se maximiza la vida útil de las celdas [45]. Bajo esta consideración, se establece que el sistema de almacenamiento no reducirá su estado de carga por debajo del 20 % ni lo incrementará por encima del 80 %, lo que permite validar cualitativamente la aproximación propuesta.

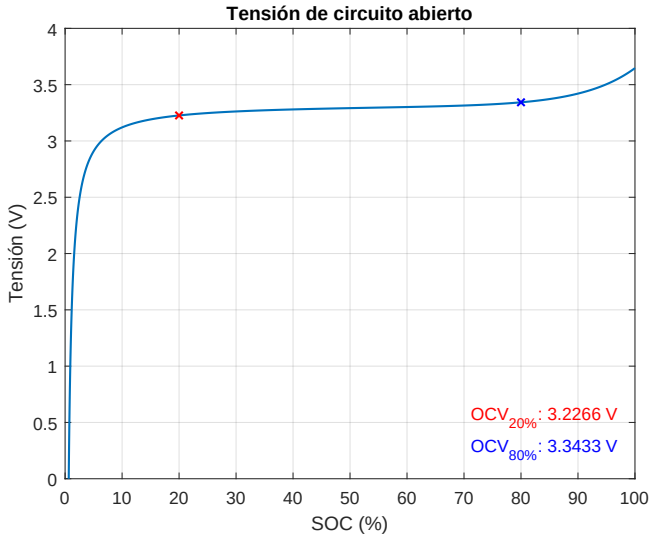

Figura 3.33: Tensión de circuito abierto estimada.

En la Figura 3.34 se ilustra el sistema de almacenamiento estudiado, advirtiendo que el banco de baterías se conecta al enlace de corriente directa por medio de un convertidor Buck – Boost y un filtro L. La estrategia de control consiste en controlar la corriente a través del filtro, pues de esta forma indirectamente se controla la corriente que circula por el banco de baterías.

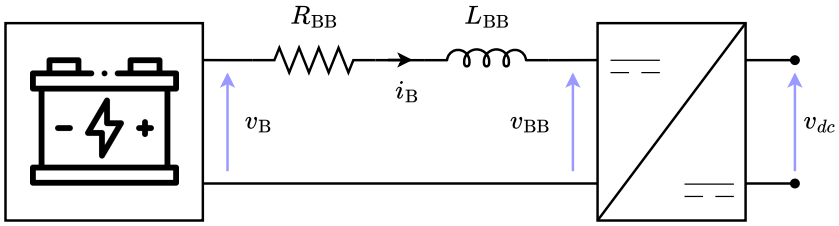

Figura 3.34: Conexión del banco de baterías al enlace de corriente directa.

La celda unitaria evidentemente no cubre la potencia nominal definida para el sistema de almacenamiento en baterías, por lo que es necesario escalar el modelo tal como se hizo para la celda de combustible y el electrolizador, considerando m ramas en paralelo constituidas por n celdas unitarias en serie. En la Ecuación (3.48) se presentan las expresiones que permiten ajustar los parámetros y variables del modelo en función de la celda base y del arreglo considerado.

$$E_{\mathrm{B,BESS}} = n\, E_{\mathrm{B,Uni}} \qquad R_{\mathrm{B,BESS}} = \frac{n}{m}\, R_{\mathrm{B,Uni}} \qquad i_{\mathrm{B,BESS}} = m\, i_{\mathrm{B,Uni}} \tag{3.48}$$

La potencia nominal del sistema de almacenamiento en baterías dependerá del arreglo escogido para el conjunto de celdas que lo componen. Si se define una potencia nominal mínima $P_{\mathrm{n,mín}}$ para el equipo en cuestión, las combinaciones de celdas que satisfacen lo requerido son aquellas que verifican (3.49).

$$m \geq \frac{P_{\mathrm{n,mín}}}{n\, v_{\mathrm{n,Uni}}\, i_{\mathrm{n,Uni}}} \tag{3.49}$$

La capacidad del controlador para alcanzar el punto de operación nominal puede verse comprometida por una configuración $n \times m$ inadecuada. Se establece como criterio que el controlador debe ser capaz de imponer corriente nominal a lo largo de todo el rango operativo comprendido entre un estado de carga superior al 20 % e inferior al 80 %, tanto en régimen de carga como de descarga. La peor condición en descarga ocurre cuando el SOC es igual a 20 %, pues la tensión interna es mínima. En esta condición es necesario que la tensión a la salida del convertidor sea mínima pero no nula, pues de ser el caso no existe transferencia de potencia al enlace de corriente directa. Dado el escenario descrito, se define la siguiente condición presentada en (3.50).

$$\frac{n\,\mathrm{OCV}_{20\,\%} - 0{,}1\, v_{\mathrm{dc,mín}}}{\frac{n}{m} R_{\mathrm{B,Uni}} + R_{\mathrm{BB}}} \geq m\, i_{\mathrm{n,Uni}} \tag{3.50}$$

La peor condición en carga ocurre cuando el SOC es igual a 80 %, pues la tensión interna es máxima. En esta condición es necesario que la tensión a la salida del convertidor sea máxima pero no igual a la tensión del enlace de corriente directa, con la finalidad de no acotar estrictamente el rango de actuación. Dado el escenario descrito, se define una tercera condición para el problema presentada en (3.51).

$$\frac{0{,}9\, v_{\mathrm{dc,mín}} - n\,\mathrm{OCV}_{80\,\%}}{\frac{n}{m} R_{\mathrm{B,Uni}} + R_{\mathrm{BB}}} \geq m\, i_{\mathrm{n,Uni}} \tag{3.51}$$

Con el objetivo de limitar la potencia disipada en la componente resistiva del filtro, se debe establecer una cota de pérdidas asociada al punto de operación nominal. Dicha cota se define en función de un porcentaje de pérdidas objetivo $\%_{\mathrm{Loss}}$ conforme a lo indicado en (3.52).

$$\frac{\%_{\mathrm{Loss}}\, n\, v_{\mathrm{n,Uni}}}{R_{\mathrm{BB}}\, i_{\mathrm{n,Uni}}} \geq m \tag{3.52}$$

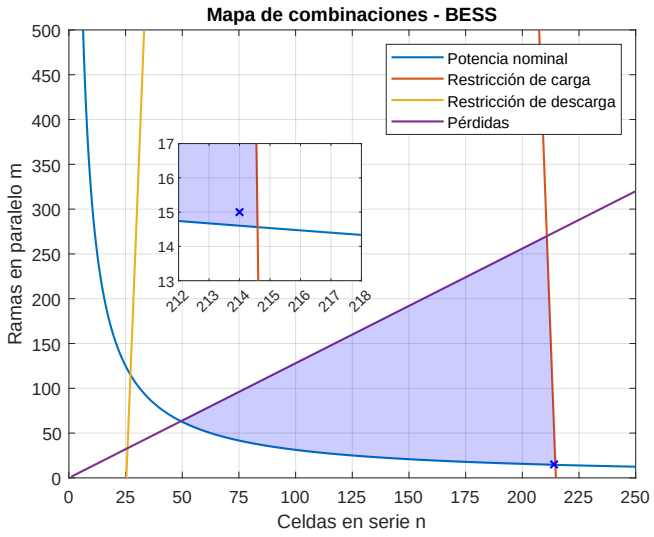

Figura 3.35: Mapa de combinaciones del sistema de almacenamiento en baterías.

Presentadas las condiciones anteriores, se define que los grados de libertad de diseño serán la potencia nominal mínima $P_{\mathrm{n,mín}}$, la resistencia del filtro R_{BB} y el porcentaje de pérdidas objetivo $\%_{\mathrm{Loss}}$. En la Figura 3.35 se presenta el mapa de combinaciones que permite visualizar los diferentes arreglos que satisfacen las condiciones anteriormente presentadas, considerando una potencia nominal mínima de 1 MW, pérdidas del 2 % y una resistencia del inductor igual a 0.5 mΩ. Se escoge el arreglo que considera 15 ramas en paralelo constituidas por 214 unidades en serie. Esta combinación cumple con todas las restricciones impuestas, encontrándose dentro del área factible denotada por color azul en la Figura 3.35. De esta manera, el sistema de almacenamiento en baterías se adjudica una potencia nominal de 1 MW con una tensión y corriente nominal de 685 V y 1500 A. Adicionalmente, se obtiene una resistencia interna equivalente de 5.71 mΩ.

La Ecuación (3.53) describe la dinámica de la corriente que circula por el filtro L, que es la misma que circula por la batería. Es importante destacar que esta expresión se encuentra normalizada de acuerdo con el sistema por unidad definido en el Apéndice A.

$$\frac{l_{\mathrm{BB}}}{\omega_b}\,\frac{\mathrm{d}}{\mathrm{d}t} i_{\mathrm{B}} + r_{\mathrm{BB}}\, i_{\mathrm{B}} = v_{\mathrm{B}} - v_{\mathrm{BB}} \tag{3.53}$$

La arquitectura del controlador es idéntica a la implementada en la celda de combustible, dado que la variable de actuación del lazo corresponde a la tensión de salida del convertidor Buck – Boost, mientras que la tensión en los terminales de la batería se considera una perturbación. La evolución temporal de esta perturbación está determinada por la constante de tiempo T_{B}, por lo que, si el controlador de corriente se sintoniza con una respuesta significativamente más rápida, dicha perturbación tiende a comportarse como una señal cuasiestática desde la perspectiva del controlador, facilitando su supresión efectiva.

El dimensionamiento del inductor del filtro se realiza considerando el comportamiento dual del convertidor, el cual funciona como Buck durante la fase de carga y como Boost durante la fase de descarga. En el primer caso, la tensión de salida v_{out} coincide con la tensión nominal de la batería, mientras que en el segundo, es la tensión de entrada v_{in} la que se iguala a dicho valor, de acuerdo con los criterios clásicos de dimensionamiento.

$$L_{\mathrm{BB}} = \mathrm{máx}\left(\frac{v_{\mathrm{n,B}}\, D}{f_{sw}\, \Delta I}; \frac{v_{\mathrm{n,B}}\,(1-D)}{f_{sw}\, \Delta I}\right) \tag{3.54}$$

La elección del ciclo de trabajo responde a un compromiso equilibrado entre ambos modos de operación del convertidor, por lo que se adopta el valor $D = 0{,}5$. En estas condiciones, la expresión utilizada para dimensionar la inductancia del filtro L se presenta en (3.55). Se considera un rizado máximo permitido del 5 % respecto de la corriente nominal del equipo operando a una frecuencia de conmutación de 10 kHz, lo que conduce al dimensionamiento de un inductor de 460 μH.

$$L_{\mathrm{BB}} = \frac{v_{\mathrm{n,B}}}{2\, f_{sw}\, \Delta I} \tag{3.55}$$

La planta asociada al control de corriente es la presentada en (3.56), en la que se ha incluido el retardo asociado a la acción del convertidor mediante la constante de tiempo τ_{sw}, estimado como el inverso de la frecuencia de conmutación.

$$G_{\mathrm{B}}(s) = -\frac{1}{r_{\mathrm{BB}}}\left(\frac{1}{\frac{l_{\mathrm{BB}}}{r_{\mathrm{BB}}\,\omega_b}\, s + 1}\right)\left(\frac{1}{\tau_{sw}\, s + 1}\right) \tag{3.56}$$

Se selecciona un controlador tipo PI para la regulación de corriente, cuya función de transferencia se presenta en (3.57). Las ganancias fueron ajustadas conforme a las directrices del criterio óptimo simétrico. Es importante señalar que el controlador incorpora un esquema anti-windup, con el propósito de mitigar los efectos asociados a la imposibilidad de aplicar una tensión superior a la disponible en el enlace de corriente directa.

$$C_{\mathrm{B}}(s) = k_{c,\mathrm{B}}\left(\frac{T_{n,\mathrm{B}}\, s + 1}{T_{n,\mathrm{B}}\, s}\right) \qquad k_{c,\mathrm{B}} = -\frac{l_{\mathrm{BB}}}{2\,\omega_b\,\tau_{sw}} \qquad T_{n,\mathrm{B}} = 4\,\tau_{sw} \tag{3.57}$$

En la Figura 3.36 se muestra el desempeño del controlador de corriente de la batería, donde inicialmente la corriente del equipo es nominal en régimen de carga, para posteriormente cambiar a corriente nominal en régimen de descarga. Se observa que el desempeño del controlador es adecuado, pues existe seguimiento de referencia a lo largo de toda la maniobra. Adicionalmente, el seguimiento de referencia no se ve comprometido por las variaciones de la tensión en los terminales de la batería, lo que permite concluir que el rechazo a perturbaciones resulta efectivo.

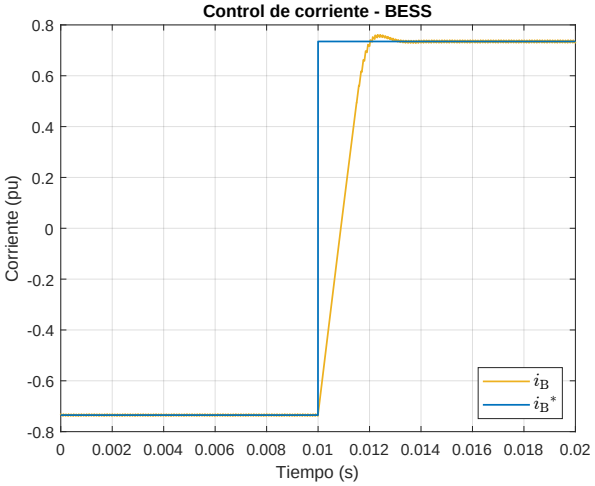

Figura 3.36: Seguimiento de referencia.

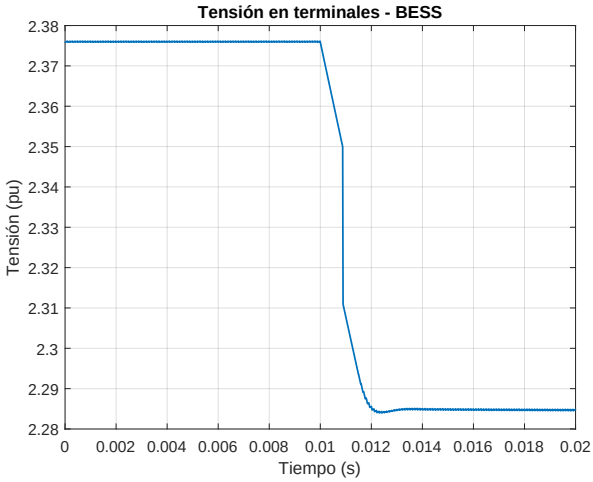

Figura 3.37: Tensión en terminales.

3.6. Enlace de corriente directa

El enlace de corriente directa permite la transferencia de energía entre los sistemas que conforman la planta de almacenamiento, utilizando un capacitor que actúa como interfaz para los convertidores estáticos. La Figura 3.38 muestra la conexión del convertidor formador de red y los sistemas de almacenamiento al enlace de corriente directa.

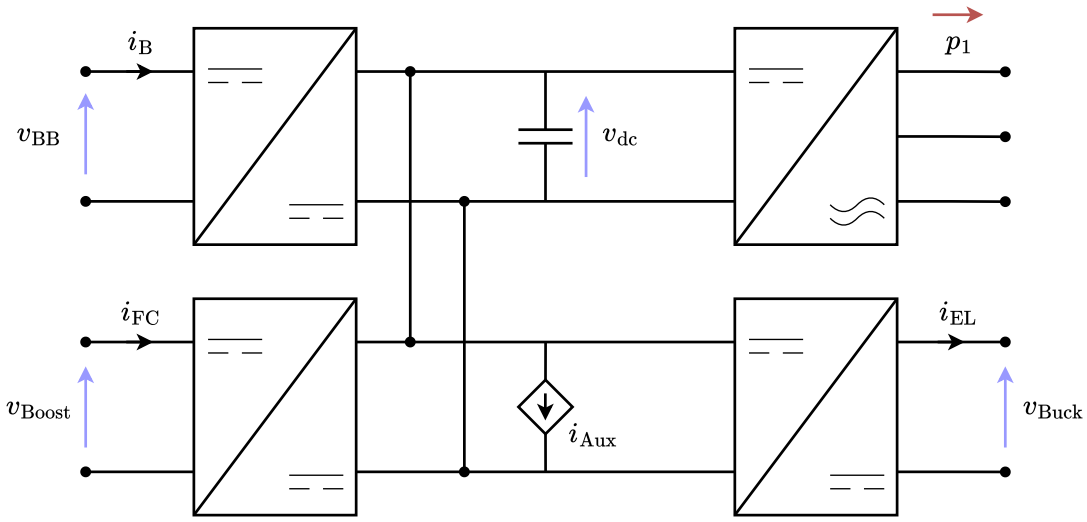

Figura 3.38: Enlace de corriente directa.

La potencia asociada a la operación de los servicios auxiliares se modela como una fuente de corriente controlada, la cual se conecta en polaridad inversa al capacitor del enlace de corriente directa. La corriente instantánea en por unidad a raíz del consumo auxiliar queda dada por

$$i_{\mathrm{Aux}} = \frac{3\,p_{\mathrm{Aux}}}{2\,v_{\mathrm{dc}}}$$

Despreciando las pérdidas de los convertidores, la potencia instantánea inyectada o retirada del enlace por cada sistema es en por unidad

$$p_{\mathrm{B}} = \frac{2}{3}\,v_{\mathrm{BB}}\,i_{\mathrm{B}} \qquad p_{\mathrm{FC}} = \frac{2}{3}\,v_{\mathrm{Boost}}\,i_{\mathrm{FC}} \qquad p_{\mathrm{EL}} = \frac{2}{3}\,v_{\mathrm{Buck}}\,i_{\mathrm{EL}}$$

La evolución temporal de la tensión del capacitor queda dada por (3.58), expresión que se encuentra normalizada de acuerdo con el sistema por unidad definido en el Anexo A. Es importante advertir que para satisfacer el balance de potencia, la tensión en el enlace de corriente directa debe permanecer constante.

$$\frac{c}{3\,\omega_b}\frac{\mathrm{d}}{\mathrm{d}t}({v_{\mathrm{dc}}}^2) = p_{\mathrm{B}} + p_{\mathrm{FC}} - p_{\mathrm{EL}} - p_{\mathrm{Aux}} - p_1 \tag{3.58}$$

Además de la necesidad de satisfacer el balance de potencia, la correcta modulación de los convertidores depende de que exista un nivel de tensión específico en el enlace. Por ello, es necesario controlar activamente la magnitud de la tensión del capacitor. Este control es llevado a cabo por el sistema de almacenamiento basado en baterías, el cual puede compensar rápidamente los desbalances causados por perturbaciones en la red. La variable de control será el cuadrado de la tensión, por lo que convenientemente se sustituye

$${v_{\mathrm{dc}}}^2 = \mu$$

La Ecuación (3.58) es linealizada en torno a un punto de operación según

$$\frac{c}{3\,\omega_b}\frac{\mathrm{d}}{\mathrm{d}t}\Delta\mu = \frac{2}{3}\Big(v_{\mathrm{BB},0}\,\Delta i_{\mathrm{B}} + i_{\mathrm{B},0}\,\Delta v_{\mathrm{BB}}\Big) + \underbrace{\Delta p_{\mathrm{FC}} - \Delta p_{\mathrm{EL}} - \Delta p_{\mathrm{Aux}} - \Delta p_1}_{\Delta p_{\mathrm{D}}} \tag{3.59}$$

A partir del lazo de control de corriente del sistema de almacenamiento en baterías, es posible definir las funciones de transferencia $H(s)$ y $T(s)$ tal como se muestra a continuación.

$$i_{\mathrm{B}}(s) = H(s)\,i_{\mathrm{B}}(s)^* \qquad\qquad v_{\mathrm{BB}}(s) = T(s)\;i_{\mathrm{B}}(s)^* \tag{3.60}$$

Al aplicar la transformada de Laplace sobre (3.59) y luego sustituir (3.60), es posible obtener una función de transferencia entre la referencia de corriente del sistema de almacenamiento basado en baterías y el cuadrado de la tensión del enlace en corriente directa. De esta manera, se define que la referencia de corriente mencionada constituye la actuación del controlador de la tensión del enlace de corriente directa, mientras que los términos ajenos al aporte de potencia por parte del sistema de baterías suponen una perturbación para el lazo. La planta de control es

$$\Delta\mu(s) = \frac{1}{s}\underbrace{\frac{2\,\omega_b}{c}\Big[v_{\mathrm{BB},0}\,H(s) + i_{\mathrm{B},0}\,T(s)\Big]}_{k_{\mathrm{eff}}(s)}\Delta {i_{\mathrm{B}}}^* \tag{3.61}$$

Es importante destacar que la ganancia dinámica $k_{\mathrm{eff}}(s)$ depende del tiempo de asentamiento del lazo de control de corriente del sistema de almacenamiento basado en baterías. Por lo tanto, si el lazo de control de tensión es significativamente más lento que el de corriente, $k_{\mathrm{eff}}(s)$ puede considerarse prácticamente constante desde la perspectiva del controlador de tensión. Si la condición se cumple, es posible aproximar

$$k_{\mathrm{eff}}(s) \sim k_{\mathrm{eff}}(0)$$

En la Figura 3.39 se muestra cómo la ganancia dinámica varía en función del punto de operación y del estado de carga del sistema de almacenamiento basado en baterías. Se observa una clara dependencia respecto del punto de operación, pues en mayor o menor medida dependiendo del estado de carga, la ganancia dinámica disminuye conforme se transita desde el régimen de carga al régimen de descarga. Este resultado pone de manifiesto la necesidad de implementar una estrategia adaptativa para garantizar un control de tensión adecuado.

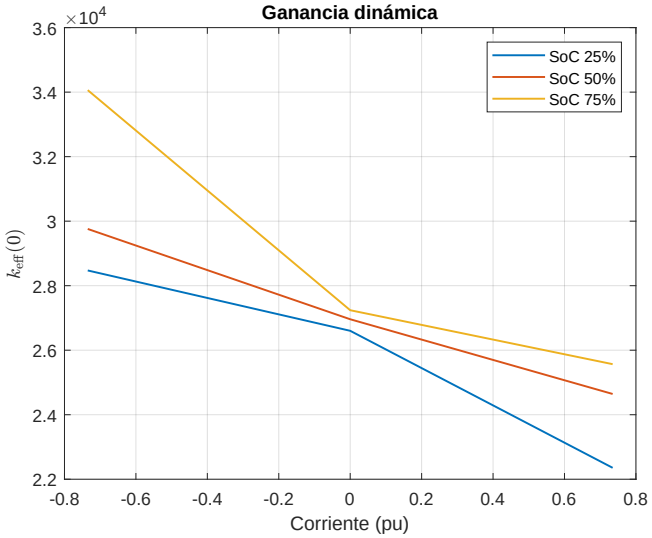

Figura 3.39: Ganancia dinámica.

Adicionalmente se observa que el estado de carga también influye en el valor de la ganancia dinámica, pues para un mismo punto de operación adopta diferentes valores. Con la finalidad de evaluar si ambas variables son igual de determinantes en la dinámica de la planta, en las Figuras 3.40 y 3.41 se presentan respectivamente las respuestas a escalón de la planta linealizada para un estado de carga fijo y para un punto de operación fijo.

