UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

Departamento de Arquitectura de Computadores y Automática

TESIS DOCTORAL

Técnicas de diseño de redes inalámbricas de sensores corporales inteligentes y energéticamente eficientes

Design techniques for smart and energy-efficient wireless body sensor networks

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Francisco Javier Rincón Vallejos

Directores

David Atienza Alonso
Marcos Sánchez-Élez Martín
Nadia Khaled

Madrid, 2013

Técnicas de diseño de redes inalámbricas de sensores corporales inteligentes y energéticamente eficientes

Design techniques for smart and energy-efficient Wireless Body Sensor Networks

Tesis Doctoral

Francisco Javier Rincón Vallejos

Departamento de Arquitectura de Computadores y Automática
Facultad de Informática
Universidad Complutense de Madrid

2012

Técnicas de diseño de redes inalámbricas de sensores corporales inteligentes y energéticamente eficientes

Design techniques for smart and energy-efficient Wireless Body Sensor Networks

Tesis presentada por

Francisco Javier Rincón Vallejos

Departamento de Arquitectura de Computadores y Automática
Facultad de Informática
Universidad Complutense de Madrid

2012

Técnicas de diseño de redes inalámbricas de sensores corporales inteligentes y energéticamente eficientes

Memoria presentada por Francisco Javier Rincón Vallejos para optar al grado de Doctor por la Universidad Complutense de Madrid, realizada bajo la dirección de D^a. Nadia Khaled (Nestlé Research Center, Lausanne, Suiza), D. David Atienza Alonso (EPFL, Suiza, y Departamento de Arquitectura de Computadores y Automática, Universidad Complutense de Madrid) y D. Marcos Sánchez-Élez Martín (Departamento de Arquitectura de Computadores y Automática, Universidad Complutense de Madrid).

Design techniques for smart and energy-efficient Wireless Body Sensor Networks

Dissertation presented by Francisco Javier Rincón Vallejos to the Complutense University of Madrid in order to apply for the Doctor's degree. This work has been supervised by Ms. Nadia Khaled (Nestlé Research Center, Lausanne, Suiza), Mr. David Atienza Alonso (EPFL, Switzerland, and Department of Computer Architecture and Automation, Complutense University of Madrid) and Mr. Marcos Sánchez-Élez Martín (Department of Computer Architecture and Automation, Complutense University of Madrid).

Madrid, 2012

Este trabajo ha sido posible gracias a la Comisión Interministerial de Ciencia y Tecnologia, por las ayudas recibidas a través de los proyectos CICYT TIN2005/5619 y CICYT TIN2008/00508

A Francisco, Sacri y Emi
. . . y, en especial, a Laura

Agradecimientos

El agradecimiento es la memoria del corazón.
Lao-Tsé

Primero me gustaría dar las gracias a los principales culpables de que ahora este escribiendo esta tesis doctoral, los que me motivaron para elegir este camino, que a veces me ha dado muchos dolores de cabeza pero también me ha proporcionado grandísimas satisfacciones. El primer responsable es Román Hermida, que durante una de sus clases de AIC nos informó sobre la existencia de unas becas de colaboración para ayudar en labores de investigación en el DACYA. Solicité una de ellas y empecé a trabajar con Marcos Sánchez-Élez, a quien tengo que agradecer el interés que me transmitió por la investigación.

También tengo que dar las gracias a David Atienza, que me propuso trabajar en un tema nuevo, cuando todavía no sabíamos muy bien qué hacer ni cuál sería el resultado, que al final me ha apasionado y ha dado lugar a esta tesis. Gracias también por invitarme a Suiza, primero al laboratorio del profesor Giovanni de Micheli, a quien tengo que agradecer su atención durante las estancias que he realizado, y luego a su propio laboratorio.

He de dar las gracias también a Nadia Khaled, que me ha dedicado toneladas de su tiempo y me ha inculcado la importancia de estudiar hasta el mas mínimo detalle, así como el valor de la pulcritud y la precisión para dar lugar a un trabajo bien hecho.

Tengo que expresar mi más sincero agradecimiento a Joaquín Recas, compañero de fatigas durante muchos periodos de esta tesis, que es también en parte suya.

No puedo olvidarme de todos mis compañeros y ex-compañeros de la UCM y EPFL. Mención especial merece Alberto, grandísimo amigo desde hace ya 11 años, siempre dispuesto para lo que haga falta, un 10 como persona. Gracias también a Pablo, Jose, Carlos, Juanan, Lucas, Íñigo, Antonio, Miguel, Poletti, Guillermo, David Cuesta, Mónica, Víctor, Rubén, Hossein, Ahmed, Ivan, Vincenzo, Mohamed, Alessandro, Shivani, Arvind, Karim, Federico, Srini, Antonio Pullini, Shashi, Ciprian, Haykel, Abhishek,

Frank, Diego y Nicolas Boichat, con quienes he compartido muchos momentos a lo largo de este periodo que permanecerán imborrables, y perdón si me olvido a alguien.

Quiero dedicar esta tesis muy especialmente a mis padres, que nunca han dudado en sacrificarse para poner a mi alcance todas las herramientas necesarias para llegar hasta aquí. Gracias por apoyarme en todas las decisiones que he tomado, aunque a veces eso implique distanciarnos un poco más. Me harían falta 10 vidas para devolverles todo lo que me han dado. También quiero agradecérselo a mi hermano, que siempre está ahí, ya sea para leerse un paper o para escucharme un rato mientras tomamos una cerveza.

Por último quiero dar las gracias y dedicar esta tesis a Laura, que es quien más la ha sufrido. Gracias por compartir conmigo las satisfacciones que me ha dado este trabajo y, sobre todo, por apoyarme cuando las cosas no salían tan bien. Gracias de corazón por estar siempre a mi lado.

What we do for ourselves dies with us.
What we do for others and the world remains and is immortal.

Albert Pike

Index

List of Figures

List of Tables

List of Algorithms

Resumen

En cumplimiento del Artículo 4 de la normativa de la Universidad Complutense de Madrid que regula los estudios universitarios oficiales de postgrado, se presenta a continuación un resumen en español de la presente tesis, que incluye la introducción, objetivos, principales aportaciones y conclusiones del trabajo realizado.

0.1. Introducción

Las redes inalámbricas de sensores [CES04] (en inglés: "wireless sensor networks" o WSN) son redes inalámbricas constituidas por una serie de sensores autónomos distribuidos que cooperan para medir parámetros físicos. Sus utilidades van desde la observación ambiental a la monitorización médica. La WSN captura e interpreta informacion relevante, que posteriormente es almacenada o enviarla a una estación base. Para llevar a cabo estas tareas cada nodo está equipado con varios sensores, un chip de radio y un microcontrolador. Como fuente de energía utilizan una batería, aunque también pueden usarse dispositivos de recuperación de energía (como paneles solares o generadores termoeléctricos), para así aumentar el tiempo de vida del nodo y minimizar su mantenimiento [PS05].

La Figura 1 muestra la arquitectura típica de una WSN. La red se compone de varios nodos (*sensor nodes*) que miden una serie de parámetros y envían la información recogida a un nodo central (*mesh node*). Cada nodo puede alcanzar uno o más nodos centrales, ya sea directamente o a través de otros nodos de la red, siguiendo un esquema multisalto. Los nodos centrales pueden estar conectados entre sí dando lugar a una red de más alto nivel a la que puede accederse por medio de estaciones de control (*management station*), ya sea usando una conexión inalámbrica punto a punto con uno de los nodos centrales o a través de Internet.

Las WSNs pueden estar equipadas con cualquier tipo de sensor, dando lugar a multitud de aplicaciones. Se usan tanto en el ámbito comercial como industrial, para medir parámetros que son demasiado difíciles o caros de monitorizar usando sensores con cables. Estas redes pueden desplegarse en

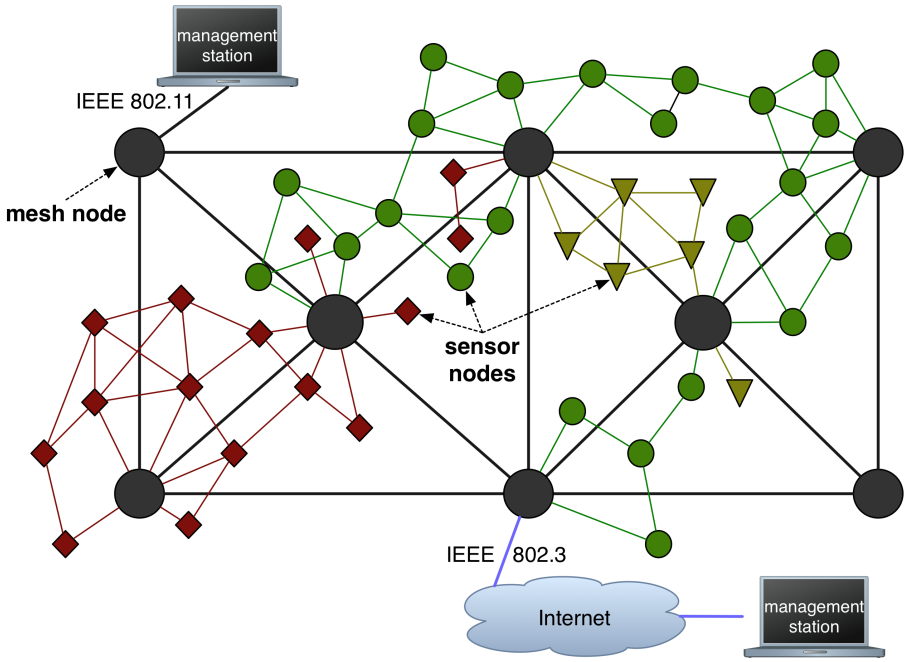

Figura 1: Arquitectura típica de una WSN

áreas selváticas, donde pueden operar durante años sin necesidad de recargar o cambiar las baterías de los nodos. Entre sus aplicaciones encontramos:

- Investigación biológica: seguimiento de especies salvajes en su hábitat natural.
- Monitorización ambiental: polución (aire, agua, tierra), detección de incendios forestales, predicción de terremotos y otros desastres naturales.
- Respuesta frente a accidentes: dar soporte a los equipos de emergencia para identificar riesgos o peligros, localizar supervientes.
- Agricultura: seguimiento de cultivos, gestión de microclimas para mejorar la producción de vinos, control de fertilización.
- Sector industrial: control y seguimiento del proceso de fabricación, predicción de fallos mecánicos, control de calidad.
- Sector comercial: control de stock, seguimiento de productos, control de calidad.
- Arquitectura: espacios inteligentes, domótica, detección de intrusos.
- Transporte: control de los sistemas internos de coches, barcos y aviones.

- Autoridades locales: información sobre el tráfico, análisis y coordinación.
- Militar: vigilancia, identificación de enemigos, seguimiento de objetivos, soporte para operaciones logísticas.
- Medicina: seguimiento de personal médico y pacientes, monitorización de señales biométricas, control de administración de medicamentos.

Este trabajo se centra en el estudio de un subconjunto de WSNs para aplicaciones biomédicas, llamadas redes inalámbricas de sensores corporales (en inglés: "wireless body sensor networks" o WBSNs), que tienen un enorme potencial para transformar la manera en la que interaccionamos con las tecnologías de la información [HJB+09], para así obtener un mayor beneficio de su uso. Como caso práctico, y sin pérdida de generalidad, este trabajo se centra en el estudio del electrocardiograma (ECG). Las siguientes secciones muestran más en detalle las ventajas de las WBSNs y su aplicación para monitorización de ECG.

0.1.1. Redes Inalámbricas de Sensores Corporales

El envejecimiento de la población, así como su estilo de vida cada vez más sedentario, están incrementando la aparición de enfermedades crónicas como enfermedades cardiovasculares, hipertensión y diabetes. Según la Organización Mundial de la Salud, las enfermedades cardiovasculares causan el 30 % de las muertes a nivel mundial (alrededor de 17,5 millones de personas en 2005). La diabetes afecta actualmente a 180 millones de personas en todo el mundo, pudiendo llegar a 360 millones en 2030 según los expertos. De hecho, más de 2300 millones de personas tendrán sobrepeso en 2015. Además, se está produciendo un rápido aumento en el número de afectados por enfermedades neurodegenerativas como el Parkinson o el Alzheimer [PW10]. Resumiendo, hoy en día, las enfermedades no contagiosas son responsables del 63 % de las muertes a nivel mundial y se prevé que el coste causado por las mismas puede llegar al 75 % del producto interior bruto en el año 2030.

Las crecientes necesidades de atención médica están ejerciendo una enorme presión sobre el sistema sanitario. Por otra parte, la escasez de personal cualificado, así como el endurecimiento de los presupuestos han agravado la inminente crisis sanitaria. Estas tendencias económicas, sociales, demográficas ponen de relieve la necesidad de aprovechar los avances tecnológicos para lograr soluciones viables y eficaces que mejoren la calidad de vida de los pacientes. Por otra parte, como se muestra en la Figura 2, el 50 % de los determinantes de la salud están relacionados con el estilo de vida. Estos cambios exigen un cambio de paradigma en la prestación de asistencia sanitaria, donde las tecnologías de la información y los sistemas de monitorización de pacientes son claves para la llegada de la medicina preventiva personalizada.

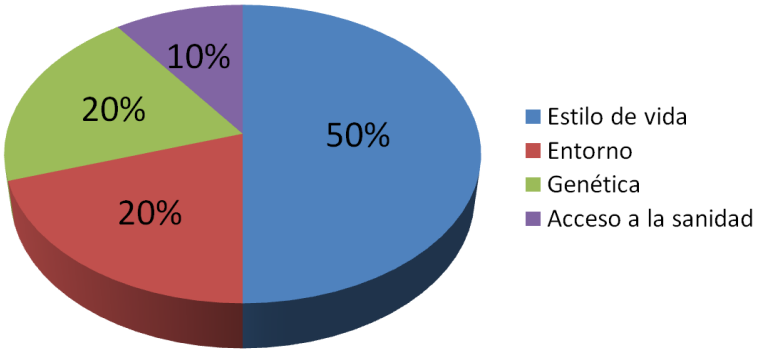

Figura 2: Determinantes de la salud (fuente: Institute for the future, Center for disease control and prevention, 2006)

Estos sistemas son capaces de recoger datos de los pacientes y mejorar la eficacia de las terapias. También pueden apoyar la toma de decisiones clínicas y proporcionar una vía de comunicación cómoda y barata entre los pacientes y el personal sanitario.

Las WBSNs para monitorización, diagnóstico y detección de emergencias, están ganando popularidad y están llamadas a cambiar profundamente la asistencia sanitaria en los próximos años. El uso de estas redes permite una supervisión continua, contribuyendo a la prevención y el diagnóstico precoz de enfermedades, al tiempo que mejora la autonomía del paciente con respecto a otros sistemas de monitorización actuales, más invasivos y de mucho mayor tamaño. Sin embargo, la adopción de las WBSNs debe superar grandes retos técnicos y sociales (forma y tamaño, duración de la batería, fiabilidad, seguridad, privacidad, compatibilidad, facilidad de uso, etc.). Aunque las WBSNs comparten algunos de estos retos con las WSNs genéricas, han surgido muchos interrogantes que requieren de nuevas líneas de investigación. Por lo tanto este tipo de redes merece un análisis por separado, ya que la naturaleza de las señales, el tipo de algoritmos de procesamiento de señal y las limitaciones inherentes de las aplicaciones sanitarias hacen a estas redes muy diferentes del resto de WSNs.

A diferencia de las WSNs genéricas, que se componen de un gran número de nodos distribuidos que realizan la misma tarea, las WBSNs cuentan con unos pocos nodos (a menudo menos de 10) que cubren el cuerpo de un ser humano, cada uno de los cuales está dedicado a una tarea específica. Por ejemplo, un nodo que mide aceleración en el tobillo para el análisis de la marcha no puede medir también el electroencefalograma (EEG), ya que tanto la ubicación como el hardware de muestreo de las señales son totalmente diferentes.

Desde el punto de vista de la adquisición de datos, las señales comúnmente utilizadas en WBSNs, tales como EEG, ECG o aceleración, tienen

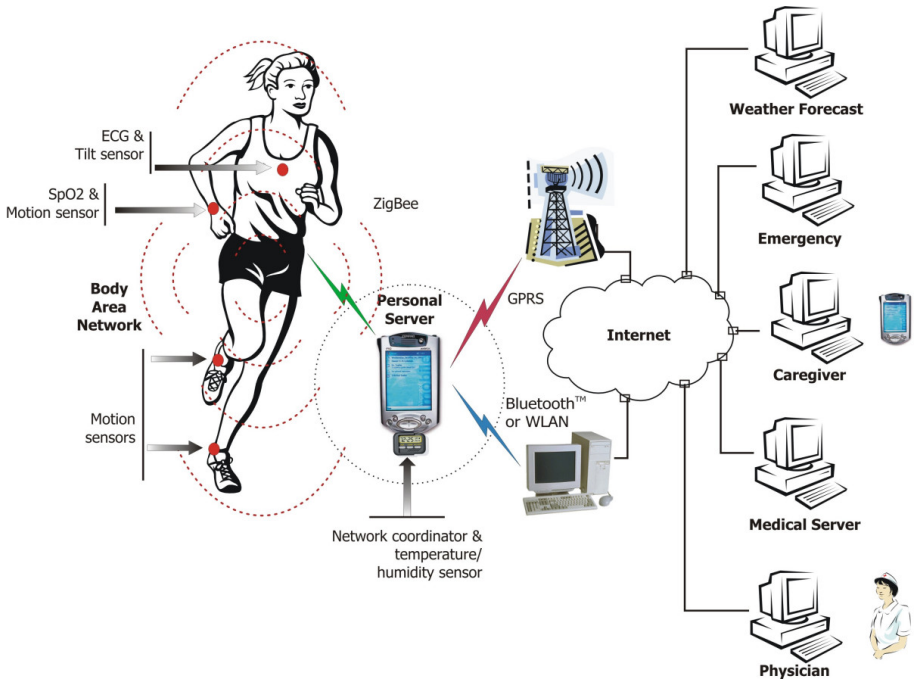

Figura 3: Arquitectura de una WBSN

componentes en frecuencias relativamente altas, lo que requiere el uso de altas velocidades de muestreo (desde 100 Hz a 1 kHz). Esto conduce a la generación de una enorme cantidad de información en cada nodo en comparación con las WSNs genéricas, que necesita ser almacenada, procesada o transmitida a una estación base o coordinador.

Para lograr una adopción generalizada, los nodos de una WBSN deben ser mínimamente invasivos, lo que significa que deben tener un reducido tamaño que los haga cómodos de usar y por tanto no alteren la actividad normal del paciente. Esto implica el uso de baterías más pequeñas, que obligan a buscar un compromiso entre consumo de energía y la fidelidad, rendimiento y latencia de las aplicaciones. Esta tesis estudia estos compromisos con el fin de mejorar la eficiencia energética de las WBSNs, prolongando por tanto su vida útil. Al mismo tiempo se busca explotar la limitada capacidad de procesamiento y memoria de los nodos, para lograr WBSNs inteligentes que sean capaces de proporcionar información en tiempo real sobre el estado de salud del paciente.