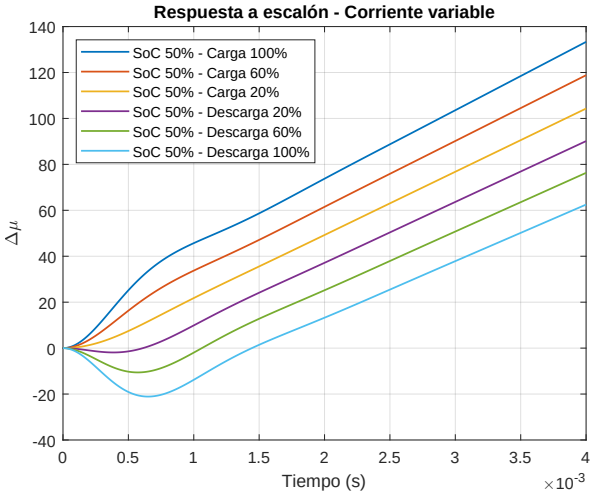

Figura 3.40: Respuesta a escalón - SoC fijo.

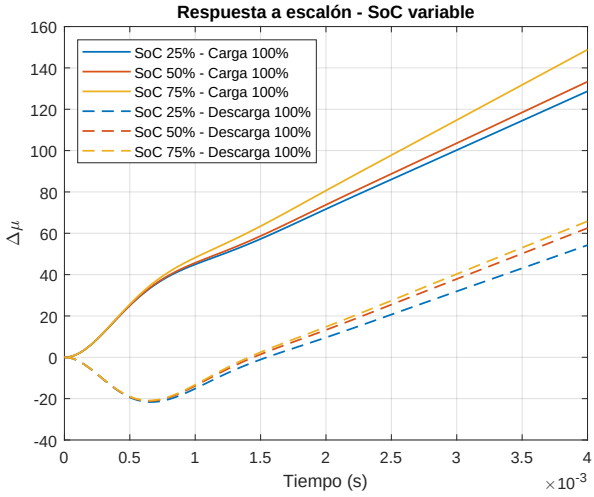

Figura 3.41: Respuesta a escalón - i_{B} fijo.

A partir de los gráficos previamente presentados, se concluye que el punto de operación es mucho más significativo que el estado de carga en la dinámica temprana de la planta, pues al sensibilizar esta última variable para un mismo punto de operación la respuesta a escalón no es sustancialmente diferente. De esta manera, se define que las ganancias del controlador serán ajustadas bajo la regla adaptativa impuesta por la curva de ganancia dinámica para un estado de carga igual al 50 %. Entonces, la sintonización del controlador queda

$$k_{p,\mathrm{dc}} = \frac{2\,\xi_{\mathrm{dc}}\,\omega_{n,\mathrm{dc}}}{k_{\mathrm{eff}}(i_{\mathrm{B}})} \qquad\qquad k_{i,\mathrm{dc}} = \frac{{\omega_{n,\mathrm{dc}}}^2}{k_{\mathrm{eff}}(i_{\mathrm{B}})}$$

donde ξ_{dc} y $\omega_{n,\mathrm{dc}}$ corresponden respectivamente al coeficiente de amortiguamiento base del lazo y a la frecuencia natural base del lazo.

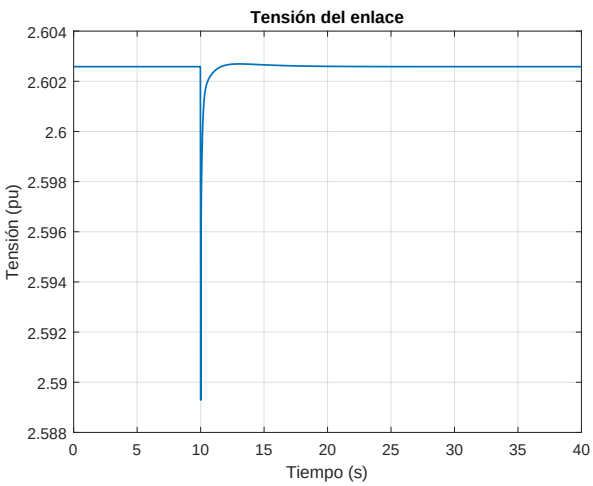

Figura 3.42: Tensión del enlace.

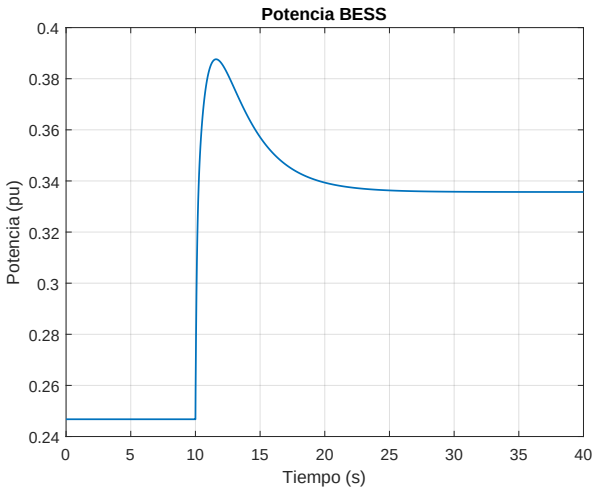

Figura 3.43: Potencia del banco de baterías.

En la Figura 3.42 se muestra el comportamiento de la tensión del enlace frente a un incremento en la potencia demandada por el convertidor formador de red en el instante $t = 10\,\text{s}$, evidenciando un correcto seguimiento de la referencia. Adicionalmente, la Figura 3.43 presenta la potencia suministrada por el sistema de baterías, la cual se incrementa para compensar el desbalance energético introducido por la red. Estos resultados se obtuvieron utilizando un coeficiente de amortiguamiento base de 0.707 y una frecuencia natural base de 600 rad/s, lo que permitió una operación estable del lazo de control de tensión.

La simulación del enlace de corriente directa se realizó empleando un modelo promedio de la planta de almacenamiento, centrado en el comportamiento global del sistema sin considerar los efectos de alta frecuencia, como la conmutación. Por esta razón, en las Figuras 3.42 y 3.43 no se observa ruido en las señales. Un modelo que incluya la conmutación implicaría una elevada demanda computacional, lo que resulta problemático para el entrenamiento del agente, pues requeriría tiempos de simulación extremadamente largos por episodio. Además, en la práctica esto no se justifica, ya que es la dinámica de baja frecuencia la que determina el intercambio de potencia entre los sistemas de almacenamiento y el convertidor. En la Figura 3.44 son comparados los modelos promedio y conmutado de la celda de combustible, el electrolizador y el sistema de baterías; como se puede observar, el modelo promedio describe de manera suficiente la dinámica de los sistemas conmutados. Dado este resultado, los entrenamientos del agente serán llevados a cabo en el modelo promedio de la planta de almacenamiento.

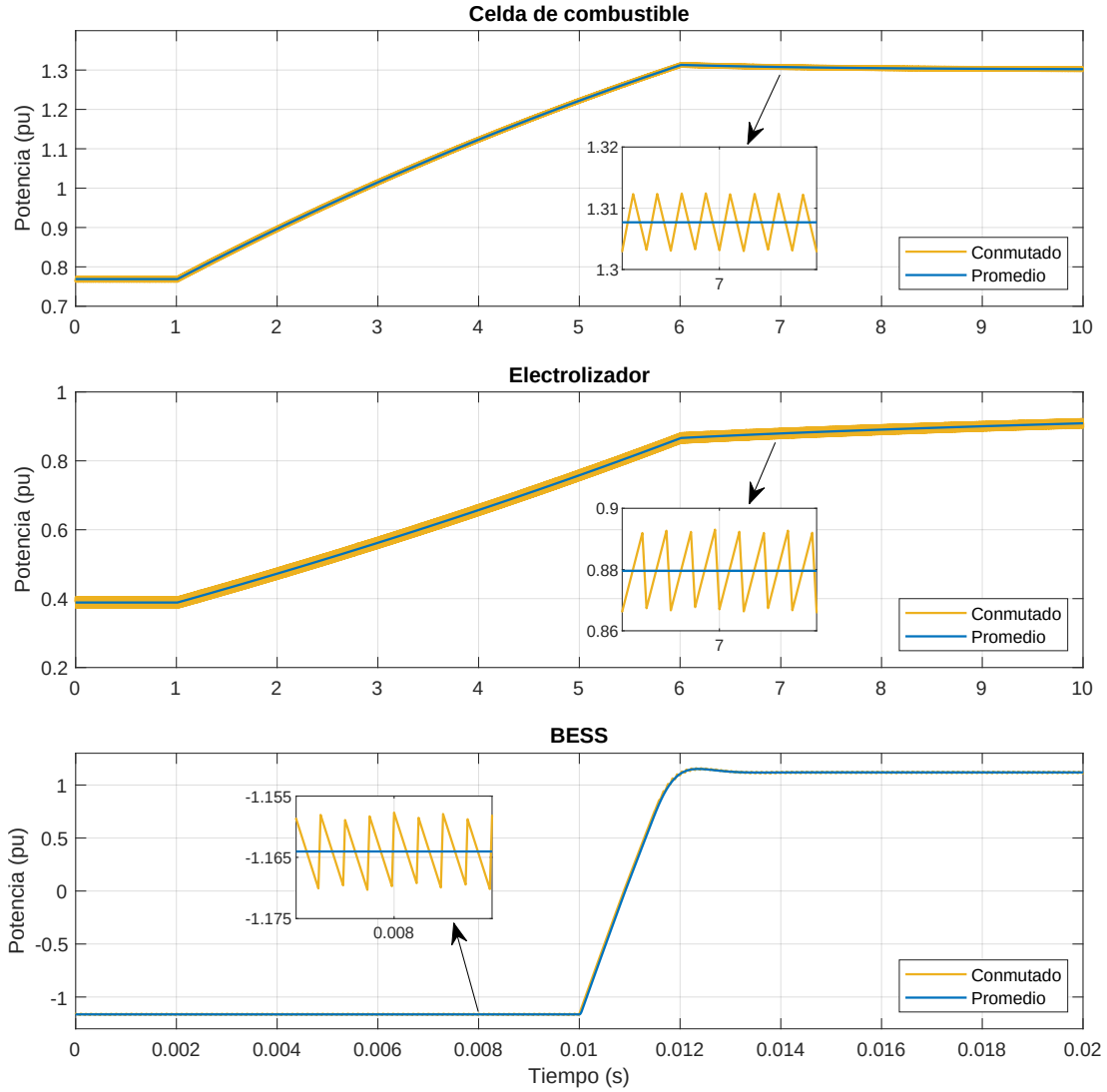

Figura 3.44: Validación de modelos promedio.

Esta página se ha dejado intencionadamente en blanco.

Capítulo 4

Aprendizaje por refuerzo

En este capítulo se presenta la implementación del algoritmo DDPG, el que tiene por objetivo llevar a cabo el entrenamiento del agente responsable de gestionar los intercambios de energía en la planta de almacenamiento. En primer lugar, se describe la política de control referida, para posteriormente abordar el diseño de la función de recompensa utilizada durante el entrenamiento del agente. Finalmente, es presentada la configuración del algoritmo que permitió alcanzar un aprendizaje satisfactorio.

4.1. Política de control

El funcionamiento adecuado de la planta de almacenamiento depende de que el balance de potencia se mantenga en todo momento. Para satisfacer esta condición se pueden establecer múltiples consignas; sin embargo, según los objetivos que se busquen alcanzar, algunas resultarán más adecuadas que otras. En el presente caso de estudio se definen tres criterios no negociables para el diseño de la política, los cuales se enuncian a continuación.

- El estado de carga del sistema de almacenamiento basado en baterías no puede descender por debajo del 20 % y no puede estar por sobre el 80 %, con la finalidad de preservar la vida útil de las celdas [45].
- Siempre debe existir capacidad disponible en el sistema de almacenamiento basado en baterías, tanto para carga como para descarga, pues este sistema es responsable de compensar los desbalances transitorios provocados por los requerimientos de la red.
- La planta de almacenamiento tiene como objetivo proporcionar control primario de frecuencia; en este contexto, los servicios auxiliares son suministrados por la celda de combustible, con el fin de ofrecer flexibilidad total en la absorción o entrega de potencia a la red.

Una de las políticas que cumple con los criterios previamente definidos consiste en que el sistema de almacenamiento basado en baterías se encargue únicamente de compensar los desbalances transitorios, mientras que los demás sistemas asumen la carga estacionaria según el régimen operativo. Cuando la red requiere potencia, la celda de combustible genera la energía necesaria para satisfacer tanto los requerimientos de la red como los servicios auxiliares, mientras que el electrolizador permanece en régimen de espera. Por el contrario, si la red entrega potencia a la planta de almacenamiento, el electrolizador absorbe la energía proporcionada por la red, mientras que la celda de combustible suministra la potencia requerida para los servicios auxiliares. Aunque esta política es completamente funcional, no tiene como objetivo mejorar la gestión del recurso primario. Cuando la red demanda energía, sólo existe un punto de operación en el cual la celda de combustible puede alcanzar su máxima eficiencia. En cambio, si la red entrega energía a la planta de almacenamiento, la celda de combustible nunca puede operar en su punto óptimo, ya que los servicios auxiliares consumen como máximo aproximadamente 0.23 pu.

El problema central radica en que siempre existe una demanda fija de servicios auxiliares que obliga el despacho de la celda de combustible, forzando el consumo no óptimo del recurso primario. Por simplicidad, en el presente caso de estudio se considerará la condición de operación más desfavorable, en la que tanto la celda de combustible como el electrolizador permanecen encendidos en todo momento, ya sea en operación activa

o en régimen de espera. No se analizarán escenarios en los que la celda de combustible o el electrolizador se apaguen por completo, ya que la decisión de una inactivación total depende del despacho definido por el operador de red, considerando que los tiempos de arranque en frío superan la ventana temporal del control primario de frecuencia [46].

La política propuesta se basa en la siguiente filosofía. El objetivo de incorporar una planta de almacenamiento de energía en un sistema eléctrico con alta penetración de energías renovables es, por un lado, absorber los excedentes de energía de la red, generalmente provenientes de la generación solar durante el día, y, por otro lado, liberarlos cuando la generación solar no está disponible, es decir, durante la noche. De esta manera, la planta de almacenamiento tendrá indiscutiblemente horarios de carga y descarga. Durante el régimen de carga, el electrolizador absorberá toda la potencia proveniente de la red, mientras que la celda de combustible operará en su punto óptimo. Estas consignas generan un excedente de potencia en el enlace de corriente directa, pues tal como se mencionó anteriormente, el punto óptimo de operación de la celda de combustible siempre es superior al máximo consumo de servicios auxiliares. Dicho excedente de potencia será absorbido por el sistema de almacenamiento basado en baterías, siempre que el estado de carga no supere una cota superior bien definida. Si el estado de carga supera la cota superior, la celda de combustible deberá suministrar únicamente la potencia necesaria para cubrir los servicios auxiliares. Por otra parte, durante el régimen de descarga la celda de combustible se despachará en su punto óptimo, mientras que el electrolizador permanecerá en espera. Si el punto óptimo es inferior al requerimiento agregado de la red y los servicios auxiliares, el sistema de almacenamiento en baterías proveerá la potencia faltante siempre que el estado de carga se encuentre por sobre una cota inferior bien definida. Si el punto óptimo es superior al requerimiento agregado de la red y los servicios auxiliares, el excedente será absorbido por el banco de baterías siempre que el estado de carga se encuentre por debajo de la cota superior anteriormente referida. En caso de que no se satisfagan las condiciones del estado de carga, la celda de combustible no podrá operar en su punto óptimo y deberá satisfacer la potencia demandada por la red y los servicios auxiliares. A continuación se presenta la política de control propuesta en forma de pseudocódigo para una comprensión más directa.

Pseudocódigo 1 Política de control propuesta

if $[p_1 > 0]$ **then**
  $p_{\mathrm{EL}} = 0$
  if $[\mathrm{SoC} \leq 40\,\%] \vee [(\mathrm{SoC} \geq 60\,\%) \wedge ({p_{\mathrm{FC}}}^* \geq p_1 + p_{\mathrm{Aux}})]$ **then**
    $p_{\mathrm{FC}} = p_1 + p_{\mathrm{Aux}}$
  else
    $p_{\mathrm{FC}} = {p_{\mathrm{FC}}}^*$
  end if
else if $[p_1 < 0]$ **then**
  $p_{\mathrm{EL}} = -p_1$
  if $[\mathrm{SoC} \geq 60\,\%]$ **then**
    $p_{\mathrm{FC}} = p_{\mathrm{Aux}}$
  else
    $p_{\mathrm{FC}} = {p_{\mathrm{FC}}}^*$
  end if
else
  $p_{\mathrm{FC}} = p_{\mathrm{Aux}}$
  $p_{\mathrm{EL}} = 0$
end if

Esta política mejora la gestión del recurso primario, pues siempre que se satisfagan las condiciones del estado de carga, la celda de combustible puede operar en su punto óptimo y los excedentes de energía son almacenados en un dispositivo de eficiencia superior como lo es el banco de baterías. Más aún, este último no se saturará permanentemente en la cota superior de almacenamiento pues durante el régimen de descarga liberará energía. Se ha definido una cota inferior y superior del 40 % y 60 % respectivamente, con la finalidad de mantener una holgura del 20 % en los extremos de la banda de operación destinada exclusivamente a asegurar la compensación de los desbalances transitorios. Es importante advertir que el sistema podrá mejorar aún más la gestión del recurso primario si la banda de compensación es acotada,

pero ello supone comprometer la disponibilidad del banco de baterías para hacerse cargo de los desbalances transitorios. La optimización de la banda de compensación se reserva para estudios futuros, pues depende de los requerimientos del control primario de frecuencia en la barra de conexión.

4.2. Función de recompensa

La función de recompensa utilizada durante el entrenamiento del agente debe estar alineada con la política de control enunciada anteriormente, y más aún, debe ser lo suficientemente robusta para que el agente no encuentre políticas con las que maximice la recompensa sin satisfacer la política de control. Dada la asignación de consignas por parte de la política de control, la función de recompensa deberá contar con una componente dedicada a anular el error entre un valor medido y un valor esperado. Adicionalmente, deberá contar con una componente destinada a penalizar el uso del electrolizador cuando la planta de almacenamiento se encuentra en régimen de descarga.

Sea e el error entre una consigna cualquiera y su valor medido. Con la finalidad de anular el error, pueden formularse dos componentes que persiguen el mismo objetivo, las cuales se presentan en (4.1).

$$f_{\text{Square}}(e) = 1 - k_{\text{Square}}\, e^2 \qquad\qquad f_{\text{Asym}}(e) = \frac{k_{\text{Asym}}}{k_{\text{Asym}} + |e|} \tag{4.1}$$

Ambas formulaciones maximizan la recompensa cuando el error es eliminado; sin embargo, exhiben un comportamiento diferente durante la fase de adquisición de recompensas. Ante errores significativos ambas funciones impulsan al agente a reducir el error para maximizar el retorno. En cambio, frente a errores pequeños, la formulación cuadrática atenúa progresivamente su gradiente disminuyendo los incentivos percibidos por el agente para continuar mejorando la política. Este fenómeno no se observa en la formulación asintótica, por el contrario, el gradiente constantemente insta al agente a obtener mejores recompensas como se muestra en la Figura 4.1. De esta manera, la asignación de consignas se materializará en la función de recompensa mediante componentes asintóticas.

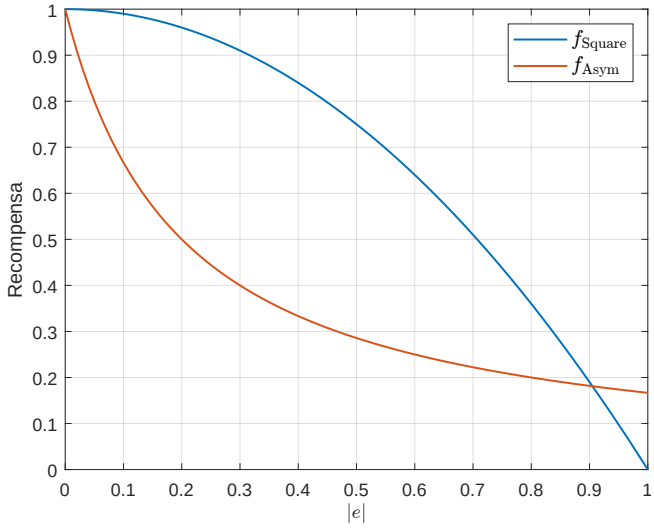

Figura 4.1: Funciones anuladoras de error.

Por otra parte, el uso del electrolizador será penalizado utilizando una componente lineal tal como se presenta en (4.2). Dentro de esta función, se define $p_{\text{EL,máx}}$ como la potencia máxima que puede alcanzar el electrolizador en por unidad.

$$f_{\text{Pen}}(p_{\text{EL}}) = 1 - \frac{1}{p_{\text{EL,máx}}}\, p_{\text{EL}} \tag{4.2}$$

Una vez presentadas las componentes esenciales de la función de recompensa, se procede a definir su formulación en función del régimen de operación y del estado de carga del banco de baterías. En régimen de descarga y con un estado de carga inferior o igual al 40 %, o bien, en régimen de descarga, con un estado de carga superior o igual al 60 % y un requerimiento agregado de la red y de los servicios auxiliares igual o inferior al punto óptimo de la celda de combustible, la función de recompensa queda dada por (4.3).

$$r_1(p_{\mathrm{FC}}, p_1, p_{\mathrm{Aux}}, p_{\mathrm{EL}}) = w_{1,1}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{FC}} - p_1 - p_{\mathrm{Aux}}|}\right] + w_{1,2}\left[1 - \frac{1}{p_{\mathrm{EL,máx}}}\, p_{\mathrm{EL}}\right] \tag{4.3}$$

En régimen de descarga, para cualquier otro caso que no haya sido abordado por r_1, la función de recompensa queda dada por (4.4).

$$r_2(p_{\mathrm{FC}}, p_{\mathrm{EL}}) = w_{2,1}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{FC}} - {p_{\mathrm{FC}}}^*|}\right] + w_{2,2}\left[1 - \frac{1}{p_{\mathrm{EL,máx}}}\, p_{\mathrm{EL}}\right] \tag{4.4}$$

En régimen de carga, con un estado de carga del sistema de almacenamiento basado en baterías superior o igual a 60 %, la función de recompensa queda dada por (4.5).

$$r_3(p_{\mathrm{FC}}, p_1, p_{\mathrm{Aux}}, p_{\mathrm{EL}}) = w_{3,1}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{EL}} + p_1|}\right] + w_{3,2}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{FC}} - p_{\mathrm{Aux}}|}\right] \tag{4.5}$$

En régimen de carga, con un estado de carga del sistema de almacenamiento basado en baterías inferior al 60 %, la función de recompensa queda dada por (4.6).

$$r_4(p_{\mathrm{FC}}, p_1, p_{\mathrm{EL}}) = w_{4,1}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{EL}} + p_1|}\right] + w_{4,2}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{FC}} - {p_{\mathrm{FC}}}^*|}\right] \tag{4.6}$$

En caso de no existir flujo de potencia a través del convertidor formador de red, la función de recompensa queda dada por (4.7).

$$r_5(p_{\mathrm{FC}}, p_{\mathrm{Aux}}, p_{\mathrm{EL}}) = w_{5,1}\left[\frac{k_{\mathrm{Asym}}}{k_{\mathrm{Asym}} + |p_{\mathrm{FC}} - p_{\mathrm{Aux}}|}\right] + w_{5,2}\left[1 - \frac{1}{p_{\mathrm{EL,máx}}}\, p_{\mathrm{EL}}\right] \tag{4.7}$$

El coeficiente asintótico k_{Asym} regula la penalización del error, incrementando la severidad conforme el coeficiente disminuye. Por otra parte, los pesos $w_{i,j}$ son responsables de calibrar la importancia relativa de cada objetivo. Si los pesos son normalizados, la señal de recompensa queda acotada superiormente por la unidad, característica útil para el posterior análisis de los resultados del entrenamiento.