En la literatura podemos encontrar una gran variedad de aplicaciones de WBSNs, que van del estudio del sueño a la detección de estados de ánimo o dolor. Por ejemplo, Grundlehner y otros [GBPG09] presentan una WBSN para monitorizar en tiempo real el nivel de activación mental, basándose en un análisis del ECG, la respiración, la respuesta galvánica de la piel y la temperatura de la piel. Estas constantes vitales son recogidas por una

WBSN y transmitidas a un ordenador central que procesa la información y proporciona una evaluación en tiempo real del nivel de activación mental. Abbate y otros [AAC+11] presentan un sistema para detección de caídas basado en datos de aceleración con el que obtienen un 100 % de sensibilidad y especificidad.

En esta tesis se considera como caso práctico una WBSN para monitorización cardiaca, que lleva a cabo un análisis automático del ECG. Tradicionalmente, este análisis se realiza en el hospital usando voluminosos sistemas de monitorización instalados en las habitaciones, que procesan el ECG en tiempo real, o por medio de un sistema Holter, que almacena el ECG del paciente durante un largo periodo de tiempo, tras el cual los datos son descargados e interpretados en el hospital usando un software especial. Aunque más pequeños que los tradicionales monitores instalados en los hospitales, los Holters siguen teniendo un tamaño que los hace incómodos de utilizar. Recientemente, se ha dedicado un importante esfuerzo industrial y académico a crear sistemas de monitorización de ECG de pequeño tamaño, portátiles e inalámbricos, que están llamados a ser el motor de la próxima generación de sistemas móviles de monitorización cardiaca. Sin embargo, estos sistemas se limitan principalmente a la transmisión del ECG en bruto, lo que implica un gran consumo de energía que reduce su autonomía a muy pocos días. El objetivo de este trabajo es diseñar un monitor Holter inalámbrico de larga duración y reducido tamaño con capacidad de diagnóstico en tiempo real.

0.1.2. Monitorización cardiaca

La elección del ECG está principalmente motivada por el gran conocimiento de esta señal en la comunidad médica. Por otra parte, como ya he mencionado anteriormente, las enfermedades cardiovasculares son la principal causa de mortalidad en el mundo, responsables del 30 % de las muertes a nivel mundial en 2005 y de un gasto económico de miles de millones de dólares [PW10]. Se prevé que estas cifras vayan en aumento debido al rápido envejecimiento de la población mundial y a su estilo de vida poco saludable. Las WBSNs están preparadas para ofrecer soluciones a gran escala y bajo coste, para disminuir el alto gasto sanitario que, de lo contrario, puede volverse insostenible en los próximos años. Esta tesis propone un sistema automático y multiderivación de diagnóstico de ECG en tiempo real. Para maximizar la eficiencia energética de este Holter inalámbrico inteligente, esta tesis propone reducir el envío de información por radio incluyendo análisis automatizado de la señal ECG en el nodo. Los resultados obtenidos en esta tesis para el caso práctico considerado (monitorización de ECG) se puede extrapolar a otras aplicaciones que utilizan diferentes señales biométricas, ya que todas ellas tienen requisitos muy similares.

El ECG describe la actividad eléctrica del corazón, registrada por electrodos colocados sobre la piel [SL05]. Las variaciones de voltaje medidas por los

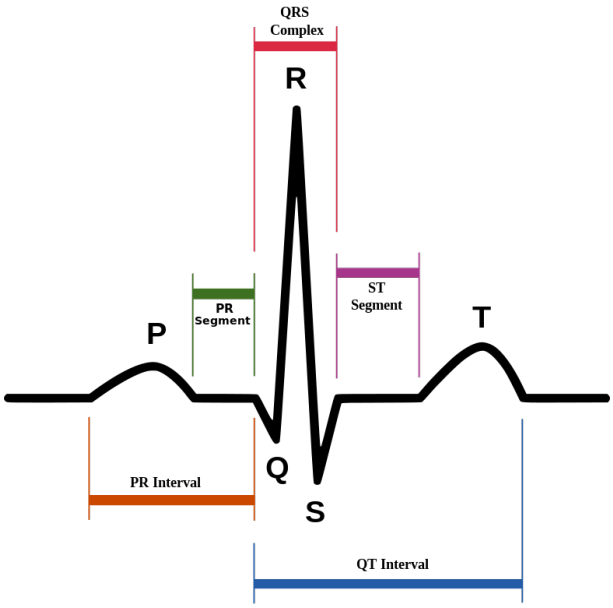

Figura 4: Fragmento de ECG que muestra las ondas que componen un latido cardiaco: complejo QRS y ondas P y T

electrodos son causadas por los potenciales de acción de las células cardiacas, causantes de la contracción del corazón. La morfología de la onda resultante contiene información que puede usarse para diagnosticar enfermedades reflejadas por alteraciones de la actividad eléctrica del corazón. El patrón de tiempo que caracteriza la ocurrencia de latidos sucesivos es también muy importante. El ECG es una señal más o menos periódica, y cada latido se compone de un complejo QRS, precedido por una onda P, y seguido por una onda T (véase la Figura 4), que se corresponden con la secuencia de eventos que definen un ciclo cardiaco.

0.1.3. Objetivos de esta tesis

El primer objetivo de esta tesis es desarrollar un sistema de monitorización de ECG, que no sólo muestre continuamente el ECG del paciente, sino que además lo analice en tiempo real y sea capaz de dar información sobre el estado del corazón a través de un dispositivo móvil. Esta información también puede ser enviada al personal médico en tiempo real. Si ocurre un evento peligroso, el sistema lo detectará automáticamente e informará de inmediato al paciente y al personal médico, posibilitando una rápida reacción en caso de emergencia. Para conseguir la implementación de dicho sistema, se desarrollan y optimizan distintos algoritmos de procesamiento de ECG en tiempo real, que incluyen filtrado, detección de puntos característicos y clasificación de arritmias.

El segundo reto de esta tesis consiste en la mejora de la eficiencia energética de la red de sensores, cumpliendo con los requisitos fidelidad y rendimiento de la aplicación. Para ello se proponen técnicas de diseño para reducir el consumo de energía, que permitan buscar un compromiso óptimo entre el tamaño de la batería y su tiempo de vida. Si el consumo de energía puede reducirse lo suficiente, sería posible desarrollar una red que funcione permanentemente. Por lo tanto, el muestreo, procesamiento, almacenamiento y transmisión inalámbrica tienen que hacerse de manera que se suministren todos los datos relevantes, pero con el menor consumo posible de energía, minimizando así el tamaño de la batería (que condiciona el tamaño total del nodo) y la frecuencia de recarga de la batería (otro factor clave para su usabilidad). Por lo tanto, para lograr una mejora en la eficiencia energética del sistema de monitorización y análisis de ECG propuesto en esta tesis, se estudian varias soluciones a nivel de control de acceso al medio (MAC) y sistema operativo (OS).

0.2. Evaluación de arquitecturas

0.2.1. Arquitecturas para WBSNs: estado del arte

El presente trabajo considera la arquitectura típica de una WBSN donde un microcontrolador ejecuta una aplicación software de procesamiento de datos, y los demás servicios se delegan en un OS. La Figura 5 ofrece una visión estructural de este tipo de nodos. Asumimos que un nodo transmite sus datos a una estación base siguiendo la topología de estrella normalmente usada en WBSNs. Cada nodo de la red genera una cantidad de tráfico constante, evitando así ráfagas de datos que pueden interferir con la transmisión periódica de los otros nodos de la red.

El *sensor* es el bloque hardware que muestrea la señal a una frecuencia que depende de las componentes en frecuencia de la señal y de otras limitaciones como el teorema de Nyquist-Shannon. A continuación, un conversor A/D digitaliza las muestras utilizando un número de bits que depende de su resolución.

Las aplicaciones (*applications*) comprenden todos los programas software utilizados para procesar los datos detectados, incluyendo filtrado, extracción de características, compresión y agregación.

El sistema operativo (*operating system*, OS) ofrece servicios tales como la interacción con el hardware o el software que gestiona el muestreo de la señal, la memoria y la transmisión por radio. El OS también gestiona un conjunto de colas, incluyendo las de la comunicación entre procesos y otra que contiene los paquetes listos para ser transmitidos. Además, el sistema operativo implementa un protocolo MAC para controlar el acceso al medio inalámbrico compartido entre los nodos de la red.

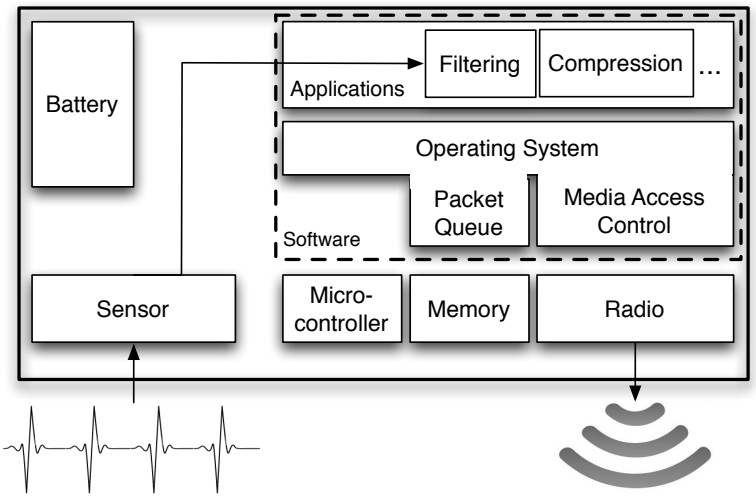

Figura 5: Diagrama de bloques de la arquitectura de referencia

El microcontrolador (*microcontroller*, μC) es el bloque que se encarga de ejecutar el OS y las aplicaciones software. Dependiendo del hardware, técnicas tales como escalamiento dinámico de voltaje [GC97] podrían estar disponibles para permitir que el microcontrolador permanezca activo durante un tiempo limitado (*ciclo de trabajo*) y para cambiar a un estado de bajo consumo cuando no hay tareas en espera de ser ejecutadas.

El banco de memoria (*memory*) almacena los datos de las aplicaciones y del sistema operativo. Aunque una memoria más grande exhibe un mayor consumo de energía [KOKE06], un tamaño demasiado limitado puede afectar a la capacidad de las colas internas, y por tanto al rendimiento de las aplicaciones y de la transmisión.

Por último, la *radio* es la encargada de modular y transmitir los datos a través del canal inalámbrico. Dependiendo de las características de la plataforma, pueden configurarse la potencia de transmisión inalámbrica y el esquema de modulación, incluso dinámicamente [SRS03], para determinar la cobertura de la comunicación con una tasa de paquetes erróneos (PER) predeterminada.

Con la proliferación de las aplicaciones para WBSN, muchas plataformas con una arquitectura que se corresponde con la representada en la Figura 5 han aparecido durante los últimos años. La primera decisión crítica que debemos tomar cuando nos enfrentamos al desarrollo de una WBSN es la selección de la plataforma más apropiada, teniendo en cuenta los requisitos de la aplicación. En este trabajo, las diferentes plataformas disponibles se van a analizar desde el punto de vista de su flexibilidad y su eficiencia. En este sentido, podemos clasificarlas en dos grupos principales: *application-specific integrated circuits (ASICs)* y plataformas *commercial off-the-shelf (COTS)*.

El uso de ASICs es siempre más eficaz en cuanto a consumo de energía y tamaño, ya que están especialmente diseñados para una aplicación muy específica. Sin embargo, el diseño de estos dispositivos requiere un gran esfuerzo mientras que ofrecen una flexibilidad muy limitada. A pesar de que las plataformas ASIC logran una gran eficiencia energética, su uso es muy limitado debido a su naturaleza "rígida", no reprogramable, por lo que solo pueden realizar una secuencia predefinida de operaciones. Los COTS son en general menos eficientes energéticamente y tienen un tamaño más grande, pero por otro lado proporcionan una mayor flexibilidad, permitiendo la ejecución de tareas más heterogéneas, debido a que normalmente usan procesadores de propósito general (GPPs). A menudo, los COTS son una buena solución ya que pueden ofrecer un tamaño, rendimiento y tiempo de vida razonables para aplicaciones WBSN, si bien tienen un coste de producción muy bajo.

Una elección equilibrada entre flexibilidad y eficiencia energética para el procesamiento de señales biométricas en tiempo real consiste en el uso de un *application-specific instruction-set processor (ASIP)*. Los ASIPs ofrecen flexibilidad para un grupo de aplicaciones específico, mientras que consiguen una mayor eficiencia energética en comparación con los microcontroladores de propósito general de bajo consumo, ya que tienen un conjunto de instrucciones reducido y personalizado para el grupo de aplicaciones correspondiente.

0.2.2. La plataforma Shimmer™

Teniendo en cuenta las ventajas e inconvenientes de cada tipo de arquitectura disponible para WBSNs, decidí utilizar el el nodo Shimmer™ como caso práctico de arquitectura. La elección de esta plataforma vino motivada principalmente por las siguientes razones:

- Gran flexibilidad, debido al uso del microcontrolador de propósito general MSP430. Este trabajo no sólo se centra en el procesamiento de señales biométricas, sino que también analiza y propone las optimizaciones a otros niveles de la arquitectura (OS, protocolo MAC, etc.). Por esta razón, las implementaciones de ASIC y de ASIP se han descartado, buscando tener un conjunto de instrucciones lo más flexible posible.

- Los recursos computacionales y de memoria de esta plataforma, así como su fuente de energía, ofrecen un ejemplo representativo de los recursos típicos de un nodo de una WBSN para las aplicaciones consideradas en este trabajo. Además, el uso de componentes COTS reduce el coste de la plataforma en comparación con ASICs o ASIPs.

- Soporte. Aunque el MSP430 no ofrece un consumo tan reducido como otras arquitecturas, el uso de este microcontrolador está muy extendido, siendo utilizado en dominios de aplicación muy heterogéneos. Por

lo tanto, hay un gran apoyo para esta plataforma en lo que se refiere a herramientas disponibles o código portado (OSs, protocolos MAC, aplicaciones, etc.).

- Gran número de campos de aplicación, ya que hay multitud de accesorios para el Shimmer™, por lo que se puede utilizar no sólo para ECG, sino también para otras aplicaciones biomédicas (EEG, electromiograma, respuesta galvánica de la piel, etc.) y no biomédicas (temperatura, humedad, luminosidad, etc.).

- Es un buen representante de una plataforma WBSN, por lo tanto todos los resultados obtenidos en esta tesis usando el nodo Shimmer™, son igualmente aplicables a otras plataformas WBSN.

0.2.3. Arquitectura de la plataforma Shimmer™

El nodo utilizado en este trabajo es el Shimmer™ [BGM+10], una pequeña plataforma inalámbrica diseñada por Intel para aplicaciones médicas. Esta plataforma tiene el objetivo a largo plazo de facilitar la investigación en tecnologías para una vida independiente, ya que puede montarse en el brazalete de un reproductor MP3 y tiene una carcasa muy robusta con un tamaño de $54 \times 35 \times 18$ mm. El nodo Shimmer™ está equipado con el microcontrolador de 16 bits y ultra bajo consumo TI MSP430F1611 [Insc], que opera a una frecuencia máxima de 8 MHz, incluye 10 kB de RAM, 48 kB de memoria Flash y otros periféricos como un convesor A/D de 8 canales y resolucion de 12 bits, una unidad de acceso directo a memoria (DMA) y un multiplicador hardware. Este nodo tambien está equipado con dos chips de radio (uno Bluetooth y otro compatible con IEEE 802.15.4), un acelerómetro de 3 ejes y un puerto de expansión utilizado para conectar una placa equipada con sensores adicionales (ECG, giroscopios, etc.). En este trabajo usamos una placa que es capaz de muestrear y acondicionar 3 derivaciones de ECG. También incorpora una ranura MicroSD™ que permite almacenar hasta 2 GB, cuatro diodos emisores de luz (LED) para fines de visualización y una batería integrada de Li-ion de 280 mAh.

Esta plataforma ha sido diseñada para programarse con el sistema operativo TinyOS [LMP+04] y sus controladores específicos. Desafortunadamente TinyOS muestra varios inconvenientes, que serán discutidos en la Sección 0.3, que lo hacen inadecuado para las aplicaciones consideradas en este trabajo. Debido a estas limitaciones, se ha portado FreeRTOS [biba], que es una sistema operativo ligero, portable y de código abierto, que da soporte para operaciones en tiempo real. Por último, hemos utilizado la herramienta de código abierto GCC 3.2.3 para el MSP430 [bibb] y Code Composer Essentials (CCE) v.3.1 [Insb], que utilizan todos los recursos hardware del microcontrolador utilizado, para compilar las aplicaciones para el nodo.

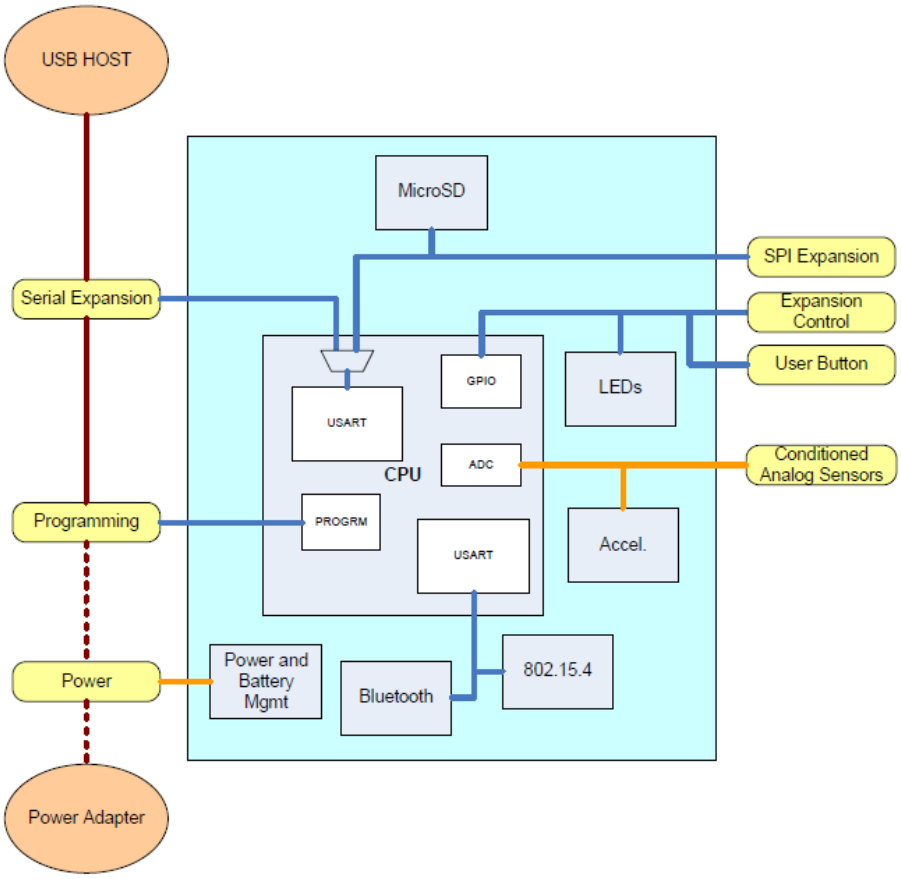

Figura 6: Esquema simplificado de las interconexiones del Shimmer™.