Una función de recompensa con naturaleza condicional degrada el aprendizaje del agente, pues una misma tupla (s, a, s') puede generar recompensas distintas según la rama lógica activa. Esto rompe la propiedad de Markov sobre la que se cimenta el algoritmo, provocando que la función de pérdidas del crítico sufra alta varianza y sesgo. Adicionalmente, los umbrales inducen discontinuidades locales en la estimación del crítico a causa de estar cerca de las fronteras lógicas. El resultado de implementar una función de recompensa condicional sin tomar acciones que mitiguen los efectos descritos, son políticas que convergen lento y/o a soluciones no óptimas, además de una elevada sensibilidad a hiperparámetros y exploración. Para afrontar los problemas derivados de la recompensa condicional, se adoptan las ideas de [47] y se implementa una máquina de recompensas que, además del valor escalar de recompensa r_t, devuelve un vector de activación $\tilde{\mathbf{H}}_t$ que identifica la rama lógica activa. Este vector se inserta en las observaciones del agente, recuperando la propiedad de Markov requerida por el algoritmo. Se define el vector de activación $\tilde{\mathbf{H}}_t$ según

$$\tilde{\mathbf{H}}_t = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 \end{bmatrix}^{\mathrm{T}} \tag{4.8}$$

donde $h_i \in \{0, 1\}$, ocurriendo la activación del estado i cuando se activa la i-ésima función de recompensa definida previamente.

4.3. Implementación del algoritmo

Como se mencionó anteriormente, el entrenamiento del agente se realiza sobre el modelo promedio de la planta de almacenamiento. Tanto el modelo referido como el algoritmo de entrenamiento fueron implementados en MATLAB, pues la extensión Reinforcement Learning Toolbox permite la integración de agentes RL en modelos Simulink, entorno de simulación en el que fue desarrollado el modelo de la planta de almacenamiento. Para la ejecución de las simulaciones se empleó un tiempo de muestreo de 50 μs, mientras que el tiempo de muestreo del agente se fijó en 100 ms. No se justifica un tiempo de muestreo más pequeño para el agente, ya que la asignación de consignas de potencia no requiere capturar dinámicas rápidas.

El vector de observaciones corresponde al conjunto de variables que representan el estado actual del entorno, y que son utilizadas por el agente para tomar decisiones. La selección cuidadosa de estas observaciones es fundamental, ya que determina la capacidad del agente para percibir correctamente el estado del sistema, identificar situaciones relevantes y generar acciones efectivas. Además, se recomienda normalizar las observaciones, típicamente a un rango acotado como $[0, 1]$ o $[-1, 1]$, lo que contribuye a estabilizar el entrenamiento, evitar saturación de las funciones de activación en las redes neuronales y acelerar la convergencia del agente. En teoría, el mínimo de observaciones que deben ser provistas al agente son aquellas que permiten satisfacer la propiedad de Markov sobre la que se cimenta el algoritmo, pero tal como se menciona en [20], la elección final no sigue una regla única y suele requerir juicio y experiencia.

La construcción del vector de observaciones se realiza siguiendo la filosofía que se presenta a continuación. El balance de potencia en el enlace de corriente directa está definido por (4.9), considerando que el agente únicamente puede actuar sobre la potencia inyectada por la celda de combustible y la potencia absorbida por el electrolizador. Los servicios auxiliares no pueden modificarse, ya que dependen directamente de la operación de los sistemas basados en hidrógeno, mientras que la potencia gestionada por el banco de baterías se encuentra regulada por el controlador de tensión del enlace de corriente directa.

$$\Delta p = p_{\mathrm{FC}} + p_{\mathrm{B}} - p_{\mathrm{EL}} - p_{\mathrm{Aux}} - p_1 \tag{4.9}$$

Es importante destacar que, aunque el agente no tenga la capacidad de actuar directamente sobre la potencia gestionada por el banco de baterías, sí debe disponer de información sobre la compensación realizada por el controlador de tensión. Esta información se captura mediante el remanente hiperbólico $\Delta\rho$, definido en (4.10). Cuando el agente sigue la política de control, el remanente de potencia que debe compensar el banco de baterías es pequeño, o al menos acotado, lo que garantiza que la observación opera principalmente en su zona lineal. En cambio, si el agente no sigue la política de control, la potencia que debería gestionar el banco de baterías aumenta significativamente, provocando la saturación de la observación. En síntesis, esta observación permite al agente asociar su saturación con recompensas bajas, mientras que, cuando no se encuentra saturada, proporciona información relevante sobre el balance de potencia. Dada la naturaleza de la tangente hiperbólica, no es necesario normalizar esta observación.

$$\Delta\rho = \tanh(p_{\mathrm{FC}} - p_{\mathrm{EL}} - p_{\mathrm{Aux}} - p_1) \tag{4.10}$$

Al vector de observaciones se incorporan también la potencia aportada por la celda de combustible p_{FC}, la potencia absorbida por el electrolizador p_{EL}, la potencia requerida para la operación de los servicios auxiliares p_{Aux} y la potencia retirada o suministrada por el convertidor formador de red p_1. La inclusión de estas observaciones permite ampliar la percepción del entorno por parte del agente y garantiza que se cumpla la propiedad de Markov en el balance de potencia. Asimismo, la observación de la potencia del convertidor formador de red proporciona información directa sobre el régimen de operación que se está llevando a cabo. De este conjunto de observaciones sólo se normaliza la potencia aportada por la celda de combustible y la potencia suministrada por el electrolizador, dividiendo ambas señales por sus valores nominales en por unidad. El resto de observaciones no son normalizadas pues naturalmente se encuentran dentro del rango recomendado.

Se define que las acciones del agente corresponden a las referencias de los controladores de corriente de la celda de combustible y del electrolizador, pues a través de ellas el agente varía las consignas de potencia. La potencia entregada por cada sistema depende del producto entre la tensión aplicada por el controlador y la corriente circulante; por ello, incluir la tensión referida previamente en las observaciones proporciona al

agente una medida del esfuerzo adicional que debe ejercer a través de la corriente de referencia para alcanzar la potencia deseada. Dado que la máxima tensión que los convertidores pueden imponer en sus terminales está limitada por la tensión del enlace de corriente directa, estas observaciones se normalizan con respecto a la tensión nominal del enlace.

La última observación física incluida en el vector de observaciones es el estado de carga del banco de baterías, el cual proporciona al agente información sobre el régimen de operación cuando se combina con la potencia retirada o suministrada por el convertidor formador de red. Además, esta observación indica cuánto margen de estado de carga se tiene antes de alcanzar alguno de los umbrales lógicos definidos previamente. El estado de carga se presenta al agente en su forma decimal, por lo que no requiere normalización. De este modo, los vectores de observaciones y acciones del agente quedan definidos según lo indicado por (4.11) y (4.12) respectivamente. Las acciones del agente se establecen en el rango $[0, 1]$, para luego ser escaladas mediante los valores nominales de corriente en por unidad de cada sistema.

$$\mathbf{O_t} = \begin{bmatrix} \Delta\rho & p_{\mathrm{FC}} & p_{\mathrm{EL}} & p_{\mathrm{Aux}} & p_1 & v_{\mathrm{Boost}} & v_{\mathrm{Buck}} & \mathrm{SoC} & h_1 & h_2 & h_3 & h_4 & h_5 \end{bmatrix}^{\mathrm{T}} \tag{4.11}$$

$$\mathbf{A_t} = \begin{bmatrix} {i_{\mathrm{FC}}}^* & {i_{\mathrm{EL}}}^* \end{bmatrix}^{\mathrm{T}} \tag{4.12}$$

La configuración de las redes neuronales constituye un elemento fundamental en el desempeño del agente, pues ellas determinan la capacidad del modelo para aproximar la función de valor y la política en entornos de alta dimensionalidad. Tanto la red del actor, encargada de generar las acciones, como la del crítico, responsable de evaluar su calidad, requieren arquitecturas adecuadamente seleccionadas en términos de profundidad, número de neuronas y funciones de activación, a fin de garantizar una representación estable y eficiente del espacio de estados y acciones. Una configuración inadecuada puede conducir a una convergencia lenta o incluso inestable, mientras que un diseño apropiado permite capturar con mayor precisión las relaciones no lineales del sistema, favoreciendo así el aprendizaje óptimo y la generalización del agente frente a distintas condiciones operativas.

Para la red neuronal del actor se definió la arquitectura descrita a continuación. La capa de entrada se compone de trece neuronas que reciben las observaciones del agente. Posteriormente se incorporan dos capas ocultas completamente conectadas, cada una con 64 neuronas y función de activación `ReLU`. La capa de salida consta de dos neuronas con función de activación `Sigmoid`, lo que permite restringir las acciones al rango operativo deseado. Además, se implementó normalización por lotes entre capas (BN, por sus siglas en inglés) con el objetivo de reducir el desplazamiento covariante interno y mejorar la estabilidad del aprendizaje [25].

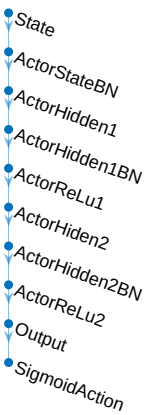

Figura 4.2: Red del actor.

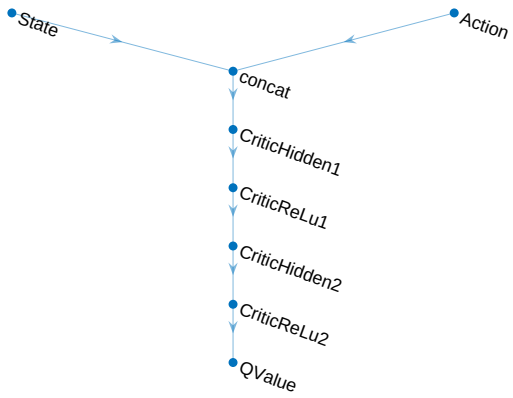

Figura 4.3: Red del crítico.

En cuanto a la red neuronal del crítico, su capa de entrada está formada por la concatenación de los vectores de observaciones y acciones del agente. La red incluye dos capas ocultas completamente conectadas, cada una con 128 neuronas y función de activación `ReLU`. La capa de salida consta de una única neurona sin función de activación, dado que su propósito es aproximar el valor de $Q(s, a)$. Aplicar una función de activación en la salida de la red crítica podría inducir errores en la estimación de la función de valor al limitar su rango efectivo.

Se define que el optimizador para el ajuste de los pesos y sesgos de las redes neuronales será `Adam`, tal como se implementó en [25]. En el entrenamiento de agentes basados en métodos actor - crítico, la selección de la tasa de aprendizaje representa un compromiso entre velocidad de convergencia y estabilidad del entrenamiento. Una tasa de aprendizaje alta permite descender rápidamente por la superficie de error durante las primeras etapas, acelerando la adquisición de conocimiento, pero al mismo tiempo puede provocar oscilaciones en los gradientes e impedir la convergencia hacia una política estable. Por el contrario, una tasa de aprendizaje baja mejora la estabilidad numérica y reduce la varianza de las actualizaciones, aunque a costa de una convergencia más lenta y una posible detención prematura en mínimos locales. Dentro de este equilibrio, se recomienda que la tasa de aprendizaje del crítico sea superior a la del actor, ya que el primero debe adaptarse con mayor rapidez a los cambios de política para proporcionar evaluaciones actualizadas del valor $Q(s, a)$. De este modo, el actor recibe señales de gradiente más coherentes y puede ajustar su política de manera gradual, evitando que errores momentáneos en la estimación del crítico generen inestabilidades en la exploración [48]. Sin embargo, en el presente trabajo esta configuración no produjo un desempeño adecuado, por lo que ambas tasas de aprendizaje se igualaron, obteniéndose un comportamiento más estable durante el proceso de entrenamiento. En la Tabla 4.1 se presentan los parámetros ajustados para el aprendizaje de las redes neuronales del algoritmo.

Tabla 4.1: Parámetros de aprendizaje.

Parámetro	**Actor**	**Crítico**
Tasa de aprendizaje	1×10^{-4}	1×10^{-4}
Saturación de gradiente	1.00	1.00
Metodo de saturación	`l2norm`	`l2norm`
Factor de regularización L2	1×10^{-3}	1×10^{-4}

En cualquier algoritmo basado en aprendizaje por refuerzo, la función de valor $Q(s, a)$ se define como la esperanza matemática del retorno descontado evaluado en el estado s_t y la acción a_t.

$$Q(s_t, a_t) = \mathbb{E}\left[G_t \mid s_t, a_t\right] \tag{4.13}$$

Luego, el máximo valor que puede tomar la función $Q(s, a)$ queda dada por la adquisición máxima de recompensas. Considerando que el entorno sobre el que se entrena el agente es determinista, y que en la práctica el horizonte de interacciones es infinito, la máxima adquisición de recompensas se obtiene cuando

$$r_{t+k+1} = r_{\text{máx}}, \quad \forall\, k \tag{4.14}$$

Sustituyendo en el retorno descontado, es posible aproximar

$$Q_{\text{máx}} \approx \sum_{k=0}^{\infty} \gamma^k \, r_{\text{máx}} = \frac{r_{\text{máx}}}{1-\gamma} \tag{4.15}$$

El objetivo del agente es ajustar su política para maximizar las recompensas adquiridas, por lo que es esperable que la función de valor alcance de forma estable un valor cercano a $Q_{\text{máx}}$ una vez que el entrenamiento haya concluido. Es por esta razón que es útil acotar las recompensas entre 0 y 1, pues de esta manera es posible evaluar el desempeño del agente en una escala normalizada y comparable entre distintos entornos o configuraciones. Además, una recompensa acotada garantiza la estabilidad numérica del cálculo del retorno descontado y del valor $Q(s, a)$, evitando que este crezca sin límite a medida que se aproxima γ a uno. En consecuencia, la definición de una recompensa limitada contribuye tanto a la interpretabilidad del proceso de aprendizaje como a la robustez del entrenamiento del agente.

El ajuste de los hiperparámetros del algoritmo constituye un proceso iterativo, dado que cada problema de control presenta características propias que determinan una sensibilidad distinta frente a cada grado de libertad. En consecuencia, la selección óptima de estos valores no puede generalizarse, sino que requiere de un proceso de calibración adaptado al comportamiento del entorno y a los objetivos de aprendizaje. A continuación, se presenta una guía resumida que describe la influencia de los principales hiperparámetros sobre el desempeño del agente, sirviendo como referencia para el ajuste y la configuración final del algoritmo.

- **Tasa de descuento** γ: determina la relevancia de las recompensas futuras frente a las inmediatas. Valores altos promueven estrategias con visión de largo plazo, mientras que valores bajos inducen comportamientos más miopes y reactivos.
- **Desviación estándar del ruido de exploración** σ: controla la amplitud de las acciones aleatorias aplicadas durante la exploración. Un valor alto favorece la búsqueda de nuevas trayectorias, aunque puede dificultar la convergencia; en cambio, valores bajos reducen la exploración y aceleran la estabilización de la política.
- **Tasa de decaimiento del ruido de exploración** Θ: regula la rapidez con que el ruido retorna a su valor medio, afectando la persistencia temporal de las perturbaciones. Una tasa elevada produce un ruido de corta duración y más errático, mientras que una baja genera correlaciones temporales más largas y exploraciones más suaves.
- **Tamaño de la memoria de experiencias**: define la cantidad de transiciones almacenadas para el entrenamiento del agente. Memorias grandes aumentan la diversidad de experiencias y la estabilidad del aprendizaje, aunque incrementan el costo computacional.
- **Tamaño del mini-lote aleatorio**: especifica cuántas muestras se utilizan en cada actualización de los pesos y sesgos de las redes neuronales. Mini - lotes grandes reducen la varianza de los gradientes y estabilizan el aprendizaje, pero disminuyen la frecuencia de actualización; mini-lotes pequeños aceleran las iteraciones, aunque con mayor ruido en las actualizaciones.
- **Tasa de actualización de las redes objetivo** τ: controla el grado de suavizado entre las redes principales y las redes objetivo. Valores pequeños generan actualizaciones lentas pero estables, mientras que valores altos provocan una respuesta más rápida con riesgo de inestabilidad numérica.

En la Tabla 4.2 se presenta la configuración de hiperparámetros con la cual se logró un entrenamiento satisfactorio del agente. Cabe destacar que alcanzar dicha configuración requirió un extenso proceso de prueba y ajuste. Adicionalmente, es importante advertir que para la tasa de descuento definida se estima un máximo teórico de 33.33 en la función de valor.

Tabla 4.2: Hiperparámetros definidos.

Parámetro	**Valor**
Tasa de descuento γ	0.97
Desviación estándar σ	0.20
Tasa de decaimiento Θ	$1{,}6 \times 10^{-5}$
Memoria de experiencia	1×10^{6}
Mini - lote aleatorio	128
Tasa de actualización τ	1×10^{-3}

En cuanto a la modalidad de entrenamiento, se establecieron tandas de 10 000 episodios, cada uno con una duración de 60 segundos de simulación, durante los cuales el agente interactúa con el sistema bajo distintos escenarios operativos. En dichos escenarios se modifican las condiciones iniciales de operación y los requerimientos de potencia de la red, con el propósito de ampliar la diversidad de experiencias y favorecer la capacidad de generalización del agente. A lo largo del entrenamiento se monitorean las recompensas totales obtenidas en cada episodio, la media móvil de las recompensas acumuladas en los últimos 50 episodios, la estimación inicial de la red crítica y las pérdidas asociadas tanto a la red del actor como a la del crítico.

Como criterio de detención del entrenamiento se considera el aplanamiento de la estimación inicial de la red crítica, dado que en esta condición el gradiente suministrado al optimizador de la red del actor tiende a cero. Este comportamiento sugiere que el crítico ha alcanzado una evaluación estable de la política actual y, en consecuencia, estima que no es necesario continuar ajustándola. No obstante, este criterio no garantiza que el entrenamiento haya sido satisfactorio, sino únicamente que la política ha dejado de evolucionar.

4.4. Evaluación del desempeño

El criterio de detención del entrenamiento se establece en función de la valoración que realiza el crítico sobre la calidad de la política aprendida. No obstante, dado que esta evaluación representa únicamente la percepción interna del algoritmo, resulta necesario comprobar si la política obtenida mantiene un desempeño consistente bajo condiciones diferentes de las experimentadas durante el entrenamiento. Con este propósito, se lleva a cabo un proceso de validación en seis escenarios operativos independientes, orientados a examinar las distintas capacidades del agente frente a la política condicional previamente definida. Dichos escenarios permiten verificar que el agente cumpla correctamente con los regímenes de carga y descarga establecidos, así como con los límites impuestos sobre el estado de carga, garantizando que la política aprendida no solo exhiba estabilidad numérica, sino también coherencia con las restricciones físicas y operativas del sistema.

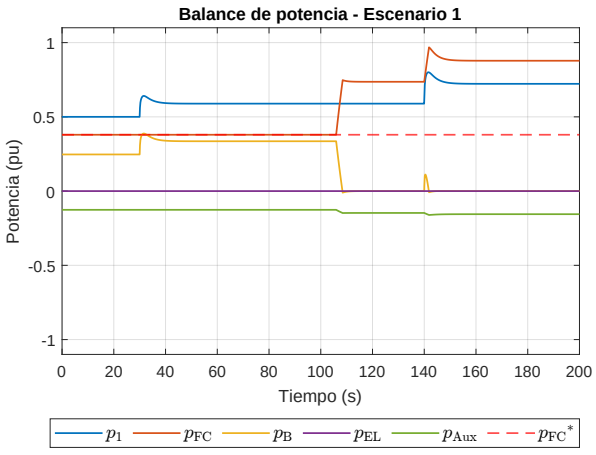

Figura 4.4: Balance de potencia - Escenario 1.

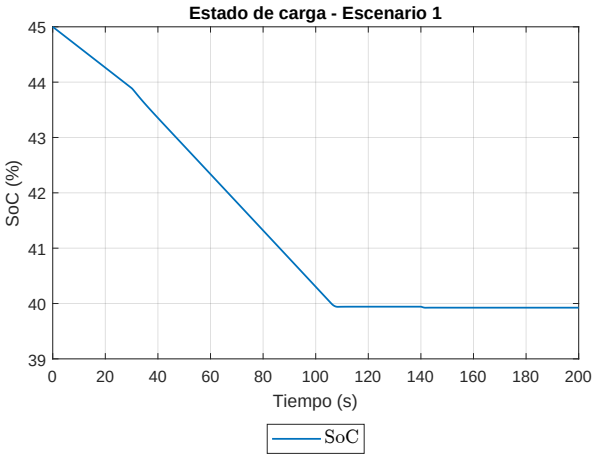

Figura 4.5: Estado de carga - Escenario 1.

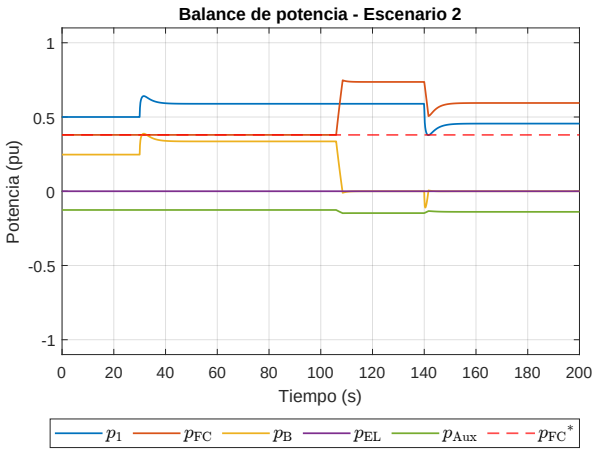

Figura 4.6: Balance de potencia - Escenario 2.

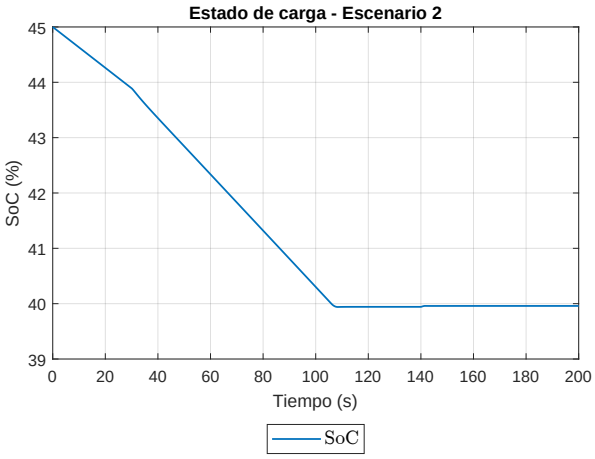

Figura 4.7: Estado de carga - Escenario 2.