0.2.3.1. Estimación del consumo de energía

Para justificar aún más la elección de la plataforma Shimmer™ y motivar las técnicas de eficiencia energética propuestas en esta tesis, se presentan los resultados de consumo de energía de un nodo WBSN en dos escenarios de aplicación: ECG streaming (un nodo situado en el cuerpo muestrea 2 derivaciones de ECG y transmite los datos a una estación base) y Rpeak (aplicación que detecta cuando un latido se produce y sólo transmite la ubicación de los picos R al coordinador de la WBSN).

Estos resultados [RPR+08] se obtuvieron utilizando una plataforma desarrollada en IMEC Netherlands [PGdVT07], cuya arquitectura hardware es muy similar a la del Shimmer™. Esta plataforma es capaz de muestrear hasta 24 canales de EEG y 1 canal de ECG usando un ASIC de ultra-bajo consumo (10,5 mW a 3,0 V). El microcontrolador usado es el TI MSP430F149, mientras que la responsabilidad de la comunicación inalámbrica recae en la Nordic nRF2401, un radio de bajo consumo de 2,4 GHz (10,5 mA con una

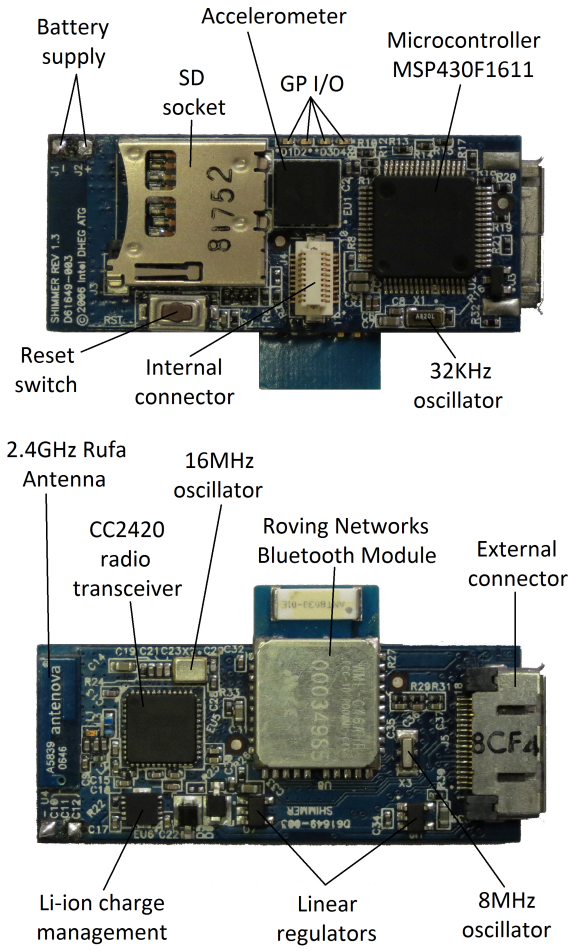

Figura 7: Foto de la placa base del Shimmer™

potencia de salida de -5 dBm y 18 mA en modo de recepción).

En cuanto al software, el nodo cuenta con un OS integrado, TinyOS, encima del cual se construye la arquitectura software completa, incluyendo el protocolo MAC. Este estudio propone el uso de un protocolo time division multiple access (TDMA), que permite a varios usuarios compartir el mismo canal de frecuencia dividiendo el tiempo en diferentes franjas. Cada nodo tiene una franja asignada, que puede utilizar para intercambiar datos con el coordinador de la WBSN. El protocolo está regulado por el coordinador, que periódicamente emite señales de sincronización (beacons) a todos los nodos.

La Figura 8 presenta el consumo de energía de los principales bloques hardware de la arquitectura para los dos escenarios considerados (ECG streaming y Rpeak) asumiendo una frecuencia de muestreo de 200 Hz. Esta figura muestra que el bloque dedicado a muestrear la señal tiene un consumo de

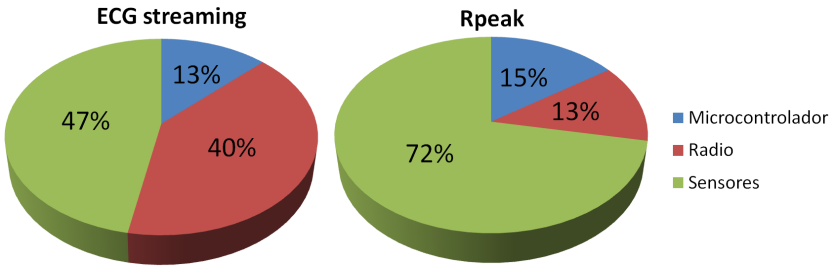

Figura 8: Consumo de energía de los bloques principales de una plataforma WBSN en dos escenarios de aplicación: ECG streaming y Rpeak

energía muy elevado, lo que se puede disminuir con el uso de la Shimmer™, ya que el consumo de energía se reduce a 6,6 mW. La radio es también un componente con gran consumo de energía, especialmente en el escenario de ECG streaming, pero su consumo se puede reducir mediante el uso de un protocolo MAC más optimizado y un sistema operativo más eficiente. Además, la Figura 8 también demuestra que el consumo de energía de la radio puede reducirse significativamente si el microcontrolador del nodo se utiliza para procesar la señal ECG y así disminuir la cantidad de datos que debe ser transmitida al coordinador de la WBSN.

Por estas razones, esta tesis propone varias técnicas de diseño para WBSNs inteligentes y energéticamente eficientes, que serán detalladas en las secciones siguientes. El objetivo final es diseñar un Holter inalámbrico inteligente de larga duración con capacidad de diagnóstico en tiempo real.

0.3. Optimizaciones de la WBSN relativas a la comunicación

Un nodo sensor está equipado con una radio, que permite al dispositivo enviar la información previamente recogida y procesada al coordinador de la WBSN o recibir órdenes de él. El protocolo de control de acceso al medio (MAC), que gestiona el acceso al canal inalámbrico, tiene que enfrentarse a varios problemas de comunicación. Los más importantes son los siguientes:

- Colisiones. Se producen cuando dos o más dispositivos acceden al canal inalámbrico para enviar información al mismo tiempo. Esto provoca la corrupción de los paquetes que se envían, que tienen que ser transmitidos de nuevo.
- Overhearing. Se produce cuando un nodo A envía datos a un nodo B, y hay un nodo C que recibe este paquete.
- Idle listening. Este problema se da cuando un nodo A no sabe cuándo otro nodo B le va a enviar un paquete. Esto causa que el nodo A

necesite mantener la radio en modo de recepción por un periodo de tiempo durante el cual no está recibiendo ninguna información.

Todos estos problemas aumentan significativamente el consumo de energía, ya sea por retransmisión de paquetes, recepción de datos que no son para el nodo o escuchar el canal cuando no se está enviando ninguna información al nodo. Un protocolo MAC para WBSNs tiene que considerar todos estos problemas y tratar de minimizar su ocurrencia para reducir el consumo de energía del nodo tanto como sea posible. Además, el protocolo MAC implica la transmisión de paquetes de control (mensajes de sincronización, acuses de recibo para indicar que un paquete ha sido recibido correctamente, etc.), cuya transmisión también debe reducirse a fin de minimizar el consumo de energía de la radio.

0.3.1. Protocolo MAC propuesto

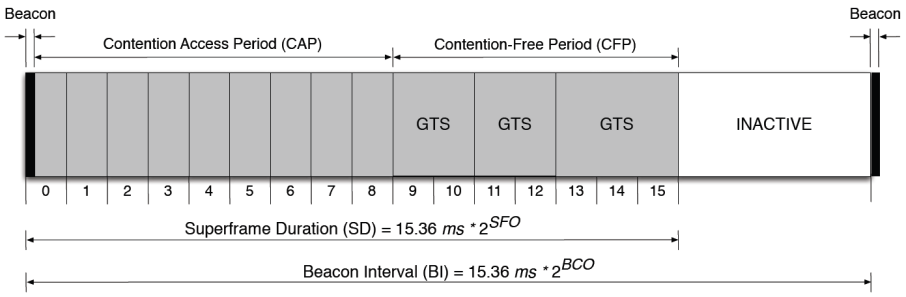

Figura 9: Estructura del superframe IEEE 802.15.4

Con el fin de aprovechar al máximo las capacidades de bajo consumo de este chip de radio, y por lo tanto reducir el consumo de energía del sistema en su conjunto, este trabajo propone un protocolo MAC especialmente optimizado para adaptarse a los requisitos de las aplicaciones para WBSN aquí consideradas. El nodo Shimmer™ implementa una sencilla pila TCP/IP para la radio CC2420, en lugar del estándar IEEE 802.15.4 [Std03]. Para optimizar el consumo energético del nodo, decidimos implementar una versión reducida del modo beacon-enabled del protocolo IEEE 802.15.4 usando slots de tiempo garantizados (en inglés: "guaranteed time slots" o GTS) [Std03] y una topología de estrella (cada nodo sensor sólo puede enviar datos a un dispositivo que actúa como coordinador de la WBSN).