En las Figuras 4.4 y 4.5 se presenta respectivamente el balance de potencia y el estado de carga asociado al primer escenario de evaluación. Las condiciones iniciales del sistema son una demanda de 0.5 pu de potencia por parte del convertidor formador de red, y un estado de carga en el sistema de baterías de 45 %. La política de control indica que bajo esta condición inicial la celda de combustible debe despacharse en su punto óptimo, mientras que el electrolizador debe permanecer en espera sin consumir potencia. En el instante 30 s ocurre una perturbación en la red que provoca un incremento en la demanda de potencia por parte del convertidor. Posteriormente, entre los instantes 100 s y 120 s se alcanza el límite inferior del estado de carga del banco de baterías, implicando que la celda de combustible debe abandonar su punto óptimo para inyectar la potencia requerida por la red y los servicios auxiliares. Finalmente, en el instante 140 s nuevamente ocurre una perturbación que provoca un incremento en la demanda de potencia por parte del convertidor, cuyo desbalance lo asume de forma íntegra la celda de combustible. El primer escenario de evaluación es prácticamente idéntico al segundo, presentado en las Figuras 4.6 y 4.7, con la única diferencia de que en el instante 140 s en vez de verse incrementada la potencia demandada por el convertidor, esta es disminuida.

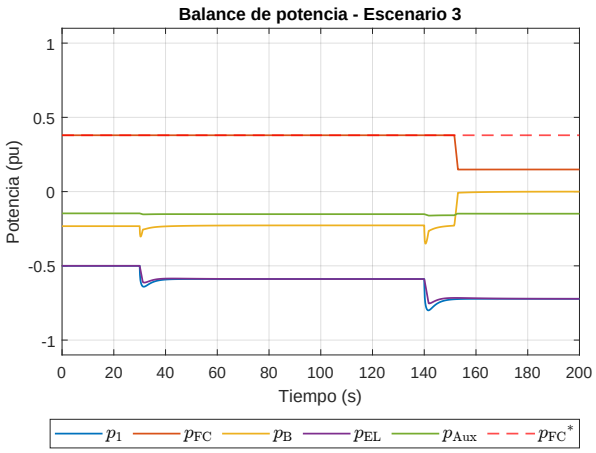

Figura 4.8: Balance de potencia - Escenario 3.

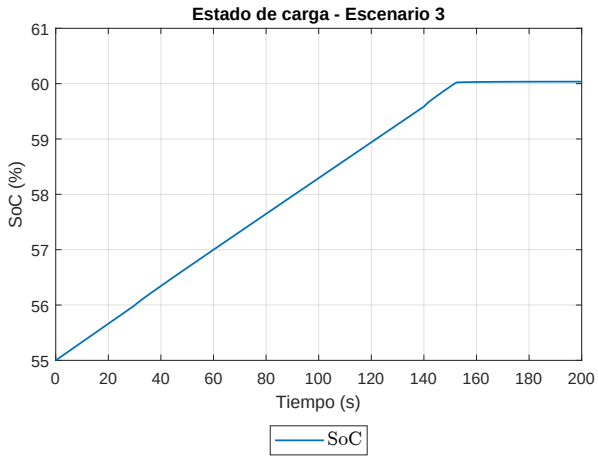

Figura 4.9: Estado de carga - Escenario 3.

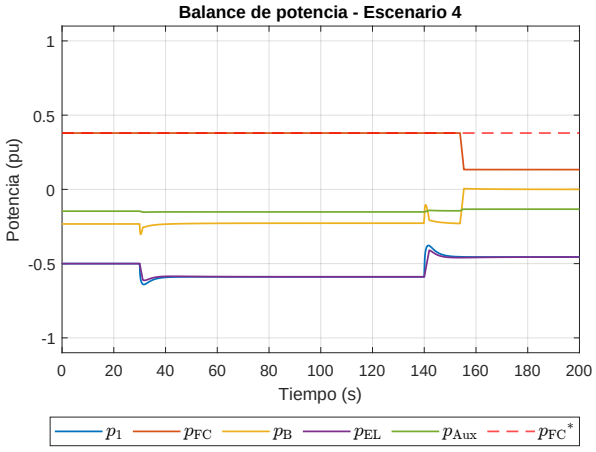

Figura 4.10: Balance de potencia - Escenario 4.

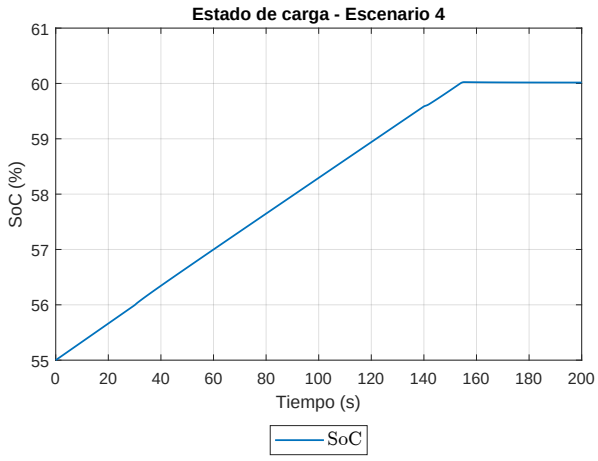

Figura 4.11: Estado de carga - Escenario 4.

En las Figuras 4.8 y 4.9 se presenta respectivamente el balance de potencia y el estado de carga asociado al tercer escenario de evaluación. Las condiciones iniciales del sistema son una entrega de 0.5 pu de potencia por parte del convertidor formador de red, y un estado de carga en el sistema de baterías de 55 %. La política de control indica que bajo esta condición inicial la celda de combustible debe despacharse en su punto óptimo, mientras que el electrolizador debe asumir la totalidad de la potencia entregada por la red. En el

instante 30 s ocurre una perturbación que provoca un incremento en la potencia aportada por el convertidor al enlace de corriente directa. Posteriormente, en el instante 140 s nuevamente ocurre una perturbación en la red que incrementa aún más el aporte de potencia por parte del convertidor; en ambos casos el desbalance es compensado por el electrolizador. Finalmente, entre los instantes 140 s y 160 s se alcanza el límite superior del estado de carga del banco de baterías, por lo que la celda de combustible abandona su punto óptimo para asumir la potencia requerida por los servicios auxiliares. El tercer escenario de evaluación es prácticamente idéntico al cuarto, presentado en las Figuras 4.10 y 4.11, con la única diferencia de que en el instante 140 s en vez de verse incrementada la potencia aportada por la red, esta se ve disminuida.

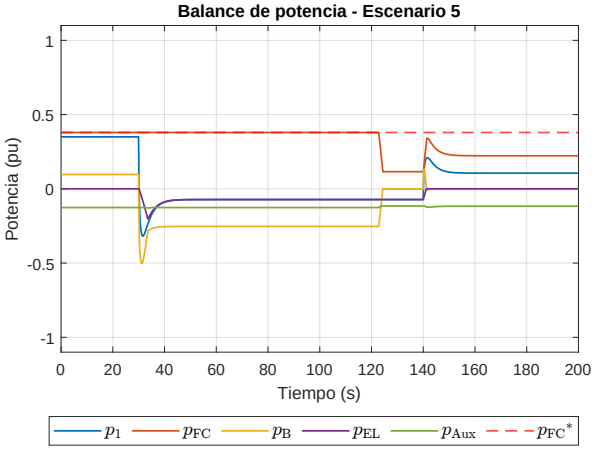

Figura 4.12: Balance de potencia - Escenario 5.

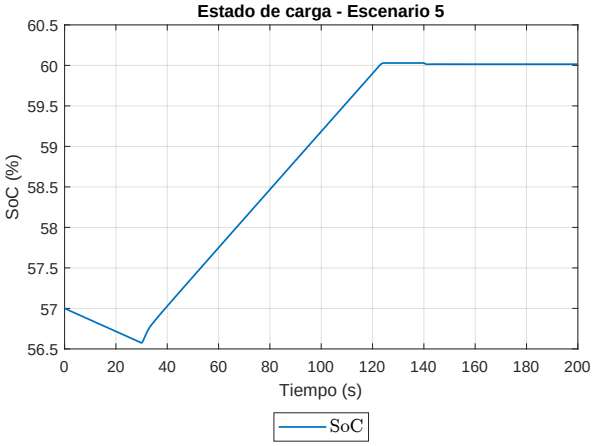

Figura 4.13: Estado de carga - Escenario 5.

En las Figuras 4.12 y 4.13 se presenta respectivamente el balance de potencia y el estado de carga asociado al quinto escenario de evaluación. Las condiciones iniciales del sistema son una demanda de 0.3 pu de potencia por parte del convertidor formador de red, y un estado de carga en el sistema de baterías de 55 %. La política de control indica que bajo esta condición inicial la celda de combustible debe despacharse en su punto óptimo, y que el electrolizador permanece en espera sin consumir potencia. En el instante 30 s ocurre una perturbación en la red tal que invierte el sentido de potencia del convertidor, por lo que el electrolizador asume toda la potencia provista por la red y la celda de combustible permanece en su punto óptimo, hasta que entre los instantes 120 s y 140 s se alcanza el límite superior del estado de carga del banco de baterías. Finalmente, en el instante 140 s se produce una perturbación en la red que invierte nuevamente el sentido del flujo de potencia del convertidor. Aunque en esta condición el estado de carga supera el 40 %, la celda de combustible no puede operar en su punto óptimo, ya que la potencia total demandada por la red y los servicios auxiliares resulta inferior a dicho punto de operación.

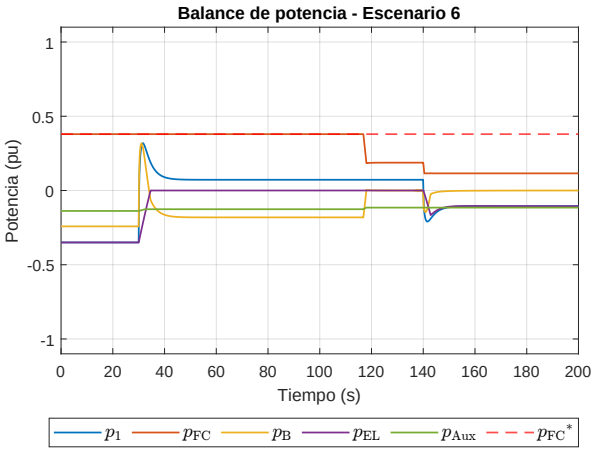

Figura 4.14: Balance de potencia - Escenario 6.

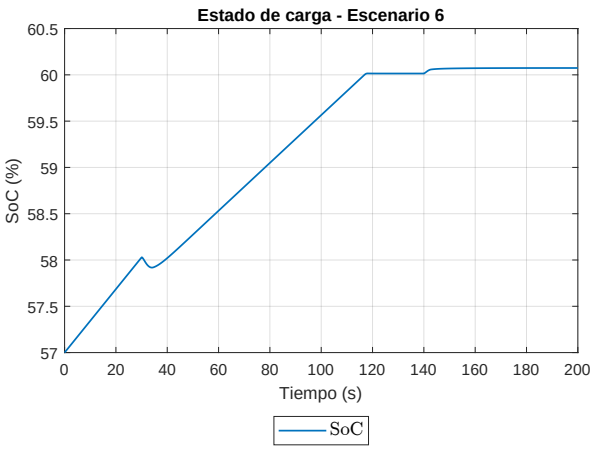

Figura 4.15: Estado de carga - Escenario 6.

En las Figuras 4.14 y 4.15 se presenta respectivamente el balance de potencia y el estado de carga asociado al sexto y último escenario de evaluación. Las condiciones iniciales del sistema son una entrega de 0.3 pu de potencia por parte del convertidor formador de red, y un estado de carga en el sistema de baterías de 57 %. La política de control indica que bajo esta condición inicial la celda de combustible debe despacharse en su punto óptimo, y que el electrolizador debe asumir la totalidad de la potencia entregada por la red. En el instante 30 s ocurre una perturbación tal que invierte el sentido de potencia del convertidor, por lo que el electrolizador anula su consumo y la celda de combustible permanece en su punto óptimo, hasta que entre los instantes 100 s y 120 s se alcanza el límite superior del estado de carga del banco de baterías. Esta última condición fuerza el abandono del punto óptimo por parte de la celda de combustible, pues la potencia total demandada por la red y los servicios auxiliares resulta inferior a dicho punto de operación. Finalmente, en el instante 140 s se produce una perturbación en la red que invierte nuevamente el sentido del flujo de potencia del convertidor, implicando que el electrolizador debe consumir la potencia entregada por la red mientras que la celda de combustible satisface la potencia requerida por los servicios auxiliares.

4.5. Generación de episodios

La generación de episodios constituye un elemento fundamental del proceso de aprendizaje, ya que determina la diversidad y calidad de las experiencias a partir de las cuales el agente actualiza sus redes neuronales. Cada episodio representa una secuencia completa de interacción entre el agente y el entorno, donde las acciones ejecutadas, las observaciones resultantes y las recompensas obtenidas conforman las transiciones almacenadas en el conjunto de experiencias. Una adecuada generación de episodios permite explorar de manera equilibrada el espacio de estados y acciones, garantizando que el agente disponga de información suficiente para aprender tanto comportamientos óptimos como estrategias de corrección frente a condiciones no vistas. En consecuencia, la cobertura y representatividad de los episodios influyen directamente en la estabilidad, velocidad de convergencia y capacidad de generalización del agente.

Cada episodio puede caracterizarse mediante cuatro variables, descritas a continuación. Al inicio de cada episodio2, la planta de almacenamiento se encuentra entregando o absorbiendo una cantidad específica de potencia desde la red, mientras que el estado de carga inicial también se encuentra definido. A partir de estas dos condiciones iniciales se determinan las restantes variables del sistema conforme a la política de control establecida. En otras palabras, la primera observación del agente sigue la política referida. Luego, en $t \in [1, 5]$ se introduce una perturbación en la red tal que incrementa o disminuye la potencia entregada o absorbida por la planta de almacenamiento. Las características descritas pueden ser codificadas en el vector $\mathbf{E}^{(k)}$, el cual describe las condiciones iniciales y la perturbación introducida en el episodio k.

$$\mathbf{E}^{(k)} = \begin{bmatrix} p_{10} & \mathrm{SoC}_0 & \Delta p_0 & t_0 \end{bmatrix}^{\mathrm{T}} \tag{4.16}$$

Para generar los episodios de entrenamiento se evaluó el muestreo por distribución uniforme e hipercubo latino (LHS, por sus siglas en inglés). Independientemente del método empleado, ambos entregan un vector con cuatro elementos que es asociado al episodio k, tal como se muestra a continuación donde $x_i \in [0, 1]$.

$$\mathbf{T}^{(k)} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}^{\mathrm{T}} \tag{4.17}$$

El muestreo mediante LHS resulta preferible frente a una distribución uniforme en la generación de episodios para entornos de aprendizaje por refuerzo, debido a su capacidad para proporcionar una cobertura más homogénea del espacio de muestreo. Mientras que una distribución uniforme puede generar concentraciones aleatorias de puntos en ciertas regiones y dejar otras sin explorar, LHS divide el espacio de cada variable en intervalos equiprobables y selecciona un punto de cada uno, garantizando así que todas las combinaciones posibles de valores sean representadas de manera equilibrada [49]. Esto reduce la varianza del muestreo y mejora la eficiencia del entrenamiento, ya que el agente se expone a una mayor diversidad de condiciones iniciales con un número reducido de episodios, favoreciendo la generalización del comportamiento aprendido.

Una métrica estadística adecuada para comparar el desempeño entre ambos métodos de muestreo es la distancia euclidiana mínima entre los vectores generados $\mathbf{T}^{(k)}$. Un valor mayor de esta distancia indica que los puntos se encuentran más separados entre sí, lo que a su vez refleja una distribución más uniforme y homogénea en el espacio de muestreo. En consecuencia, esta métrica permite evaluar de manera cualitativa la calidad de la dispersión lograda por cada método. En las Figuras 4.16 y 4.17 se muestra la distancia euclidiana mínima obtenida para distintas generaciones de vectores muestreados mediante distribución uniforme e hipercubo latino. La línea punteada representa el valor promedio de dicha distancia considerando todas las generaciones. En la Figura 4.16 se ilustran conjuntos compuestos por 10 episodios por generación, mientras que en la Figura 4.17 se presentan los resultados correspondientes a generaciones de 5000 episodios.

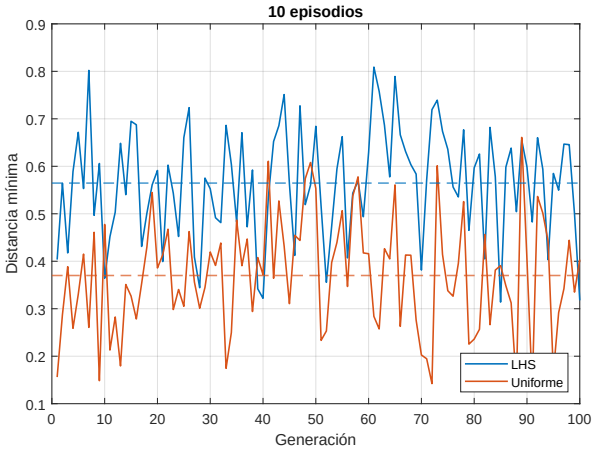

Figura 4.16: Generaciones de 10 episodios.

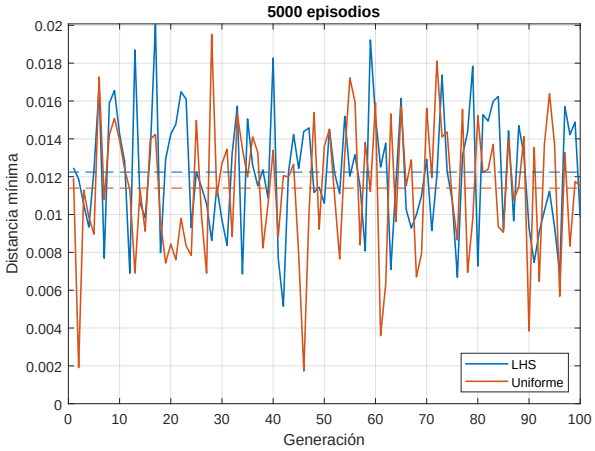

Figura 4.17: Generaciones de 5000 episodios.

En ambos casos se observa que la distancia euclidiana mínima promedio es mayor para el muestreo mediante hipercubo latino que para aquel basado en distribución uniforme, lo que confirma que el LHS genera una distribución de puntos más homogénea y mejor dispersa dentro del espacio de muestreo. Para las generaciones de 10 episodios, el muestreo mediante LHS alcanzó una distancia mínima promedio un 52.6 % superior a la obtenida con distribución uniforme, mientras que para las generaciones de 5000 episodios esta ventaja se redujo a un 7.44 %. Esta diferencia se debe a que, en conjuntos de menor tamaño, la cobertura del espacio de muestreo depende en mayor medida de la estrategia de generación utilizada, por lo que el diseño estructurado del LHS permite evitar la concentración aleatoria de puntos característica del muestreo uniforme.

De este modo, se establece que los episodios de entrenamiento serán generados mediante muestreo por hipercubo latino, obteniendo el vector $\mathbf{E}^{(k)}$ a partir de $\mathbf{T}^{(k)}$ conforme a la transformación que se presenta a continuación.

$$E^{(k)} = \begin{bmatrix} p_{1,0} \\ \mathrm{SoC}_0 \\ \Delta p_{L,0} \\ t_0 \end{bmatrix} = \begin{bmatrix} 2\,x_1 - 1 \\ 0{,}2 + 0{,}6\,x_2 \\ (x_3 - 0{,}5)(1 - |2\,x_1 - 1|^2) \\ 1 + 4\,x_4 \end{bmatrix} \tag{4.18}$$

Esta estrategia de generación de episodios garantiza que cada simulación inicie desde condiciones iniciales físicamente coherentes con la operación de la planta, evitando escenarios no realistas o inestables. Además, la introducción de una perturbación variante en tiempo y magnitud impide que el agente anticipe un evento específico, promoviendo así un aprendizaje más robusto y generalizable frente a distintas condiciones del entorno.

Esta página se ha dejado intencionadamente en blanco.

Capítulo 5

Resultados y análisis

En el presente capítulo se exponen los resultados obtenidos tras el entrenamiento del agente. En primer lugar, se analizan las curvas que describen la evolución del aprendizaje y permiten evaluar la estabilidad y convergencia del entrenamiento. Luego se examina el comportamiento del agente frente a los escenarios de validación definidos previamente, y finalmente se evalúa su desempeño al ser implementado en el modelo conmutado de la planta de almacenamiento.

5.1. Entrenamiento del agente

En la Figura 5.1 se presentan las curvas asociadas al proceso de entrenamiento del agente. Estas ilustran la evolución de las recompensas acumuladas y la estabilidad del aprendizaje a lo largo de los episodios, permitiendo evaluar la efectividad de la función de recompensa propuesta y el grado de convergencia alcanzado por la política de control.

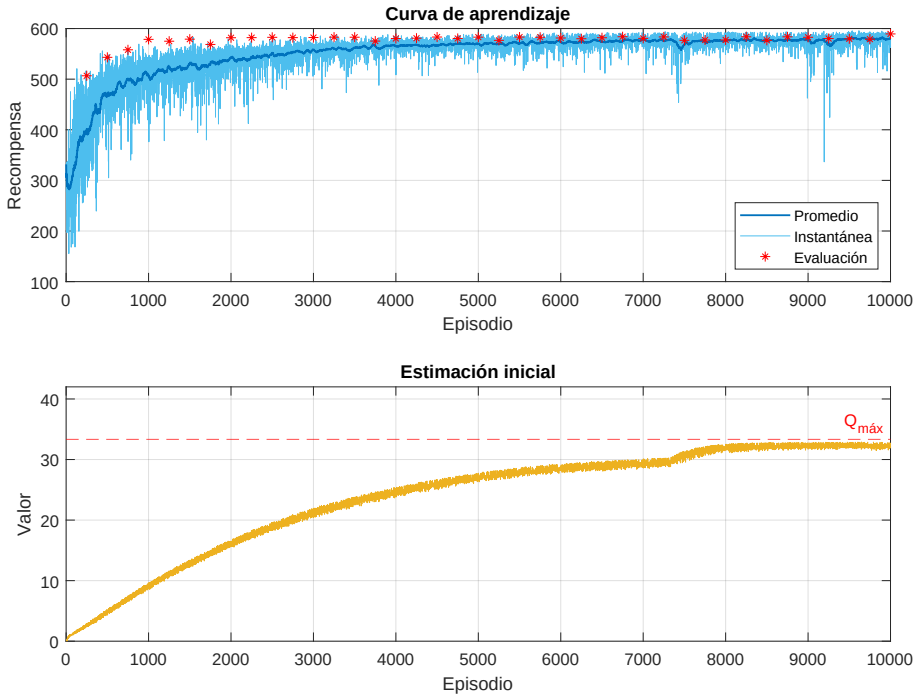

Figura 5.1: Entrenamiento del agente.

Las curvas de aprendizaje obtenidas muestran una evolución progresiva en las recompensas recolectadas por el agente, evidenciando un proceso de adaptación estable a lo largo de los episodios de entrenamiento. En las primeras etapas la recompensa acumulada instantánea presenta una alta dispersión entre episodios consecutivos, lo que refleja la fase inicial de exploración y el carácter aún aleatorio de las acciones tomadas por el agente. Conforme avanza el entrenamiento, esta dispersión disminuye y las recompensas promedio comienzan a incrementarse de forma sostenida, indicando que la política del actor logra capturar patrones de decisión que conducen a un desempeño más consistente. La convergencia de la curva hacia un nivel superior y relativamente estable sugiere que el agente adquirió una estrategia capaz de generalizar adecuadamente frente a distintos escenarios del entorno, sin manifestar sobreajuste ni inestabilidades evidentes en la política aprendida.