En este modo, el coordinador envía beacons de manera periódica, que son utilizados por el resto de los nodos para sincronizar sus transmisiones con sus correspondientes GTS. Estos beacons se envían cada beacon interval (BI) = $15{,}36 \cdot 2^{BO} ms$, donde BO es beacon order, $0 \leq BO \leq 14$. Tras la

recepción del beacon, comienza un un periodo activo llamado superframe, cuya estructura se muestra en la Figura 9. La duración del superframe (SD) es igual a $15{,}36 \cdot 2^{SO} ms$, donde SO es el superframe order, que tiene que ser menor o igual a BO. En este caso, el valor del SO se ha establecido en 4, lo que implica una SD de 245,76 ms dividida en 16 intervalos de tiempo iguales. Con el fin de permitir que varios nodos sensores formen parte de la red, se han asignado dos intervalos de tiempo por GTS, o lo que es lo mismo, 30,2 ms por GTS. Usando esta configuración, hasta 8 nodos pueden estar simultáneamente dentro de la WBSN. Observemos que, si bien el superframe se divide en un periodo de contención de acceso (CAP) y un periodo libre de contención (CFP), el protocolo propuesto sólo se utiliza este último periodo, ya que se busca una operación libre de colisiones.

El tamaño de los paquetes utilizados en este protocolo es de 127 bytes (11 para la cabecera, 114 de datos y 2 para la suma de verificación). Por lo tanto, cada nodo puede enviar un máximo de cuatro paquetes durante su GTS, porque se necesitan 7,36 ms para llenar el buffer de transmisión de la radio CC2420 y enviar el paquete a través del aire, como se mostrará en la Sección 0.3.3.1.

El protocolo propuesto utiliza un enfoque TDMA, que ha demostrado ser ventajoso en WBSNs con topología de estrella, ya que su naturaleza centralizada elimina la aparición de colisiones y overhearing. Por otra parte, este protocolo usa muy pocos paquetes de control (sólo se envían los beacons para sincronizar a los nodos con sus GTS y los ACKs que el coordinador envía a los nodos para confirmar la recepción de los paquetes transmitidos). Además, el idle listening se reduce drásticamente, ya que los nodos saben con mucha precisión cuándo tienen que estar en modo de recepción para escuchar los beacons y ACKs enviados por el coordinador.

Este protocolo requiere una sincronización muy precisa para mantener la coordinación de la red y minimizar los periodos de idle listening, ya que solo una buena sincronización permitirá a los nodos ajustar con precisión los instantes de tiempo en los que tienen que encender su radio para esperar la recepción del beacon. Además, cada nodo de una WBSN puede muestrear una o más constantes vitales (monitorización multiparamétrica), sobre las que se aplican un conjunto de algoritmos con el fin de acondicionar la señal o extraer información significativa, que posteriormente se pueden combinar para realizar un análisis global que tenga en cuenta todas las señales. Estos algoritmos también pueden cambiar de forma dinámica, en función del rendimiento deseado en cada momento. El uso de un OS ayuda a lograr una gestión eficiente de los recursos hardware del nodo. Por otra parte, el OS debe proporcionar una planificación eficaz de las distintas tareas ejecutadas en el microcontrolador, tales como algoritmos de procesamiento de señal, control del protocolo MAC o gestión del muestreo de la señal, así como ofrecer comunicación entre procesos. Finalmente, el uso de un OS permite desacoplar

los recursos hardware y software del nodo, proporcionando una mayor flexibilidad, puesto que el programador no tiene que preocuparse del hardware sobre el que van a ejecutarse sus aplicaciones. Por esta razón, introducimos el uso de un OS.

0.3.2. Sistema operativo

Un OS para WBSNs debe dar soporte para operaciones en tiempo real, que le permitan hacer frente a las altas frecuencias de muestreo requeridas por algunas señales biométricas y a la estricta sincronización impuesta por el protocolo MAC propuesto anteriormente. El OS debe también ser multitarea, ya que un nodo ejecuta diferentes tareas software simultáneamente. Estas tareas comprenden el muestreo de la señal, la gestión del protocolo MAC y la ejecución de los algoritmos que se aplican sobre las distintas señales adquiridas. Por último, la sobrecarga del OS en cuanto a computación y uso de memoria debe ser lo más pequeña posible, pues los recursos de procesamiento y almacenamiento de los nodos WBSN son muy limitados.

Tras descartar el uso de un OS orientado a WSNs como TinyOS [LMP+04] o Contiki [DGV04], principalmente porque no ofrecen soporte para operaciones en tiempo real como requieren las aplicaciones consideredas en este trabajo, se decidió utilizar FreeRTOS [biba], que aunque no está específicamente diseñado para usarse en WSNs, cumple todos los requisitos para ser utilizado en dispositivos de una WBSN.

FreeRTOS [biba] es un sistema operativo para sistemas empotrados ligero, portable y de código abierto, que da soporte para operaciones en tiempo real. Sus principales ventajas son:

- Tiene un completo planificador de tareas, con asignación de prioridades.
- Incluye varios modelos de memoria.
- Implementa colas, semáforos, monitores, mutexes para la comunicación y sincronización entre tareas o entre tareas e interrupciones.
- Soporta temporizadores software eficientes.
- Potente funcionalidad de trazado de la ejecución.
- Opciones de detección de desbordamiento de la pila.
- Herramientas de desarrollo gratuitas para muchas de las arquitecturas soportadas.

FreeRTOS está diseñado para ser pequeño y simple. El núcleo se compone de sólo tres o cuatro archivos en C. Por otra parte, todas las características

anteriormente enumeradas pueden seleccionarse por separado para ser incluidas o no, lo que permite crear una versión totalmente personalizada del OS que sólo incluye las funcionalidades necesarias, redudiendo por tanto el tamaño del programa. Esta posibilidad permite la creación de aplicaciones multitarea y en tiempo real muy ligeras. Por estas razones, se ha portado FreeRTOS al Shimmer™, y se han implementado las siguientes tareas:

- Alta prioridad: gestión del protocolo MAC.
- Prioridad media: captura, digitalización y almacenamiento de la señal.
- Baja prioridad: algoritmo de procesamiento de señal, incluyendo también el empaquetado de los datos que deben ser enviados por radio.

Cuando ninguna de estas tareas se está ejecutando, el planificador pone al microcontrolador en modo de bajo consumo. Esto permite una gestión eficiente de los recursos de la plataforma en cuanto a rendimiento y consumo de energía, al mismo tiempo que se tiene una reducida sobrecarga asociada con el OS.

FreeRTOS está portado a varios microcontroladores. En particular, existe soporte para el microcontrolador TI MSP430 y para la radio CC2420, compatible con el estándar IEEE 802.15.4, incluida en el Shimmer™.

0.3.3. Caracterización del consumo de energía del Shimmer™

En esta sección se presenta una caracterización de la temporización y el consumo de energía de todos los componentes del Shimmer™ en sus diferentes modos de trabajo. Con esta información, que no está completamente disponible en los datasheets de los componentes, es posible hacer una estimación muy precisa del consumo de energía del Shimmer™ (y por tanto de su vida útil) bajo determinadas circunstancias de trabajo.

0.3.3.1. Marco de trabajo para los experimentos

Este trabajo se basa en la plataforma Shimmer™, aunque todos los resultados obtenidos para esta plataforma se pueden extrapolar a cualquier otra plataforma para WBSNs. En adelante, asumimos el uso de FreeRTOS y el protocolo MAC propuesto en la Sección 0.3.1. Para medir el consumo de energía, se coloca una resistencia de 10,3 Ω en la alimentación del nodo. La tensión se mide con un osciloscopio, y se calculan la corriente correspondiente y el consumo de energía. Utilizando esta configuración, se puede medir el consumo de energía de cada componente del nodo por separado habilitando el componente cuyo consumo queremos medir y deshabilitando el resto de los componentes de la plataforma. De esta forma se obtiene un perfil muy detallado del consumo de energía de cada bloque en todos sus modos de trabajo.

Además, este perfil proporciona información de, por ejemplo, cuánto tiempo se requiere para transferir un paquete del microcontrolador a la radio, cuánto tiempo se tarda en enviar un paquete por radio, el tiempo requerido para pasar de un modo de trabajo a otro, etc.

Para todos los experimentos se han utilizado dos nodos: uno de ellos se encarga de muestrear el ECG, procesarlo y enviar los resultados a otro, que actúa como coordinador. Dado que usamos un esquema GTS, los resultados obtenidos con esta configuración puede extrapolarse a una red con más nodos, pues las comunicaciones de los demás nodos no interfieren entre sí.

La Figura 10 muestra la traza de consumo de energía de un nodo que ejecuta una simple aplicación de streaming, en el que la señal ECG completa se envía al coordinador de la WBSN sin realizar ningún procesamiento de la misma. En cuanto a la radio, se pueden distinguir tres fases principales en la Figura 10: recepción del beacon, modo de bajo consumo (LP) y transmisión. Durante la recepción del beacon, la radio pasa al modo de recepción antes de la llegada del beacon, y entonces lo recibe. De esta forma, en la Figura 10 se pueden apreciar dos partes claramente diferenciadas en la fase de recepción del beacon: (1) la primera parte tiene una duración de 1,39 ms y un consumo de 72,39 mW, y corresponde a tener la radio en modo de recepción esperando el beacon mientras que el microcontrolador está inactivo; (2) la segunda parte tiene una duración de 0,97 ms y un consumo de 82,59 mW, y corresponde a tener la radio en modo de recepción que transfiere al microcontrolador el beacon recibido, el microcontrolador está por tanto activo para analizar el beacon recibido. Después de la recepción del beacon, la radio nodo entra en un modo LP hasta el comienzo de su GTS. Finalmente, durante su GTS, el nodo transmite la señal ECG al coordinador. La Figura 10 muestra la transmisión de dos paquetes de datos. Durante el envío de un paquete se distinguen dos zonas: (1) la primera se corresponde con el periodo durante el cual el microcontrolador envía el paquete al chip de radio CC2420 (que se encuentra en modo inactivo), consumiendo 16,8 mW durante 2,85 ms; (2) la segunda zona representa el periodo durante el cual la radio transmite el paquete por el aire mientras que el microcontrolador cambia de nuevo modo inactivo, disipando 51,92 mW durante 4,51 ms. Después de transmitir el paquete, se aprecia cómo la radio pasa automáticamente a modo de recepción, para esperar un ACK del coordinador que indica que el paquete se ha recibido correctamente. La recepción del ACK dura 0,75 ms y tiene un consumo asociado de 72,39 mW. Durante todo el proceso, el microcontrolador está continuamente actualizando el contador de ticks del sistema (cada 0,32 ms) y la tarea de muestreo lee una muestra del ADC cada 4 ms. El muestreo se aprecia claramente durante la fase de bajo consumo de la radio. La Tabla 1 resume los consumos del microcontrolador y la radio para los diferentes modos de trabajo de nuestra plataforma, mientras que la Tabla 2 muestra la temporización de las fases de la radio anteriormente mencionadas.