En paralelo, la evolución de las estimaciones iniciales del valor Q al comienzo de cada episodio evidencia el proceso de consolidación del crítico. En los episodios más tempranos del entrenamiento estos valores se mantienen cercanos a la inicialización de la red, lo que indica que el crítico aún no ha desarrollado una representación consistente del entorno y, por tanto, sus predicciones iniciales de retorno esperado son bajas y poco informativas. A medida que el entrenamiento progresa y el crítico actualiza sus pesos a partir de la información acumulada en la memoria de experiencias, las estimaciones iniciales de Q se elevan gradualmente y tienden a estabilizarse cerca del valor máximo estimado $Q_{\text{máx}}$, reflejando una mejora en la capacidad del agente para anticipar las recompensas potenciales antes de comenzar cada nuevo episodio. La coherencia entre esta estabilización y el ascenso sostenido de la recompensa promedio sugiere que el crítico alcanza un nivel de generalización adecuado, permitiendo que el actor inicie cada episodio desde un punto de referencia más preciso y contribuya así a un proceso de aprendizaje más eficiente y estable.

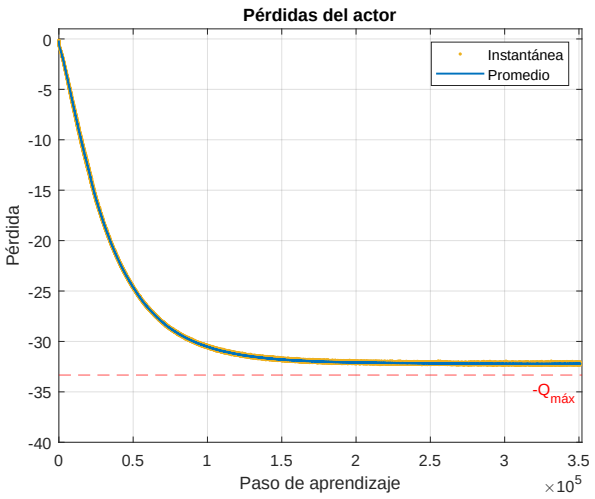

Figura 5.2: Pérdidas del actor.

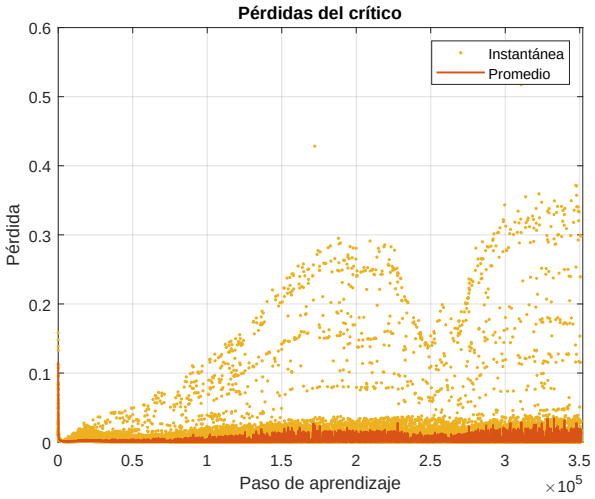

Figura 5.3: Pérdidas del crítico.

En las Figuras 5.2 y 5.3 se presentan respectivamente las pérdidas del actor y del crítico durante el proceso de aprendizaje. La evolución de la pérdida asociada al actor muestra una trayectoria descendente que culmina en una banda estrecha de valores negativos y estables. Cabe señalar que esta magnitud no corresponde a una función de error en sentido estricto, sino a una métrica auxiliar utilizada para monitorear el desempeño del actor a lo largo del entrenamiento. En su formulación teórica, esta métrica se define como el negativo del valor esperado $Q(s, \mu(s))$ estimado por el crítico; en la práctica, dicho valor se aproxima mediante el promedio sobre el minilote de muestras s_i extraídas aleatoriamente de la memoria de experiencias en cada paso de aprendizaje.

$$L_{\text{actor}} = -\mathbb{E}_{s\sim\mathcal{D}}[\,Q(s,\mu(s))\,] \;\approx\; -\frac{1}{N}\sum_{i=1}^{N} Q(s_i,\mu(s_i)) \tag{5.1}$$

En este contexto, una disminución en las pérdidas del actor implica un aumento del valor esperado $Q(s, \mu(s))$, lo que indica que la política está seleccionando acciones con mayor retorno estimado. A medida

que avanza el entrenamiento, esta métrica tiende a estabilizarse en un valor ligeramente superior al negativo del máximo teórico de retorno, lo que sugiere que el agente alcanza una política próxima al desempeño ideal impuesto por la función de recompensa. Este comportamiento es coherente con el carácter aproximado del proceso de aprendizaje, condicionado tanto por la naturaleza estocástica de las actualizaciones como por la precisión limitada del crítico al estimar $Q(s, a)$. Asimismo, la reducción progresiva de la variabilidad entre la pérdida instantánea y la pérdida promedio evidencia que las actualizaciones de política se realizan de forma consistente.

En el caso del crítico, la pérdida presenta una disminución pronunciada durante las primeras etapas del entrenamiento, seguida de una meseta de valores bajos con fluctuaciones acotadas a lo largo de las iteraciones. Este patrón es coherente con un crítico que primero corrige el sesgo inicial de sus estimaciones del valor $Q(s, a)$ y luego entra en un régimen estacionario con errores de predicción reducidos. Las oscilaciones residuales de la pérdida instantánea son esperables debido a la variabilidad de las transiciones almacenadas en la memoria de experiencias y a la naturaleza de la actualización por diferencia temporal. Lo relevante es que la pérdida promedio no deriva al alza ni presenta inestabilidades, lo cual se alinea con el incremento y posterior estabilización de las recompensas promedio, así como con la estabilización de las estimaciones iniciales de $Q(s, a)$ al comienzo de cada episodio. En conjunto, la convergencia de la pérdida del actor y la estabilización de la pérdida del crítico en niveles bajos confirman que ambos componentes alcanzaron un equilibrio de aprendizaje estable, coherente con la mejora sostenida de las recompensas y con la estabilidad observada en las curvas de evaluación.

5.2. Escenarios de validación

A continuación son presentados los resultados del desempeño del agente en los escenarios de evaluación previamente definidos. Estos escenarios fueron diseñados para medir la capacidad del agente de generalizar lo aprendido a situaciones no vistas durante el entrenamiento, lo que permite analizar la efectividad de la política en diferentes contextos.

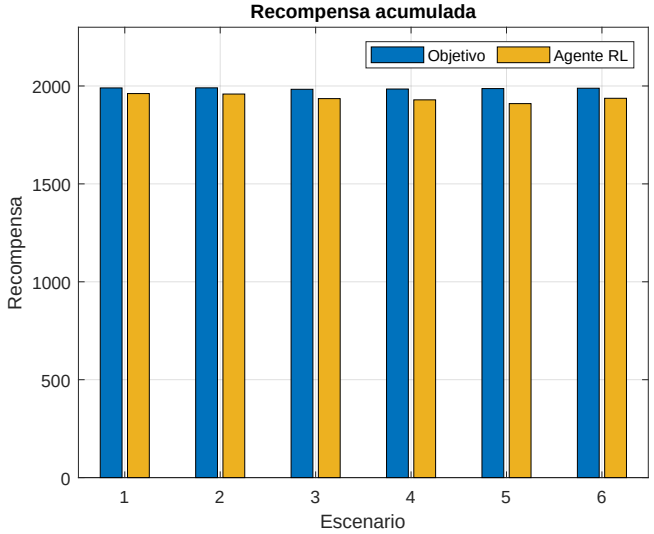

Figura 5.4: Recompensa acumulada no descontada.

En la Figura 5.4 se comparan las recompensas acumuladas no descontadas obtenidas por el agente con las recompensas alcanzadas por la política objetivo. Las barras amarillas representan el desempeño real del agente, mientras que las barras azules muestran el desempeño objetivo. En general, el agente presenta un desempeño consistente a lo largo de los distintos escenarios, con diferencias estables entre las recompensas obtenidas y el objetivo. La desviación mínima respecto de la política objetivo es de un 1.44 % en el Escenario 1, mientras que la desviación máxima alcanza un 3.86 % en el Escenario 5. Estas pequeñas diferencias podrían explicarse por una exploración limitada durante el entrenamiento, lo que impide que el agente descubra

mejores políticas, o por las particularidades de cada escenario, como la variabilidad en las recompensas o comportamientos dinámicos complejos que dificultan la aproximación al desempeño ideal. Sin embargo, el agente muestra una sólida capacidad para generalizar y adaptarse a los diferentes contextos del entorno.

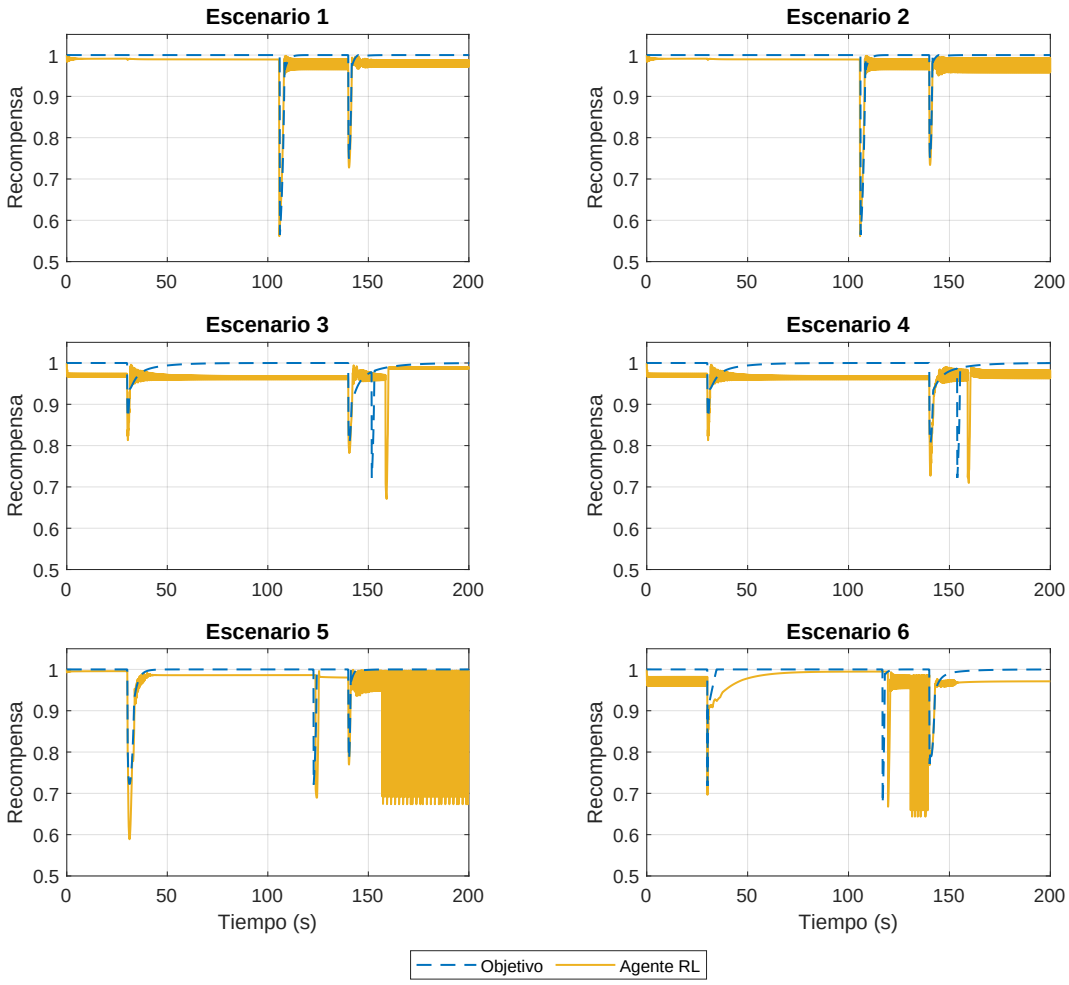

Figura 5.5: Recompensas instantáneas por escenario.

En la Figura 5.5 se presentan las recompensas instantáneas obtenidas por el agente en cada escenario, comparándolas con las recompensas de la política objetivo. En general, el agente ha adoptado una política que busca maximizar las recompensas, aunque no con la misma eficacia que la política objetivo. Si bien el agente se acerca a la unidad en la mayoría de los episodios, sus recompensas instantáneas suelen estar por debajo, en mayor o menor medida, lo que ayuda a explicar las desviaciones observadas en las recompensas acumuladas no descontadas. Además, se observan oscilaciones en las recompensas instantáneas, especialmente en los Escenarios 5 y 6, lo que sugiere que las acciones del agente son oscilantes. Esto indica cierta inestabilidad en el proceso de aprendizaje o en la política aplicada en esos escenarios, lo que resalta la importancia de analizar el comportamiento dinámico de cada uno.

La Figura 5.6 presenta el desempeño del agente durante el primer escenario de evaluación. En términos generales, el agente muestra un buen desempeño en el entorno, ya que sigue en gran medida la política de control definida, respetando el límite inferior del estado de carga y optimizando el despacho de la celda de combustible siempre que es posible. Este buen desempeño resulta en gran medida esperado, ya que este escenario fue el que presentó la menor desviación mínima en el análisis de las recompensas acumuladas no descontadas. Sin embargo, al alcanzar el límite inferior del estado de carga, la potencia proporcionada por la celda de combustible comienza a oscilar levemente, lo que se refleja en las ligeras fluctuaciones de la tensión del enlace de corriente directa.

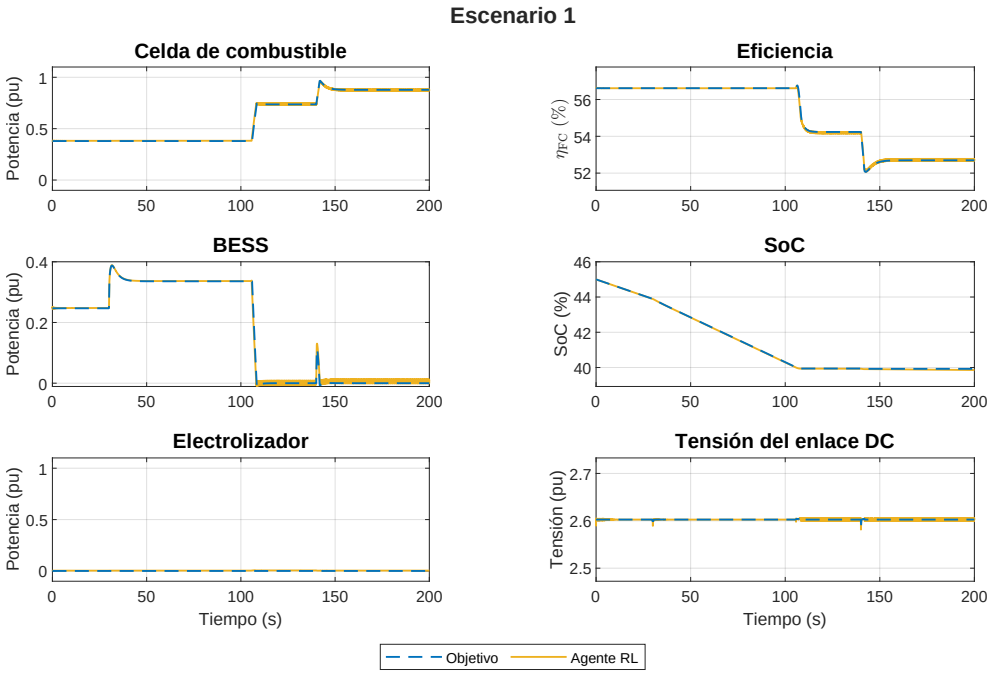

Figura 5.6: Comportamiento dinámico del Escenario 1.

Durante el segundo escenario de evaluación, ilustrado en la Figura 5.7, el desempeño del agente no muestra diferencias sustanciales respecto al escenario previamente analizado. Este comportamiento es esperado, ya que la única diferencia entre ambos escenarios es que el requerimiento de potencia de la red disminuye en lugar de aumentar llegado el instante 140 s. Ante este cambio, el agente responde reduciendo la potencia de la celda de combustible de acuerdo con la política de control definida. Sin embargo, al igual que en el primer escenario, se observan leves oscilaciones en la potencia de la celda de combustible una vez alcanzado el límite inferior del estado de carga del banco de baterías.

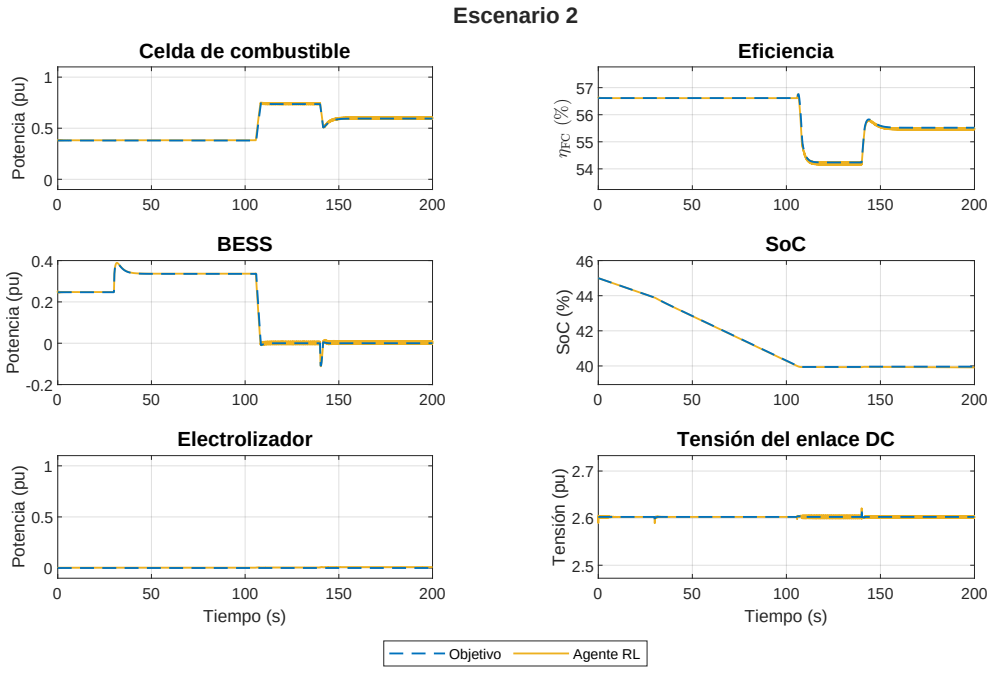

Figura 5.7: Comportamiento dinámico del Escenario 2.

En la Figura 5.8 se muestra el desempeño del agente durante el tercer escenario de evaluación. En términos generales, el desempeño es satisfactorio, aunque se observan nuevamente oscilaciones en las potencias de la celda de combustible, así como también en el electrolizador. Este fenómeno retrasa la carga del banco de baterías, lo que a su vez retrasa el cambio de consigna en la celda de combustible, y también da lugar a oscilaciones notables en la tensión del enlace de corriente directa. Este análisis es extensible a los resultados del Escenario 4, en el que sólo cambia la perturbación impuesta por la red en el instante 140 s.

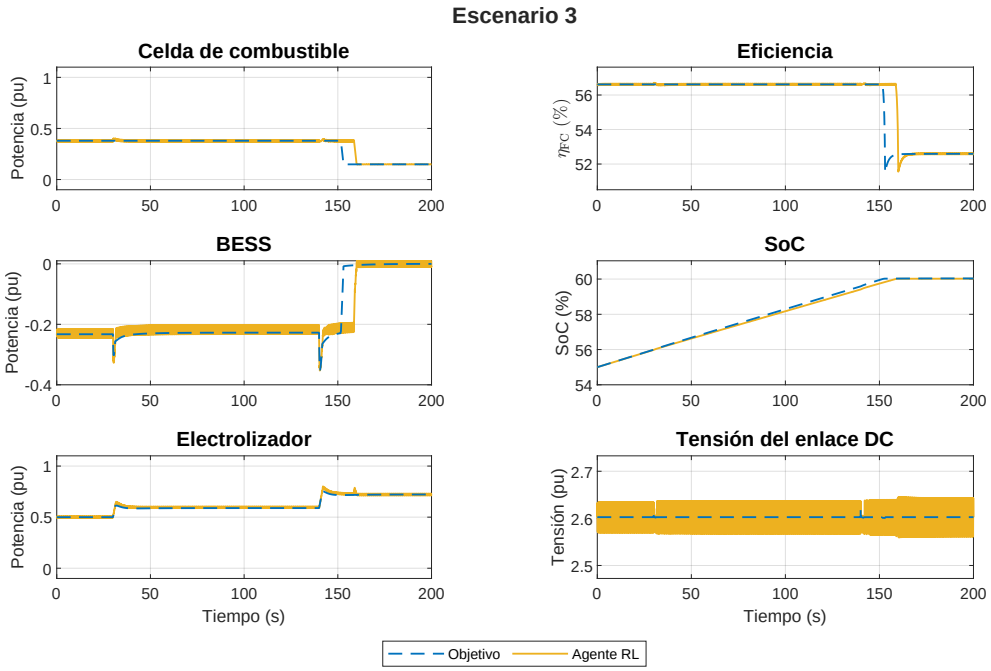

Figura 5.8: Comportamiento dinámico del Escenario 3.

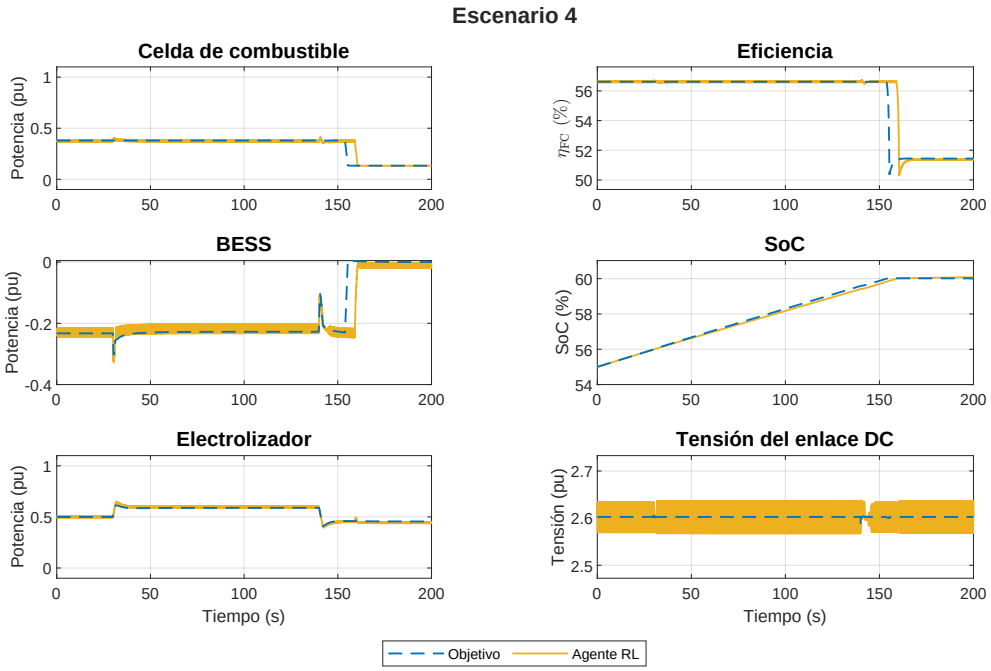

Figura 5.9: Comportamiento dinámico del Escenario 4.

Durante el quinto escenario de evaluación, ilustrado en la Figura 5.10, el desempeño del agente demuestra ser robusto frente al cambio en el sentido del flujo de potencia en el convertidor formador de red, utilizando de manera oportuna la energía provista por la red a través del electrolizador. Además, a lo largo de toda la maniobra mantiene la celda de combustible en su punto óptimo, siempre que el estado de carga esté por debajo del límite superior. No obstante, las oscilaciones vuelven a presentarse de manera notable luego de alcanzarse el límite superior del estado de carga.

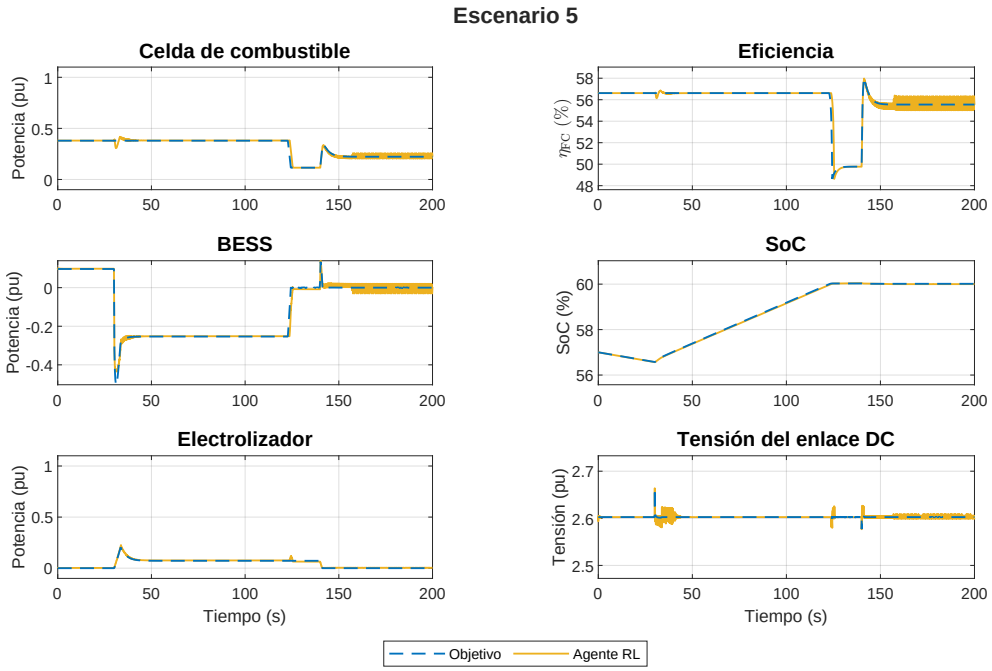

Figura 5.10: Comportamiento dinámico del Escenario 5.

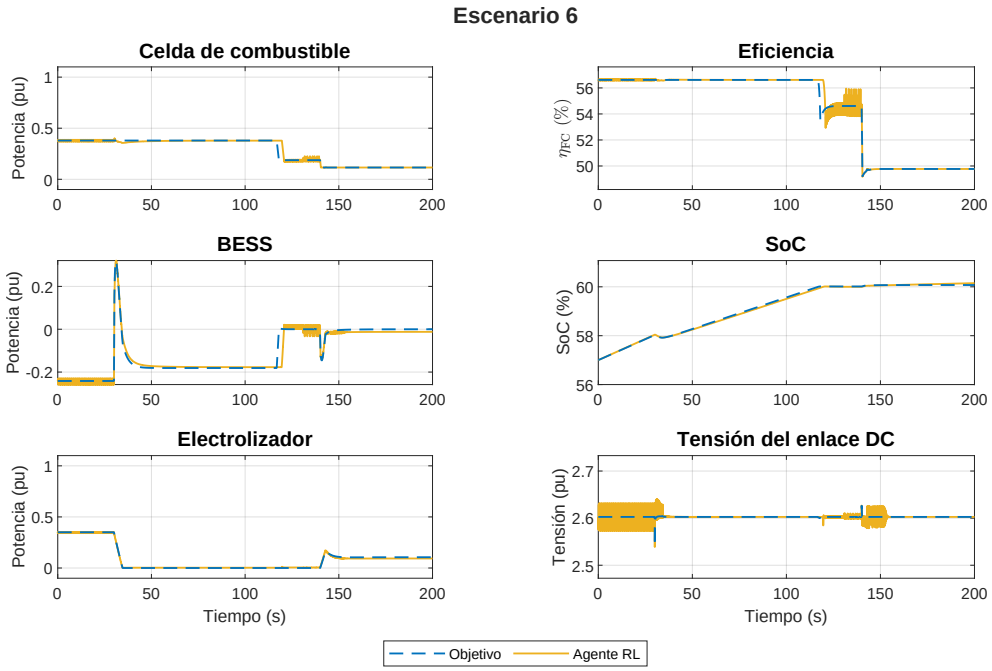

Figura 5.11: Comportamiento dinámico del Escenario 6.

Finalmente, en la Figura 5.11 se presenta el desempeño del agente en el sexto y último escenario de evaluación. Al igual que en el escenario anterior, el agente responde adecuadamente al cambio en el sentido del flujo de potencia en el convertidor formador de red, desactivando el consumo del electrolizador cuando la red requiere potencia. Además, mantiene la celda de combustible en su punto óptimo hasta que el requerimiento de la red y los servicios auxiliares es menor que la potencia óptima, ya que sobrepasar este umbral implicaría incrementar el estado de carga del banco de baterías por encima del límite definido. Este comportamiento indica que el agente ha aprendido a retirar el despacho óptimo no sólo cuando el estado de carga es inferior o igual a 40 %, sino también cuando este es igual o superior al 60 % y los requerimientos de los servicios auxiliares y la red son menores que el punto óptimo.

A partir de la revisión de los escenarios de evaluación, es posible concluir que las oscilaciones observadas en la adquisición instantánea de recompensas se deben a que la política del agente fue adaptada de tal forma que sus acciones son oscilantes, lo que se refleja en las fluctuaciones de potencia del electrolizador y la celda de combustible. Se podría plantear que por alguna razón el agente ha relacionado estas oscilaciones con el hecho de haber alcanzado alguno de los límites en el estado de carga, ya que ocurren prácticamente en todos los escenarios bajo esta condición. Sin embargo, también se observan oscilaciones en los Escenarios 3 y 4 antes de alcanzar el límite superior del estado de carga. A partir de este análisis, se puede concluir que el fenómeno oscilatorio está más relacionado con la adaptación de la política durante el entrenamiento que con una correlación directa con los umbrales del estado de carga. En efecto, la literatura indica que el comportamiento oscilatorio de las acciones es una de las características que presentan los algoritmos basados en redes neuronales, particularmente aquellos con métodos de aprendizaje por refuerzo profundo como DDPG [50].

Las oscilaciones en las acciones del agente influyen indiscutiblemente en la recolección de recompensas durante un episodio, aunque no constituyen la única causa por la cual las recompensas acumuladas no alcanzan los valores de la política objetivo. En todos los escenarios se observa que la recompensa no alcanza exactamente la unidad de forma estacionaria; sin embargo, en los Escenarios 3 y 4 este efecto se acentúa entre los instantes $t \in [50, 125]$. Esto ocurre porque el seguimiento de las consignas es menos preciso que en los demás casos, lo que se traduce en una tasa de carga ligeramente inferior del banco de baterías. Esta variable resulta especialmente representativa del desempeño del agente, ya que cualquier desviación respecto de la referencia obliga al banco de baterías a compensar el desbalance de potencia mediante la acción del controlador de tensión.

En este contexto, la capacidad de seguimiento del agente se ve limitada por las características propias del algoritmo DDPG. Los errores de seguimiento persistentes surgen de la manera en que la política y el crítico son aproximados mediante redes neuronales feedforward, las cuales al carecer de memoria, no pueden corregir los errores acumulados en el tiempo. A ello se suma la ausencia de una acción integradora en la política aprendida, lo que impide eliminar completamente los errores en régimen permanente [51]. Además, al disponer de un tiempo de entrenamiento finito y enfrentar una alta variabilidad en las recompensas, el agente tiende a estabilizar políticas subóptimas. Este comportamiento se ve agravado por la tendencia del crítico a sobreestimar los valores de la política, lo que lleva al actor a considerar óptimas acciones que en realidad conservan pequeñas desviaciones respecto de la referencia [26]. Este fenómeno se ve reflejado en las curvas de aprendizaje, donde la estimación inicial del crítico al final del entrenamiento se aproxima considerablemente al valor teórico máximo, aún cuando el desempeño efectivo de la política es subóptimo. En conjunto, estos factores explican que el seguimiento de referencia no sea completamente efectivo, provocando la intervención del controlador de tensión para restablecer el equilibrio de potencia y la recolección no óptima de recompensas en todos los escenarios.

En resumen, los resultados obtenidos a partir de los escenarios de validación muestran que el agente fue capaz de aprender a discriminar de manera efectiva entre diferentes situaciones, adaptándose correctamente a las condiciones cambiantes del entorno. Aunque se observaron algunas oscilaciones en las recompensas, especialmente en los Escenarios 5 y 6, estas no afectan gravemente el desempeño general del agente ni provocan inestabilidad en la operación. La implementación de la máquina de recompensas demostró ser exitosa, ya que permitió al agente aprender políticas más robustas y coherentes con los requisitos operativos del sistema. Sin embargo, los problemas asociados a la sobreestimación de los valores de la política, característicos del algoritmo DDPG, afectan a la precisión del seguimiento de las consignas, lo que resalta áreas de mejora en la política aplicada.

5.3. Desempeño en modelo conmutado

Con el fin de evaluar la capacidad de generalización del agente entrenado, resulta necesario analizar su desempeño en un modelo conmutado que represente de forma más realista el comportamiento del sistema físico. A diferencia del modelo promedio utilizado durante el entrenamiento, el modelo conmutado incorpora las conmutaciones propias de los convertidores, lo que introduce variaciones rápidas en las señales eléctricas. Para asegurar que las observaciones entregadas al agente conserven una naturaleza continua y eviten la influencia de componentes de alta frecuencia, se implementó un filtro pasa bajos en su entrada. Este filtro posee una frecuencia de corte de 50 Hz, valor que ofrece un compromiso adecuado entre la necesidad de que la dinámica del filtro decaiga con mayor rapidez que el tiempo de muestreo del agente, y la capacidad de atenuar las oscilaciones de alta frecuencia presentes en las señales medidas. Los escenarios en los que es estudiado el desempeño del agente inmerso en el modelo conmutado son detallados a continuación.

- **Escenario 1:** En el instante 0 s el convertidor formador de red se encuentra desconectado de la red trifásica, y procede a sincronizar su tensión v_2 con la tensión del punto de conexión. Una vez alcanzada la sincronización, en el instante 1 s se realiza la conexión a la red. Por su parte, el estado de carga inicial del banco de baterías es del 42 %. En el instante 5 s se produce una perturbación en la red requiriendo de la extracción de potencia desde la planta de almacenamiento.

- **Escenario 2:** En el instante 0 s el convertidor formador de red se encuentra desconectado de la red trifásica, y procede a sincronizar su tensión v_2 con la tensión del punto de conexión. Una vez alcanzada la sincronización, en el instante 1 s se realiza la conexión a la red. Por su parte, el estado de carga inicial del banco de baterías es del 58 %. En el instante 5 s se produce una perturbación en la red requiriendo de la absorción de potencia por parte de la planta de almacenamiento.

- **Escenario 3:** En el instante 0 s el convertidor formador de red se encuentra desconectado de la red trifásica, y procede a sincronizar su tensión v_2 con la tensión del punto de conexión. Una vez alcanzada la sincronización, en el instante 1 s se realiza la conexión a la red. Por su parte, el estado de carga inicial del banco de baterías es del 50 %. En el instante 5 s se produce una perturbación en la red requiriendo de la extracción de potencia desde la planta de almacenamiento. Posteriormente, en el instante 30 s ha lugar una nueva perturbación en la red requiriendo de la absorción de potencia por parte de la planta de almacenamiento.

A continuación se presentan los resultados correspondientes al primer escenario analizado. En este caso, se observa que el comportamiento del agente no difiere de manera sustancial respecto de la operación obtenida en el modelo promedio, lo que preliminarmente indica una adecuada capacidad de adaptación frente a las conmutaciones del inversor y las variaciones rápidas presentes en el modelo conmutado.

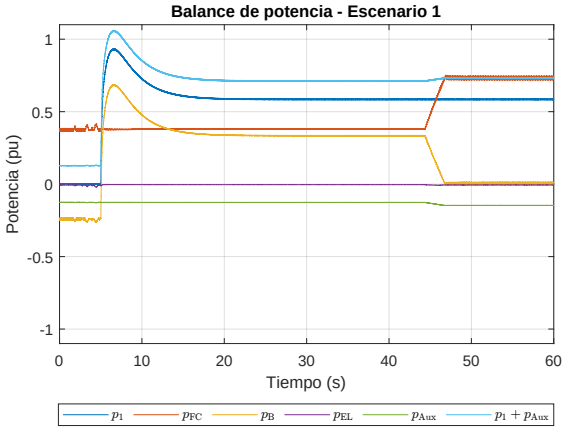

Figura 5.12: Balance de potencia - Caso 1.

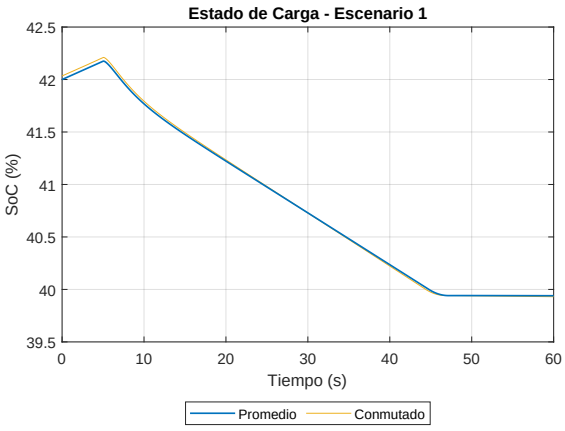

Figura 5.13: Estado de carga - Caso 1.

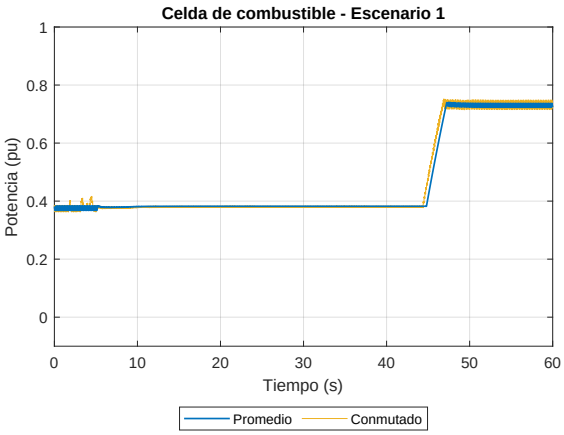

Figura 5.14: Celda de combustible - Caso 1.

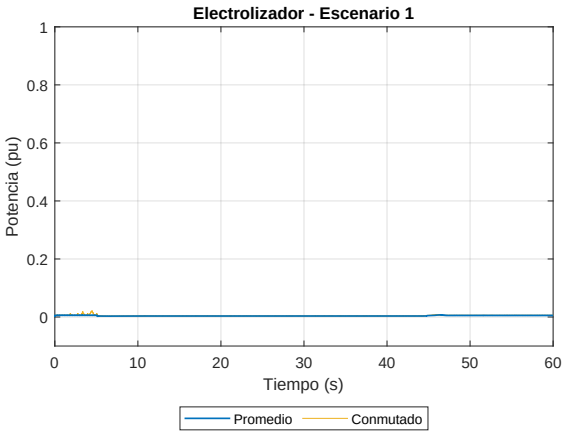

Figura 5.15: Electrolizador - Caso 1.

El desempeño del electrolizador y de la celda de combustible resulta prácticamente idéntico al observado en el modelo promedio. Esto se verifica tanto en las curvas de potencia de cada equipo como en la evolución del estado de carga del sistema de baterías, el cual no presenta diferencias significativas entre ambos modelos. Este resultado sugiere que la asignación de consignas por parte del agente es en esencia equivalente, manteniendo una distribución coherente de potencia entre los distintos subsistemas.

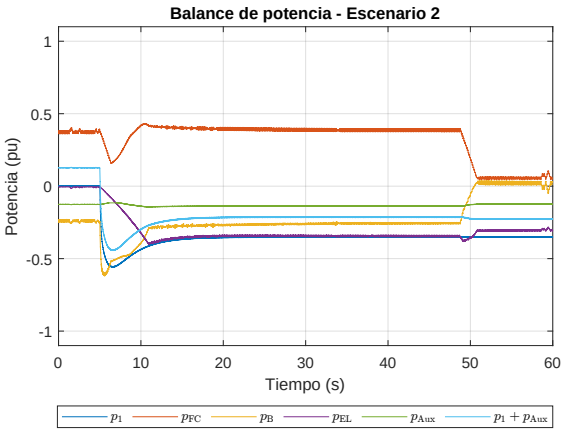

Figura 5.16: Balance de potencia - Caso 2.

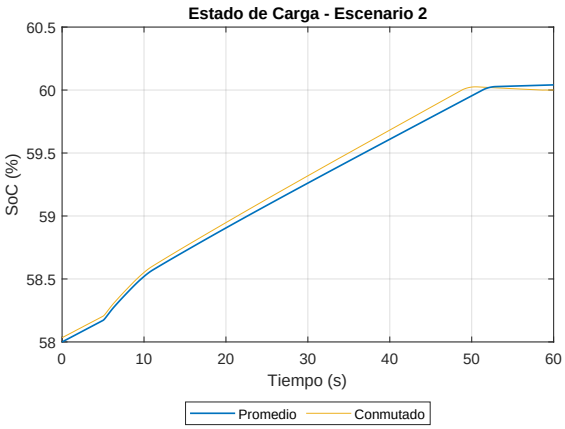

Figura 5.17: Estado de carga - Caso 2.

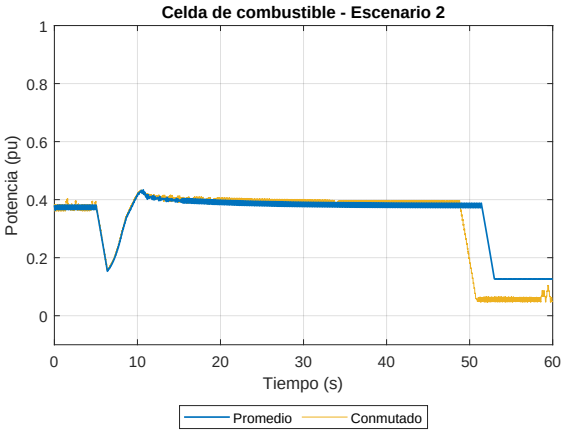

Figura 5.18: Celda de combustible - Caso 2.

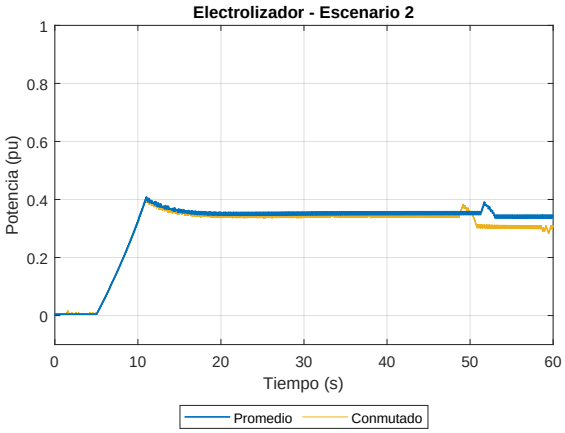

Figura 5.19: Electrolizador - Caso 2.

Desde la Figura 5.16 a la Figura 5.19 se presenta el desempeño del agente en el segundo escenario definido. Se observan diferencias evidentes entre la operación obtenida en el modelo promedio y la correspondiente al modelo conmutado, particularmente hacia el final del escenario, lo que sugiere que las conmutaciones de los convertidores y las componentes de alta frecuencia influyeron de manera significativa en las acciones del agente. Las diferencias observadas obedecen, de manera indiscutible, a que el agente percibió un estado distinto al del modelo promedio, hecho que resulta coherente considerando que la política aprendida es de carácter determinista. Este resultado pone de manifiesto que entrenar un agente mediante el algoritmo DDPG en un modelo promedio para luego implementarlo en un modelo conmutado no constituye una buena práctica, pese a las ventajas computacionales que ello ofrece. El carácter determinista de la política provoca que las acciones sean altamente sensibles a las observaciones, por lo que para alcanzar un desempeño satisfactorio, es más robusto entrenar al agente directamente en el entorno conmutado o, en su defecto, emplear una heurística de aprendizaje que no dependa de forma determinista de las observaciones, como PPO u otra alternativa que permita conservar las ventajas computacionales del modelo promedio sin comprometer la coherencia del control.

En los siguientes gráficos se presentan los resultados correspondientes al tercer y último escenario evaluado en el modelo conmutado. Se observa que el desempeño del agente es notablemente similar al obtenido en el modelo promedio, ya que las potencias inyectadas y consumidas por la celda de combustible y el electrolizador son prácticamente idénticas en términos agregados. La única diferencia apreciable se encuentra en la evolución del estado de carga del banco de baterías; sin embargo, considerando la similitud de las potencias asociadas a la celda y al electrolizador en ambos modelos, dicha diferencia puede considerarse tolerable y no compromete la coherencia general del comportamiento del sistema.

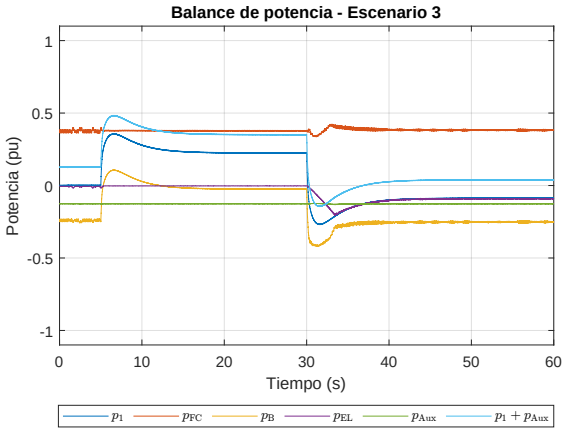

Figura 5.20: Balance de potencia - Caso 3.