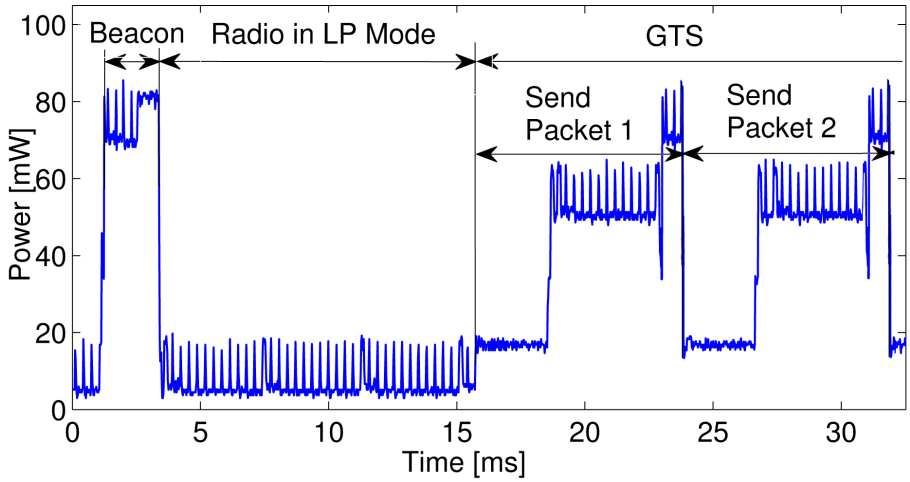

Figura 10: Potencia disipada por el Shimmer™ durante el muestreo, procesamiento y transmisión

Tabla 1: Caracterización del consumo del Shimmer™ (en mW)

	CPU	Radio	Total
Radio en modo de recepción (μC inactivo)	6,60	65,79	72,39
Radio en modo de recepción (μC activo)	16,80	65,79	82,59
Transmisión del paquete (del μC a la radio)	16,80	0,00	16,80
Transmisión del paquete (por el aire)	6,60	45,32	51,92
Muestreo (μC y radio inactivos)	6,60	0,00	6,60
Muestreo (μC activo and radio inactiva)	16,80	0,00	16,80

Tabla 2: Temporización del Shimmer™ (en ms)

Recepción del beacon (radio en recepción y μC inactivo)	1,39
Recepción del beacon (radio en recepción y μC activo)	0,97
Transmisión del paquete (del μC a la radio)	2,85
Transmisión del paquete (por el aire)	4,51
Recepción del ACK	0,75

0.3.3.2. Caso base: ECG streaming

Aunque una plataforma WBSN está equipada con un microcontrolador que le dota capacidad de procesamiento, en la mayoría de los casos este microcontrolador simplemente actúa como unidad de control para gestionar

todos los recursos hardware del nodo. En lugar de procesar los datos obtenidos por los sensores, la señal completa se envía al coordinador de la WBSN. Una de las principales aportaciones de este trabajo consiste en realizar diagnóstico a bordo del nodo para poder detectar posibles arritmias en tiempo real (tradicionalmente este procesamiento se ha realizado a posteriori). De este modo, la próxima sección propone varias técnicas que explotan el microcontrolador para reducir la cantidad de información que debe ser enviada y por tanto el consumo de energía de la radio, lo que implica una mayor vida útil del nodo sensor. Esta sección presenta un caso base en el que se realiza la transmisión de la señal ECG completa, que servirá como referencia para evaluar las posibles mejoras o penalizaciones en consumo de energía resultantes al aplicar distintas técnicas de procesamiento a bordo del nodo que se proponen en la sección siguiente.

Teniendo en cuenta la caracterización del consumo que se acaba de presentar, la Tabla 3 muestra el consumo de energía del Shimmer™ ejecutando una aplicación que simplemente envía la señal ECG en bruto al coordinador de la WBSN sin realizar ningún tipo de procesamiento previo. En este caso, si se quiere extraer información relevante a partir del ECG, el coordinador debería usar un software adicional para procesar la señal, como es el caso en los dispositivos Holter. La primera fila indica el ciclo de trabajo del microcontrolador. Dado que la aplicación no lleva a cabo ningún tipo de procesamiento, el ciclo de trabajo en este caso es 0. La segunda fila indica el intervalo de tiempo entre la generación de dos paquetes de datos consecutivos. Teniendo en cuenta que el cuerpo de un paquete es de 114 bytes, que una muestra se representa con 1,5 bytes y que la frecuencia de muestreo es de 250 Hz, la aplicación de streaming genera un paquete cada 304 ms. La tercera fila muestra el beacon interval, que se ajusta de acuerdo al intervalo de tiempo entre paquetes consecutivos. Teniendo en cuenta que la longitud de la cola de FreeRTOS dedicada a almacenar los paquetes listos para ser transmitidos esta limitada a un máximo de 4 paquetes, BCO debe ser menor o igual a 6. Se elige el valor más alto, para reducir tanto como sea posible el consumo de energía en la radio, por lo tanto, el beacon interval es 983ms. La cuarta fila informa del consumo total de energía por segundo, que es 7,70mJ en este caso. Finalmente, la última fila se calcula directamente a partir de la cuarta, asumiendo que la fuente de energía del Shimmer™ es una batería de Li-ion de 280 mAh a 3,7 V, y muestra la vida útil del sistema, que es 134,6 horas.

La Figura 11 muestra el porcentaje del consumo total de energía del Shimmer™ debido a muestreo y comunicación por radio para una aplicación que hace streaming del ECG. El consumo del microcontrolador no se considera en la figura, ya que esta aplicación no requiere ningún tipo de tratamiento. El consumo de energía debido al muestreo del ECG es constante, y no puede reducirse a menos que reemplacemos el ADC por otro de menor

Tabla 3: Vida útil del Shimmer™ haciendo streaming del ECG

Ciclo de trabajo (%)	0
Paquete listo cada... (ms)	304
Beacon interval (ms)	983
Consumo de energía (mJ)	7,70
Vida útil (h)	134,6

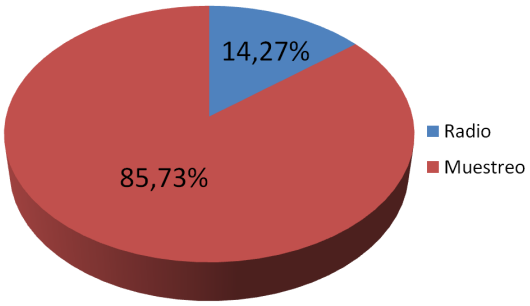

Figura 11: Consumo de energía del Shimmer™ debido a muestreo y comunicación por radio haciendo streaming del ECG

consumo. Por lo tanto, la única parte del consumo de energía que se puede reducir es la de la comunicación por radio. Para lograr esto, la siguiente sección propone varios algoritmos que extraen sólo la información relevante de la señal ECG para así reducir significativamente la cantidad de datos que debe ser transmitida, y por tanto reducir el consumo de energía de la radio. Si la sobrecarga computacional de estos algoritmos es lo suficientemente baja, la vida útil de la plataforma no se verá afectada (o incluso puede aumentarse), mientras que el sistema será capaz de realizar procesamiento inteligente del ECG y ofrecer un diagnóstico en tiempo real.

0.4. Optimizaciones a nivel de aplicación

En la Sección 0.2 hemos visto que las plataformas WBSN están equipadas con microcontroladores, que las dotan de capacidad de procesamiento y memoria. Estos microcontroladores pueden explotarse para ejecutar algoritmos de procesamiento de la señal ECG en tiempo real, tales como filtrado, compresión, extracción de puntos característicos, detección de arritmias, etc. Estos algoritmos pueden reducir significativamente la cantidad de información que debe transmitirse y por lo tanto reducir el consumo de energía derivado de la transmisión de datos. Además de la posible reducción del consumo total de energía del nodo, la ejecución de aplicaciones a bordo del

nodo puede ser de gran ayuda para los pacientes y el personal sanitario, pues proporcionan información y diagnóstico en tiempo real del estado de salud del paciente. Por lo tanto, es necesario optimizar estos algoritmos para su ejecución en tiempo real en una plataforma WBSN, con el fin de reducir la complejidad computacional y la memoria requeridas por los mismos y por lo tanto minimizar el impacto del microcontrolador en el consumo de energía total del nodo.