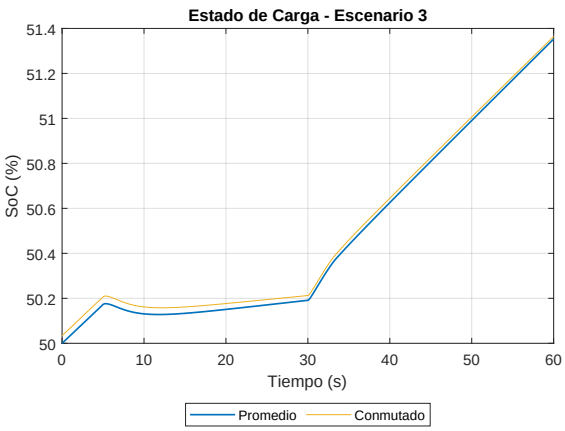

Figura 5.21: Estado de carga - Caso 3.

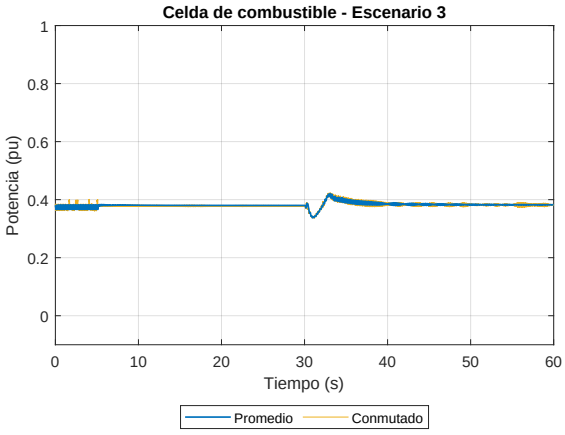

Figura 5.22: Celda de combustible - Caso 3.

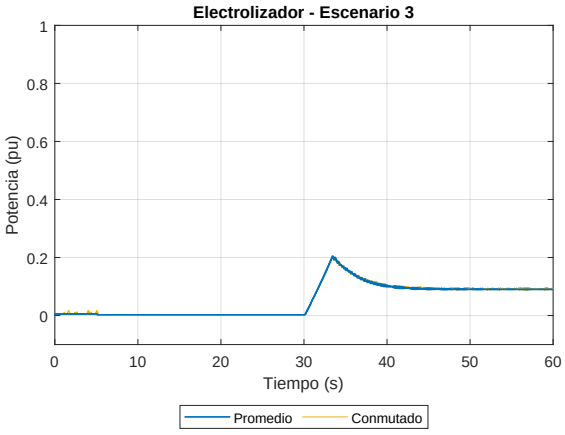

Figura 5.23: Electrolizador - Caso 3.

En la Figura 5.24 se presenta la recolección instantánea de recompensas obtenida por el agente tanto en el modelo promedio como en el modelo conmutado, para los tres escenarios previamente evaluados. Se aprecia que en los Escenarios 1 y 3 la recolección de recompensas es muy similar, lo que resulta coherente con los resultados presentados anteriormente. En cambio, en el Escenario 2 se observa una ligera degradación del desempeño del agente en el modelo conmutado, ya que hacia el final del escenario la recompensa obtenida es inferior a la registrada en el modelo promedio. Este resultado refuerza la conclusión de que la metodología utilizada para el entrenamiento del agente no es del todo adecuada.

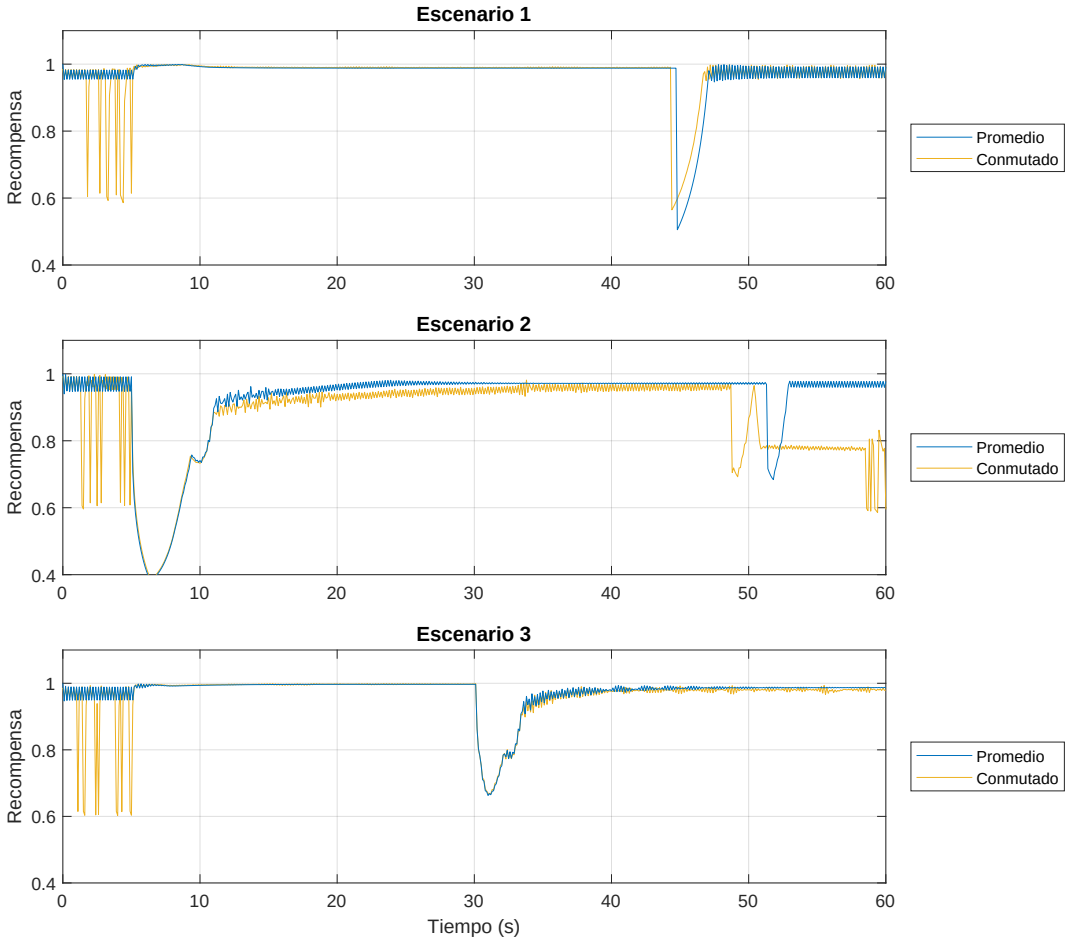

Figura 5.24: Recompensas instantáneas.

Finalmente, en la Figura 5.25 se muestra la evolución de la tensión del enlace de corriente directa para los tres escenarios evaluados en el modelo conmutado. El comportamiento observado es estable, manteniéndose dentro de los márgenes de operación definidos a pesar de las perturbaciones introducidas por las acciones del agente y el ruido propio de la conmutación. Este resultado evidencia que las oscilaciones presentes en las acciones del agente no representan, en principio, un riesgo para la estabilidad ni inducen fenómenos de resonancia en la la tensión del enlace de corriente directa. No obstante, ello no implica que dichas oscilaciones deban ser aceptadas, pudiendo implementarse estrategias de suavizado como las propuestas en [50] para mejorar la calidad dinámica del control.

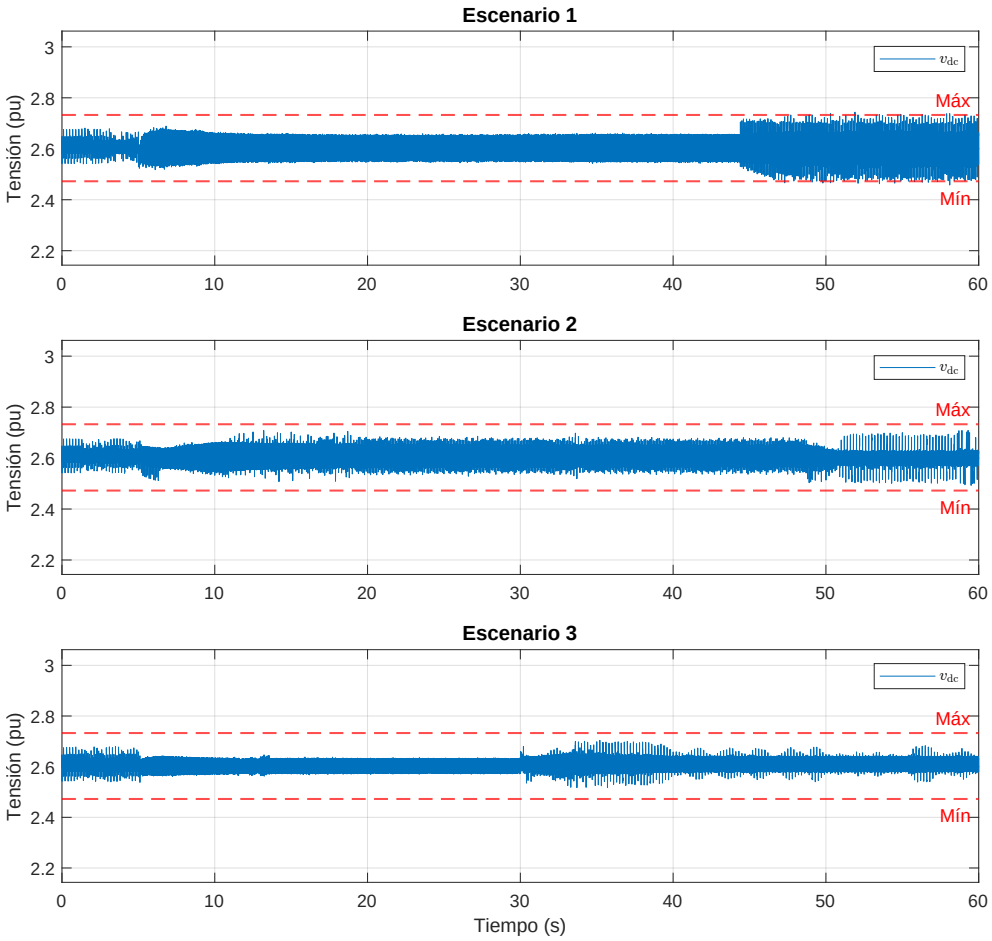

Figura 5.25: Tensión del enlace en corriente directa.

Esta página se ha dejado intencionadamente en blanco.

Capítulo 6

Conclusiones

En el presente trabajo se diseñó un sistema de control cuya finalidad es gestionar los intercambios energéticos de una planta de almacenamiento híbrida, compuesta por un electrolizador, una celda de combustible y un banco de baterías, todos conectados a una red trifásica mediante un convertidor formador de red. El sistema de control se implementó mediante un agente de aprendizaje por refuerzo entrenado mediante el algoritmo de gradiente de política determinista con redes profundas (DDPG, por sus siglas en inglés). Este agente sigue una política condicional diseñada para optimizar la gestión del hidrógeno y cumplir las principales restricciones operativas de la planta, entre ellas las rampas de toma de carga, los límites de eficiencia y la disponibilidad permanente del banco de baterías para amortiguar los transitorios de potencia. La política es implementada en el agente durante su entrenamiento mediante una máquina de recompensas, cuyo propósito es guiar el aprendizaje hacia acciones que respeten la estructura condicional definida y las consignas óptimas. Para contener el costo computacional de entrenar al agente en un entorno con fenómenos de alta frecuencia, se empleó un modelo promedio como ambiente de entrenamiento, capaz de capturar las dinámicas realmente relevantes para los intercambios energéticos.

El entrenamiento del agente fue satisfactorio, demostrando que las funciones de recompensa diseñadas, los hiperparámetros configurados y la generación de episodios mediante muestreo por hipercubo latino, fueron definiciones y metodologías apropiadas para alcanzar la convergencia de la política. Adicionalmente, utilizar funciones de recompensa acotadas dio paso a un proceso de aprendizaje estable y de sencilla interpretación, permitiendo observar explícitamente la convergencia de la estimación inicial del crítico al valor máximo teórico. El uso de una máquina de recompensas, en lugar de una función de recompensa condicional aislada, resultó efectivo para abordar los problemas descritos en la literatura relacionados a políticas condicionales. Esto se evidenció en un proceso de aprendizaje consistente y en la correcta identificación de los umbrales lógicos durante la validación, tanto en el modelo promedio como en el modelo conmutado. Si bien los resultados fueron en su mayoría satisfactorios, durante la validación y la evaluación del desempeño del agente en el modelo conmutado, fue posible observar algunas de las debilidades del algoritmo de aprendizaje adoptado. En el marco de los escenarios de validación se observaron oscilaciones en las acciones del agente que, aunque no comprometieron la estabilidad de la operación, resultan indeseables desde la perspectiva del control, y se constató además un seguimiento de consignas menos preciso de lo deseado, evidenciado por la ausencia de recompensas máximas en régimen estacionario. Finalmente, la evaluación del desempeño del agente en el modelo conmutado sugiere que, para este caso de estudio, el uso de un modelo promedio durante el entrenamiento no es una buena práctica, pues se observaron diferencias importantes en el comportamiento del agente entre ambos modelos, al menos en uno de los escenarios analizados.

El aporte de la planta de almacenamiento al control de frecuencia se materializó mediante la acción coordinada del convertidor formador de red, y de la estrategia adaptativa aplicada en el control de la tensión del enlace de corriente directa. En el convertidor trifásico se adoptó una arquitectura de control capaz de aportar inercia virtual sin recurrir a un lazo de seguimiento de fase, replicando así parte de los resultados reportados en el artículo donde se introduce esta estrategia de control. La sintonización del lazo interno mediante un algoritmo de partículas mostró un desempeño consistente y contribuyó a mitigar las dificultades de ajuste habitualmente reportadas para este tipo de controladores. Asimismo, el convertidor se modeló tanto en su versión promedio como en su versión conmutada, verificándose de manera indirecta la efectividad

de operar sin seguimiento de fase en ambos marcos. En paralelo, el control de la tensión del enlace de corriente directa adoptó una estrategia adaptativa que habilitó una respuesta rápida frente a las solicitudes de potencia del convertidor, fortaleciendo la estabilidad transitoria del sistema. De manera análoga al convertidor formador de red, se verificó indirectamente que la estrategia de control empleada resultó funcional tanto en el modelo promedio como en el modelo conmutado.

A partir de la revisión del estado del arte se establecieron modelos para los sistemas de almacenamiento, con lo cual fue posible describir de manera integral el comportamiento dinámico del sistema estudiado. Estos modelos capturaron los tiempos de respuesta más lentos de la celda de combustible y del electrolizador, la dependencia de la tensión interna del banco de baterías con su estado de carga, las rampas de toma de carga y las eficiencias globales de los subsistemas basados en hidrógeno, además de la capacidad de almacenamiento de las baterías. Asimismo, dichos modelos fueron escalados con datos experimentales reportados en la literatura, lo que añadió realismo a la operación de la planta de almacenamiento. Por último, los modelos promedio de cada componente se validaron al contrastarlos con sus versiones conmutadas, observándose una concordancia suficiente para los objetivos de análisis y control planteados.

A partir de los hallazgos, tanto favorables como desfavorables, se proponen a continuación líneas de trabajo futuro orientadas a ampliar y mejorar los resultados presentados en este documento.

- Al evaluar el desempeño del agente en el modelo conmutado se advirtió que las acciones del Escenario 2 difirieron de manera apreciable entre el modelo promedio y su símil conmutado. Este comportamiento es coherente con la naturaleza determinista de la política utilizada en DDPG: si las acciones cambiaron, fue porque las observaciones cambiaron. No obstante, dado que el modelo promedio es una simplificación del conmutado, cabría esperar decisiones similares. La discrepancia sugiere, por tanto, una sensibilidad no despreciable del agente a pequeñas variaciones en las observaciones y a las simplificaciones propias del modelo promedio. Para abordar este fenómeno se proponen dos líneas de trabajo; la primera consiste en utilizar un algoritmo de entrenamiento basado en una política estocástica como PPO o SAC, pues el carácter no determinista de las acciones robustece su desempeño frente a variaciones en las observaciones. La segunda línea de trabajo propone seguir utilizando el algoritmo DDPG, pero incorporando ruido controlado en las observaciones con el fin de reducir la sensibilidad de la política a pequeñas variaciones, y así mejorar su capacidad de generalización. En ambas propuestas se mantiene el uso de un modelo promedio durante el entrenamiento, con la finalidad de no incrementar el coste computacional.

- La configuración de hiperparámetros requirió un proceso de prueba y error que extendió de manera considerable los tiempos de ajuste del algoritmo. Como línea futura, se propone emplear optimización bayesiana para su configuración, pues ello reduce el número de ensayos, permite hallar configuraciones de mejor desempeño con menos simulaciones, automatiza buena parte del ajuste y ordena el proceso, haciéndolo más reproducible. En conjunto, dicha metodología acorta plazos, disminuye el esfuerzo operativo y aumenta la probabilidad de que el agente converja hacia una política satisfactoria [52].

- Durante la validación de la política se observaron un seguimiento de referencia impreciso y oscilaciones indeseadas en las acciones del agente. Para mejorar estos aspectos, se propone como trabajo futuro incorporar suavizado explícito de la política mediante términos de regularización sobre las acciones, de modo que estados similares y tiempos consecutivos generen salidas similares. Este enfoque atenúa las variaciones bruscas, reduce la sensibilidad al ruido y al desajuste de modelo, disminuyendo además la dependencia de ajustes finos en la función de recompensa [50]. En complemento, se sugiere ampliar el agente con una realimentación integral en la salida del actor, incorporando el error acumulado como señal auxiliar para reforzar el seguimiento en régimen permanente y la compensación de perturbaciones. Esta ampliación se integra con cambios mínimos en la arquitectura, es compatible con esquemas actor–crítico habituales y puede acompañarse de mecanismos de limitación y anti–saturación para mantener una dinámica estable y bien amortiguada [51].

- Los umbrales lógicos de la política de control se definieron de manera simétrica respecto del rango operativo del estado de carga; sin embargo, reducir la banda de reserva destinada a compensar transitorios puede mejorar la gestión del recurso primario. En consecuencia, como trabajo futuro se propone sensibilizar dicho umbral en función de los requerimientos del control primario de frecuencia en la barra de conexión y realizar un estudio energético - económico de la política de control, comparando su desempeño con una estrategia que asigna al sistema de baterías exclusivamente la compensación de transitorios.

- La sintonización del convertidor formador de red no fue el objetivo principal de este trabajo; el foco se mantuvo en contar con un modelo estable y funcional que habilitara la evaluación de la estrategia de control propuesta. Como línea futura, se propone profundizar la sintonización del lazo de control interno mediante algoritmo de partículas, extendiendo el análisis a otras arquitecturas de convertidores formadores de red y examinando aspectos adicionales de interés, como la dependencia con el índice de cortocircuito, la variación de la impedancia de red y la operación bajo condiciones de cortocircuito. Este estudio permitiría caracterizar con mayor detalle la robustez, los márgenes de estabilidad y el desempeño transitorio del sistema.

Esta página se ha dejado intencionadamente en blanco.

Bibliografía

[1] Hesam Pishbahar, Frede Blaabjerg, and Hedayat Saboori. Emerging grid-forming power converters for renewable energy and storage resources integration – a review. *Sustainable Energy Technologies and Assessments*, 60:103538, 2023.

[2] K.H. Ahmed, S.J. Finney, and B.W. Williams. Passive filter design for three-phase inverter interfacing in distributed generation. In *2007 Compatibility in Power Electronics*, pages 1–9, 2007.

[3] Roberto Rosso, Xiongfei Wang, Marco Liserre, Xiaonan Lu, and Soenke Engelken. Grid-forming converters: Control approaches, grid-synchronization, and future trends—a review. *IEEE Open Journal of Industry Applications*, 2:93–109, 2021.

[4] Myada Shadoul, Razzaqul Ahshan, Rashid S. AlAbri, Abdullah Al-Badi, Mohammed Albadi, and Mohsin Jamil. A comprehensive review on a virtual-synchronous generator: Topologies, control orders and techniques, energy storages, and applications. *Energies*, 15(22), 2022.

[5] Haniyeh Marefatjouikilevaee, Francois Auger, and Jean-Christophe Olivier. Static and dynamic electrical models of proton exchange membrane electrolysers: A comprehensive review. *Energies*, 16(18), 2023.

[6] Mostafa El-Shafie. Hydrogen production by water electrolysis technologies: A review. *Results in Engineering*, 20:101426, 2023.

[7] A.Z. Arsad, M.A. Hannan, Ali Q. Al-Shetwi, R.A. Begum, M.J. Hossain, Pin Jern Ker, and TM Indra Mahlia. Hydrogen electrolyser technologies and their modelling for sustainable energy production: A comprehensive review and suggestions. *International Journal of Hydrogen Energy*, 48(72):27841–27871, 2023.

[8] Mohamed Khalid Ratib, Kashem M. Muttaqi, Md Rabiul Islam, Danny Sutanto, and Ashish P. Agalgaonkar. Electrical circuit modeling of proton exchange membrane electrolyzer: The state-of-the-art, current challenges, and recommendations. *International Journal of Hydrogen Energy*, 49:625–645, 2024.

[9] Mehdi Ghazavi Dozein, Ahvand Jalali, and Pierluigi Mancarella. Fast frequency response from utility-scale hydrogen electrolyzers. *IEEE Transactions on Sustainable Energy*, 12(3):1707–1717, 2021.

[10] Lixin Fan, Zhengkai Tu, and Siew Hwa Chan. Recent development of hydrogen and fuel cell technologies: A review. *Energy Reports*, 7:8421–8446, 2021.

[11] Yujie Wang, Xingliang Yang, Zhengdong Sun, and Zonghai Chen. A systematic review of system modeling and control strategy of proton exchange membrane fuel cell. *Energy Reviews*, 3(1):100054, 2024.

[12] Fuel Cell Technologies Office. Comparison of fuel cell technologies. `https://www.energy.gov/eere/fuelcells/comparison-fuel-cell-technologies`, 2016. Revisado el 05 de Enero de 2025.

[13] Souleman Njoya M., Olivier Tremblay, and Louis-A. Dessaint. A generic fuel cell model for the simulation of fuel cell vehicles. In *2009 IEEE Vehicle Power and Propulsion Conference*, pages 1722–1729, 2009.

[14] M. J. Khan and M. T. Iqbal. Modelling and analysis of electro-chemical, thermal, and reactant flow dynamics for a pem fuel cell system. *Fuel Cells*, 5(4):463–475, 2005.

[15] Minyung Cha, Shantha G Jayasinghe, Hossein Enshaei, Rabiul Islam, Apsara Abeysiriwardhane, and Sanath Alahakoon. Power management optimisation of a battery/fuel cell hybrid electric ferry. In *2021 31st Australasian Universities Power Engineering Conference (AUPEC)*, pages 1–6, 2021.

[16] M.A. Hannan, S.B. Wali, P.J. Ker, M.S. Abd Rahman, M. Mansor, V.K. Ramachandaramurthy, K.M. Muttaqi, T.M.I. Mahlia, and Z.Y. Dong. Battery energy-storage system: A review of technologies, optimization objectives, constraints, approaches, and outstanding issues. *Journal of Energy Storage*, 42:103023, 2021.

[17] Yuqing Chen, Yuqiong Kang, Yun Zhao, Li Wang, Jilei Liu, Yanxi Li, Zheng Liang, Xiangming He, Xing Li, Naser Tavajohi, and Baohua Li. A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards. *Journal of Energy Chemistry*, 59:83–99, 2021.

[18] Olivier Tremblay and Louis-A. Dessaint. Experimental validation of a battery dynamic model for ev applications. *World Electric Vehicle Journal*, 3(2):289–298, 2009.

[19] Kaiyuan Li and King Jet Tseng. Energy efficiency of lithium-ion battery used as energy storage devices in micro-grid. In *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, pages 005235–005240, 2015.

[20] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2nd edition, 2018.

[21] Nikhil Buduma, Nithin Buduma, and Joe Papa. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly Media, 2 edition, 2022.

[22] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

[23] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[24] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4):229–256, 1992.

[25] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[26] Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1587–1596. PMLR, 2018.

[27] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1861–1870. PMLR, 2018.

[28] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 1928–1937. PMLR, 2016.

[29] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[30] Richard S. Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the Seventh International Conference on Machine Learning*, pages 216–224, 1990.

[31] Michael Janner, Josh Fu, Marvin Zhang, and Sergey Levine. When to trust your model: Model-based policy optimization. In *Advances in Neural Information Processing Systems*, volume 32, pages 12519–12530, 2019.

[32] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In *Advances in Neural Information Processing Systems*, volume 31, pages 4754–4765, 2018.

[33] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 2555–2565. PMLR, 2019.

[34] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

[35] Yuelin Luo, Tieqiang Gang, and Lijie Chen. Research on target defense strategy based on deep reinforcement learning. *IEEE Access*, 10:82329–82335, 2022.

[36] Aleksandr Reznik, Marcelo Godoy Simões, Ahmed Al-Durra, and S. M. Muyeen. LCL filter design and performance analysis for grid-interconnected systems. *IEEE Transactions on Industry Applications*, 50(2):1225–1232, 2014.

[37] IEEE. IEEE Standard for Harmonic Control in Electric Power Systems. *IEEE Std 519-2022 (Revision of IEEE Std 519-2014)*, pages 1–31, 2022.

[38] Taoufik Qoria, Ebrahim Rokrok, Antoine Bruyere, Bruno François, and Xavier Guillaud. A PLL-Free Grid-Forming Control With Decoupled Functionalities for High-Power Transmission System Applications. *IEEE Access*, 8:197363–197378, 2020.

[39] Taoufik Qoria. *Grid-forming control to achieve a 100 % power electronics interfaced power transmission systems*. Thèse de doctorat, HESAM Université, Lille, France, Novembre 2020. Préparée à l'École Nationale Supérieure d'Arts et Métiers, spécialité Génie Électrique.

[40] Damien Guilbert and Gianpaolo Vitale. Experimental validation of an equivalent dynamic electrical model for a proton exchange membrane electrolyzer. In *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe*, pages 1–6, 2018.

[41] Xiangjun Quan, Qinran Hu, Xiaobo Dou, Zaijun Wu, Ling Zhu, and Wei Li. Control of grid-forming application for fuel cell/electrolyser system. *IET Renewable Power Generation*, 14(17):3368–3374, 2021.

[42] François Parache, Henri Schneider, Christophe Turpin, Nicolas Richet, Olivier Debellemanière, Éric Bru, Anh Thao Thieu, Caroline Bertail, and Christine Marot. Impact of power converter current ripple on the degradation of pem electrolyzer performances. *Membranes*, 12(2), 2022.

[43] Bouchra Wahdame, Laurent Girardot, Daniel Hissel, Fabien Harel, Xavier Francois, Denis Candusso, Marie Cecile Pera, and Laurent Dumercy. Impact of power converter current ripple on the durability of a fuel cell stack. In *2008 IEEE International Symposium on Industrial Electronics*, pages 1495–1500, 2008.

[44] Gioacchino Musicò, Salvatore Gianluca Leonardi, Giovanni Lucà Trombetta, Giovanni Brunaccini, Francesco Salmeri, Davide Aloisio, and Francesco Sergi. A methodology for optimal energy management for efficient and profitable hydrogen production and storage. *Rendiconti Lincei. Scienze Fisiche e Naturali*, 36:385–406, 2025.

[45] Haimin Wang, Chuanwei Wang, Chenglong Jiang, Jiangang Zhou, Wenqin Liu, Zhiyuan Ji, Guodong Meng, and Feng Zhao. High-power charging strategy within key soc ranges based on heat generation of lithium-ion traction battery. *Journal of Energy Storage*, 72:108125, 2023.

[46] H. Sayed-Ahmed, Á.I. Toldy, and A. Santasalo-Aarnio. Dynamic operation of proton exchange membrane electrolyzers—critical review. *Renewable and Sustainable Energy Reviews*, 189:113883, 2024.

[47] Rodrigo Toro Icarte, Toryn Q. Klassen, Richard Valenzano, and Sheila A. McIlraith. Reward machines: Exploiting reward function structure in reinforcement learning. *Journal of Artificial Intelligence Research*, 73:173–208, 2022.

[48] Lixing Wang and Huirong Jiao. Multi-agent reinforcement learning-based computation offloading for unmanned aerial vehicle post-disaster rescue. *Sensors*, 24(24), 2024.

[49] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[50] Siddharth Mysore, Bassel Mabsout, Renato Mancuso, and Kate Saenko. Regularizing action policies for smooth control with reinforcement learning, 2021.

[51] Daniel Weber, Maximilian Schenke, and Oliver Wallscheid. Steady-state error compensation in reference tracking and disturbance rejection problems for reinforcement learning-based control, 2022.

[52] Jinhai Wang, Changqing Du, Fuwu Yan, Min Hua, Xiangyu Gongye, Quan Yuan, Hongming Xu, and Quan Zhou. Bayesian optimization for hyper-parameter tuning of an improved twin delayed deep deterministic policy gradients based energy management strategy for plug-in hybrid electric vehicles. *Applied Energy*, 381:125171, 2025.

Apéndice A

Sistema por unidad para simulaciones dinámicas

En el análisis de sistemas eléctricos en estado transitorio resulta conveniente expresar las ecuaciones en un sistema normalizado en por unidad. Para ello, la tensión, corriente y frecuencia base se definen como se muestra a continuación.

- V_b: Valor peak de la tensión nominal fase - neutro en V.
- I_b: Valor peak de la corriente nominal de línea en A.
- f_b: Frecuencia nominal en Hz.

Así, las demás cantidades base se obtienen según

- $\omega_b = 2\pi f_b$ en rad eléc/s
- $Z_b = V_b/I_b$ en Ω
- $L_b = Z_b/\omega_b$ en H
- $C_b = (Z_b\,\omega_b)^{-1}$ en F
- $S_b = 3/2\,V_b I_b$ en VA

donde

- ω_b: Frecuencia angular base
- Z_b: Impedancia base
- L_b: Inductancia base
- C_b: Capacitancia base
- S_b: Potencia aparente base

Es importante destacar que para el caso de estudio se definió convenientemente una potencia aparente base de 1 MVA, junto a una tensión base de $400\sqrt{2/3}$ V.

Esta página se ha dejado intencionadamente en blanco.

Apéndice B

Escalado de modelos

Electrolizador

En la Figura B.1 se presenta la disposición eléctrica de un conjunto de electrolizadores menores que constituyen un electrolizador mayor. La disposición considera m ramas conectadas en paralelo, donde cada una se compone de n electrolizadores menores conectados en serie. El modelo considerado para cada electrolizador menor es el que se presenta en [8] para bajas densidades de corriente.

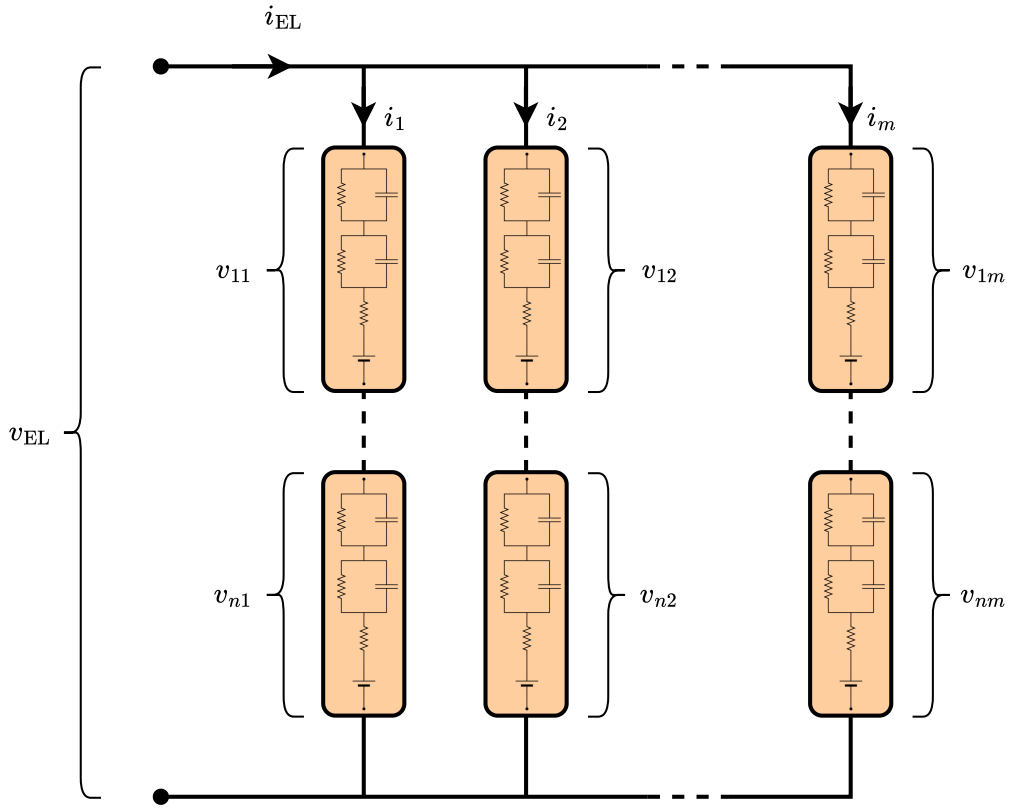

Figura B.1: Composición de un electrolizador mayor.

De acuerdo con el modelo referido, la caída de tensión en cada electrolizador menor queda dada por (B.1) en el dominio de Laplace.

$$V_{ij}(s) = \underbrace{\left(\frac{R_{\text{ano,Uni}}}{R_{\text{ano,Uni}}\, C_{\text{ano,Uni}} s + 1} + \frac{R_{\text{cat,Uni}}}{R_{\text{cat,Uni}}\, C_{\text{cat,Uni}} s + 1} + R_{\text{ohm,Uni}} \right)}_{Z(s)} I_j(s) + E_{\text{rev,Uni}}(s) \tag{B.1}$$

La corriente es idéntica en todos los electrolizadores que constituyen una rama y todos ellos son idénticos entre sí, por lo que la caída de tensión en terminales del electrolizador mayor queda descrita por (B.2).

$$\begin{aligned} V_{\text{EL}}(s) &= \sum_{i=1}^{n} \Big[Z(s)\, I_j(s) + E_{\text{rev,Uni}}(s) \Big] \\ &= n\, Z(s)\, I_j(s) + n\, E_{\text{rev,Uni}}(s) \end{aligned} \tag{B.2}$$

Por otra parte, la suma de las corrientes por cada rama constituye la corriente en terminales del electrolizador mayor, tal como se muestra en (B.3).

$$I_{\text{EL}}(s) = \sum_{j=1}^{m} I_j(s) \tag{B.3}$$

Convenientemente, considerando que todos los electrolizadores menores son idénticos, se realiza la suma vertical de las tensiones de cada rama.

$$\begin{array}{rcl} V_{\text{EL}}(s) &=& n\, Z(s)\, I_1(s) + n\, E_{\text{rev,Uni}}(s) \\ V_{\text{EL}}(s) &=& n\, Z(s)\, I_2(s) + n\, E_{\text{rev,Uni}}(s) \\ \vdots &=& \vdots \\ + \quad V_{\text{EL}}(s) &=& n\, Z(s)\, I_m(s) + n\, E_{\text{rev,Uni}}(s) \\ \hline m\, V_{\text{EL}}(s) &=& n\, Z(s) \displaystyle\sum_{j=1}^{m} I_j(s) + n\, m\, E_{\text{rev,Uni}}(s) \end{array}$$

Si se divide la suma vertical resultante por el número de ramas en paralelo, y se reemplaza lo establecido por (B.3), se obtiene la tensión del electrolizador mayor en función de su corriente a la entrada.

$$V_{\text{EL}}(s) = \underbrace{\frac{n}{m} Z(s)}_{Z(s)^*} I_{\text{EL}}(s) + \underbrace{n\, E_{\text{rev,Uni}}(s)}_{E_{\text{rev,EL}}(s)} \tag{B.4}$$

Definiendo la función de transferencia presentada en (B.5), es posible caracterizar el comportamiento del electrolizador mayor a partir de sus símiles menores que lo componen.

$$Z(s)^* = \frac{R_{\text{ano,EL}}}{R_{\text{ano,EL}} C_{\text{ano,EL}} s + 1} + \frac{R_{\text{cat,EL}}}{R_{\text{cat,EL}} C_{\text{cat,EL}} s + 1} + R_{\text{ohm,EL}} \tag{B.5}$$

Los parámetros del electrolizador mayor calculados a partir de los parámetros de sus símiles menores son presentados en (B.6).

$$\begin{aligned} R_{\text{ano,EL}} &= \frac{n}{m} R_{\text{ano,Uni}} \qquad C_{\text{ano,EL}} = \frac{m}{n} C_{\text{ano,Uni}} \qquad R_{\text{ohm,EL}} = \frac{n}{m} R_{\text{ohm,Uni}} \\ R_{\text{cat,EL}} &= \frac{n}{m} R_{\text{cat,Uni}} \qquad C_{\text{cat,EL}} = \frac{m}{n} C_{\text{cat,Uni}} \qquad E_{\text{rev,EL}} = n\, E_{\text{rev,Uni}} \end{aligned} \tag{B.6}$$

En relación con la producción de hidrógeno, el flujo molar de cada rama será n veces el flujo molar de cada electrolizador menor compuesto de n_{Uni} celdas electrolíticas. Luego, la producción neta de hidrógeno será la suma de la producción independiente de cada rama, pudiendo probarse que esta se corresponde con lo presentado en (B.7).

$$\begin{aligned} \dot{n}_{\text{H}_2\text{,out,EL}} &= \sum_{j=1}^{m} \frac{n\, n_{\text{Uni}}\, i_j(t)}{2F} \\ &= \frac{n\, n_{\text{Uni}}}{2F} \sum_{j=1}^{m} i_j(t) \\ &= \frac{n\, n_{\text{Uni}}}{2F}\, i_{\text{EL}} \end{aligned} \tag{B.7}$$

Celda de combustible

En la Figura B.2 se presenta la disposición eléctrica de un conjunto de celdas de combustible idénticas que constituyen una celda de combustible mayor. La disposición considera m ramas conectadas en paralelo, donde cada una se compone de n celdas menores conectadas en serie. El modelo definido para cada celda es el que se presenta en [13].

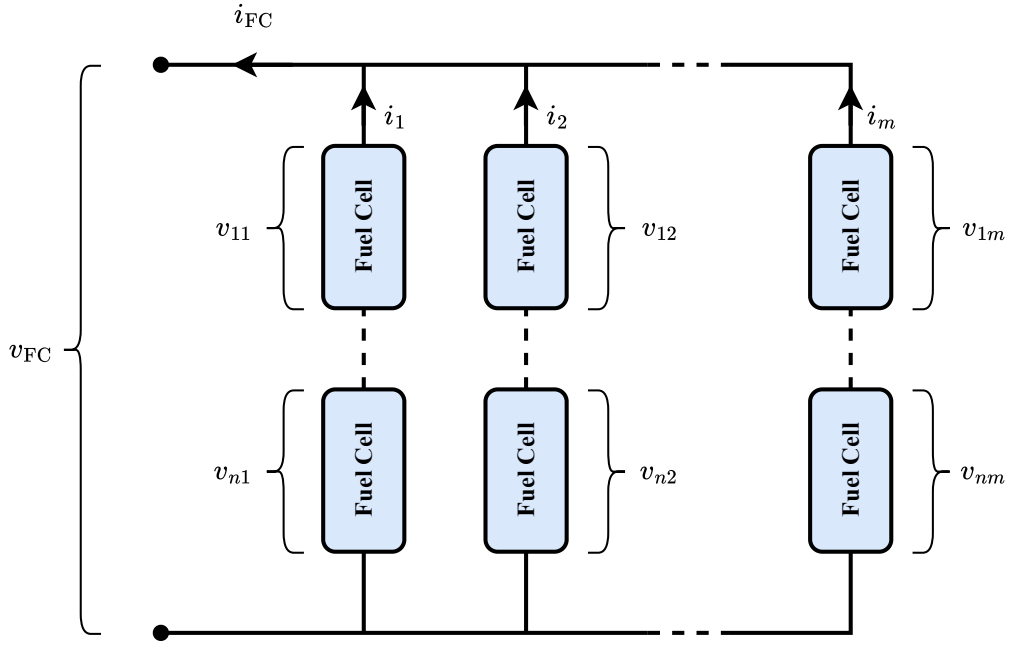

Figura B.2: Celda de combustible equivalente.

Se tiene para una única celda de la rama j

$$v_{ij} = E_{oc,\mathrm{Uni}} - \underbrace{\left(\frac{1}{s\,T_d/3+1}\right)}_{H(s)} \mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_j}{I_{o,\mathrm{Uni}}}\right) - R_{\Omega,\mathrm{Uni}}\, i_j \tag{B.8}$$

La corriente es idéntica en todas las celdas que componen una rama, además de que todas las celdas de una rama son idénticas entre sí, por lo que la caída de tensión en terminales de la celda de combustible mayor queda descrita por (B.9).

$$\begin{aligned} v_{\mathrm{FC}} &= \sum_{i=1}^{n} E_{oc,\mathrm{Uni}} - H(s)\,\mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_j}{I_{o,\mathrm{Uni}}}\right) - R_{\Omega,\mathrm{Uni}}\, i_j \\ &= n\,E_{oc,\mathrm{Uni}} - n\,H(s)\,\mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_j}{I_{o,\mathrm{Uni}}}\right) - n\,R_{\Omega,\mathrm{Uni}}\, i_j \end{aligned} \tag{B.9}$$

Por otra parte, la suma de las corrientes por cada rama constituye la corriente en terminales de la celda de combustible mayor, tal como se muestra en (B.10).

$$i_{\mathrm{FC}} = \sum_{j=1}^{m} i_j \tag{B.10}$$

Convenientemente, considerando que todas las celdas son idénticas, se realiza la suma vertical de las tensiones de cada rama.

$$\begin{array}{rcl} v_{\mathrm{FC}} & = & n\,E_{oc,\mathrm{Uni}} - n\,H(s)\,\mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_1}{I_{o,\mathrm{Uni}}}\right) - n\,R_{\Omega,\mathrm{Uni}}\, i_1 \\ v_{\mathrm{FC}} & = & n\,E_{oc,\mathrm{Uni}} - n\,H(s)\,\mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_2}{I_{o,\mathrm{Uni}}}\right) - n\,R_{\Omega,\mathrm{Uni}}\, i_2 \\ \vdots & = & \vdots \\ + \quad v_{\mathrm{FC}} & = & n\,E_{oc,\mathrm{Uni}} - n\,H(s)\,\mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_m}{I_{o,\mathrm{Uni}}}\right) - n\,R_{\Omega,\mathrm{Uni}}\, i_m \\ \hline m\,v_{\mathrm{FC}} & = & n\,m\,E_{oc,\mathrm{Uni}} - n\,H(s)\,\mathrm{NA}_{\mathrm{Uni}} \sum_{j=1}^{m} \ln\left(\frac{i_j}{I_{o,\mathrm{Uni}}}\right) - n\,R_{\Omega,\mathrm{Uni}} \sum_{j=1}^{m} i_j \end{array}$$

El supuesto de que todas las ramas son idénticas, sumado a que comparten una tensión común, posibilitan establecer que la corriente por cada una de ellas es igual, y más aún, satisfacen (B.11).

$$i_1 = i_2 = \cdots = i_m = \frac{i_{\mathrm{FC}}}{m} \tag{B.11}$$

Considerando lo mencionado anteriormente

$$\begin{aligned} \sum_{j=1}^{m} \ln\left(\frac{i_j}{I_{o,\mathrm{Uni}}}\right) &= \ln\left(\frac{\prod_{j=1}^{m} \frac{i_{\mathrm{FC}}}{m}}{I_{o,\mathrm{Uni}}{}^{m}}\right) \\ &= m\,\ln\left(\frac{i_{\mathrm{FC}}}{m\,I_{o,\mathrm{Uni}}}\right) \end{aligned} \tag{B.12}$$

De esta manera, sustituyendo la expresión anterior sumado a lo indicado por (B.10), la suma vertical da como resultado lo presentado en (B.13).

$$m\,v_{\mathrm{FC}} = n\,m\,E_{oc,\mathrm{Uni}} - n\,m\,H(s)\,\mathrm{NA}_{\mathrm{Uni}}\, \ln\left(\frac{i_{\mathrm{FC}}}{m\,I_{o,\mathrm{Uni}}}\right) - n\,R_{\Omega,\mathrm{Uni}}\, i_{\mathrm{FC}} \tag{B.13}$$

Dividiendo por la cantidad de ramas m, es posible obtener la expresión que caracteriza el comportamiento de la celda de combustible mayor a partir de las celdas menores que la componen, tal como se presenta en (B.14).

$$v_{\mathrm{FC}} = n\,E_{oc,\mathrm{Uni}} - n\,H(s)\,\mathrm{NA}_{\mathrm{Uni}}\,\ln\left(\frac{i_{\mathrm{FC}}}{m\,I_{o,\mathrm{Uni}}}\right) - \frac{n}{m}\,R_{\Omega,\mathrm{Uni}}\,i_{\mathrm{FC}} \tag{B.14}$$

En la Ecuación (B.15) se definen los parámetros de la celda de combustible mayor a partir de los parámetros de las celdas menores que la constituyen. El término asociado a la constante de tiempo no se ve alterado.

$$E_{oc,\mathrm{FC}} = n\,E_{oc,\mathrm{Uni}} \qquad I_{o,\mathrm{FC}} = m\,I_{o,\mathrm{Uni}} \qquad \mathrm{NA}_{\mathrm{FC}} = n\,\mathrm{NA}_{\mathrm{Uni}} \qquad R_{\Omega,\mathrm{FC}} = \frac{n}{m}\,R_{\Omega,\mathrm{Uni}} \tag{B.15}$$

En relación con el consumo de hidrógeno, el flujo molar de cada rama será n veces el flujo molar de cada celda de combustible menor, teniendo en consideración lo indicado por (2.8). Es importante destacar que en este caso N_{FC} corresponde al número de celdas en una celda de combustible menor. Luego, el consumo neto de hidrógeno será la suma del consumo independiente de cada rama, pudiendo probarse que este se corresponde con lo presentado en (B.16).

$$\begin{aligned}
\dot{n}_{\mathrm{H_2,in,FC}} &= \sum_{j=1}^{m} \frac{n\,N_{\mathrm{FC}}\,i_j}{2F} \\
&= \frac{n\,N_{\mathrm{FC}}}{2F} \sum_{j=1}^{m} i_j \\
&= \frac{n\,N_{\mathrm{FC}}}{2F}\,i_{\mathrm{FC}}
\end{aligned} \tag{B.16}$$