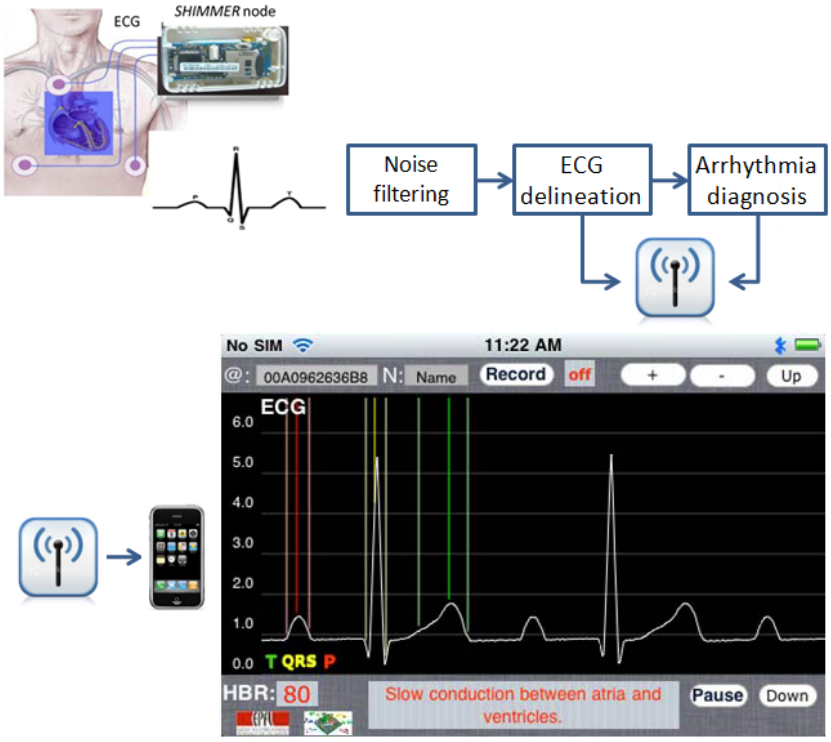

Figura 12: Arquitectura software del sistema de monitorización de ECG propuesto

La Figura 12 muestra la arquitectura software del monitor inalámbrico de ECG propuesto en este trabajo. En primer lugar, la señal ECG se muestrea y acondiciona para reducir el ruido y los artefactos (ECG filtering). Entonces, se detectan los puntos que marcan el principio, pico y final de las ondas características del ECG (ECG delineation). Finalmente, esta información se utiliza para realizar un diagnóstico de arritmias. Los resultados obtenidos después de las fases de delineación y diagnóstico de la arritmias (y, opcionalmente, la señal ECG completa) se transmiten de forma inalámbrica al coordinador WBSN, donde estos datos se pueden mostrar o almacenar.

En esta sección se presentan varios algoritmos de filtrado de ECG, delineación y detección de arritmias para su ejecución en el Shimmer™. Por último, se ofrece una evaluación del rendimiento de dichos algoritmos, así

como un estudio comparativo de su consumo de energía (y por tanto de la vida útil del nodo que los ejecuta).

0.4.1. Filtrado de ECG

Con el fin de mejorar la calidad de la señal del ECG y por tanto poder realizar un diagnóstico más preciso, es muy importante reducir tanto como sea posible la presencia de ruido, como el ruido de línea de base o ruido electromiográfico. El ruido de línea de base consiste en una actividad de baja frecuencia externa al ECG que puede interferir con el análisis de la señal, haciendo que su interpretación clínica sea incorrecta. El ruido de línea de base es a menudo inducido por el ejercicio y puede tener su origen en multitud de fuentes, que incluyen transpiración, respiración, movimientos del sujeto o contacto deficiente de los electrodos. El contenido en frecuencia del ruido de línea de base se encuentra generalmente en el rango entre 0,05 y 1 Hz [fQEWP85], pero, durante el ejercicio intenso, puede contener frecuencias más altas.

El ruido electromiográfico, que está causado por la actividad eléctrica de los músculos durante los períodos de contracción, se encuentra comúnmente en los ECGs registrados durante periodos de monitorización ambulatoria. Este tipo de ruido puede ser intermitente, debido a un movimiento repentino cuerpo. Las componentes de frecuencia de este ruido se solapan considerablemente con las del complejo QRS.

En esta sección se proponen dos técnicas para filtrado de ECG: una basada en cubic splines y otra en filtrado morfológico. La primera sólo elimina el ruido de línea de base, mientras que la segunda también incluye una fase de reducción de ruido de alta frecuencia como el electromiográfico o el causado por interferencias inducidas por líneas eléctricas cercanas.

Estimación del ruido de línea de base usando cubic splines

Esta técnica, basada en [MK77], utiliza un polinomio de tercer grado para aproximar el ruido de línea de base, que luego se resta de la señal original. Para este fin, se elige una muestra representativa (o knot) de cada latido, dentro del segmento isoeléctrico que va desde el final de la onda P al comienzo del complejo QRS. Entonces se calcula un polinomio de tercer grado que pase por tres knots consecutivos. Este polinomio, que representa una aproximación del ruido de línea de base, se resta del ECG original para obtener la señal filtrada.

Filtrado morfológico

Este método, introducido por primera vez en [SCK02], aplica varias operaciones morfológicas (erosiones y dilataciones) a la señal ECG para estimar el ruido de línea de base. Los operadores morfológicos son muy utilizados

para procesamiento de señales e imágenes debido a su gran robustez y rendimiento para extraer información de la forma de las señales, además de ser operaciones sencillas y rápidas de calcular. En este tipo de operaciones se usan dos conjuntos de datos: por un lado la señal a transformar y por otro una serie de elementos estructurales que se diseñan según las características morfológicas que se quieren extraer de la señal.

Este método consta de dos fases. En la primera se aplican varias operaciones de erosión y dilatación a la señal ECG original para estimar el ruido de línea de base. Primero se aplica una erosión seguida de una dilatación, que elimina los picos en la señal. Entonces, se aplica a la onda resultante una dilatación seguida de una erosión para eliminar los valles. El resultado final es una estimación del ruido de línea de base, que se resta a la señal original para eliminarlo. En el segundo paso, se reduce el ruido de más alta frecuencia por medio de varias operaciones adicionales de erosión y dilatación aplicadas en paralelo, usando unos elementos estructurales especiales que ayudan a mantener los picos y valles de las ondas importantes (complejo QRS y ondas P y T).

0.4.2. Delineación de ECG

Se ha dedicado un gran esfuerzo a la investigación para automatizar el análisis de señales ECG y la detección de sus ondas más significativas [SL05]. De hecho, el rendimiento de un sistema de análisis automatizado de ECG depende críticamente de una detección precisa de estas ondas (delineación de ECG). Esto ha motivado la aparición de gran variedad de técnicas para delineación de ECG.

Dos de las técnicas más sobresalientes para delineación automática de ECG se analizan en esta sección. La primera se basa en la transformada wavelet (WT) [LZT95, MAO$^+$04], y la segunda en la derivada multiescala morfológica (MMD) [SCK05]. Siguiendo estos dos enfoques, se han desarrollado varios algoritmos para la delineación de una (single-lead) o varias (multi-lead) derivaciones, que además se han optimizado para su ejecución en el Shimmer™. Esta sección describe los algoritmos propuestos y analiza su rendimiento en términos de precisión.

Para evaluar la precisión de la delineación se utiliza la base de datos QT (QTDB) [LMGM97], que contiene 105 señales ECG de 2 derivaciones, muestreadas a 250 Hz y con una duración de 15 minutos. Los puntos característicos del ECG han sido anotados manualmente por cardiólogos, lo que permite una fácil evaluación de la precisión de los algoritmos de delineación automática. Las estadísticas que se utilizan en esta evaluación son la media (m) y desviación estándar (σ) de los errores en la detección de cada uno de los puntos. Además, se cuentan los verdaderos positivos (TP), falsos positivos (FP) y falsos negativos (FN) para calcular la sensibilidad (Se) y el

valor predictivo positivo ($P+$), que según [EC508] se definen como:

$$Se = \frac{TP}{TP + FN}, \quad P+ = \frac{TP}{TP + FP} \tag{1}$$

Para evaluar la delineación single-lead, el algoritmos se ejecuta por separado para las 2 derivaciones de la QTDB y, para cada uno de los puntos, se elige la detección más próxima a la anotación manual.

0.4.2.1. Delineación single-lead basada en WT

Esta técnica lleva a cabo la detección de todos los puntos característicos (inicio, pico y final) de las ondas del ECG mediante una WT, que produce derivadas de la señal ECG de entrada en cinco escalas diádicas (de la 2^1 a la 2^5). La elección de estas escalas viene motivada por la observación de que la mayor parte de la energía de la señal ECG se encuentra dentro de ellas [MAO+04]. En particular, se ha demostrado que la energía del complejo QRS es menor en las escalas superiores a la 2^4 y que las ondas P y T tienen componentes importantes en la escala 2^5.

Una vez calculados los resultados de la WT, se aplican un conjunto de reglas para extraer los puntos importantes de la señal ECG. El primer punto detectado es el pico principal del complejo QRS, entonces el complejo QRS se delinea, detectándose su inicio y final. Luego se detectan los picos de las ondas P y T, junto con sus inicios y finales de onda.

Para este primer conjunto de experimentos, se ejecuta el algoritmo propuesto en el Shimmer™ y se calculan los cuatro valores que indican la precisión de la delineación. La Tabla 4 muestra los resultados obtenidos, junto con los del algoritmo original que funciona de forma offline [MAO+04]. La última columna indica las tolerancias máximas de desviación estándar exigidas por el CSE [fQEWP85]. Estos resultados ponen de manifiesto que la precisión del algoritmo propuesto está dentro de las tolerancias máximas de desviación estándar para todos los puntos, excepto el QRS_{onset} and P_{onset}, que solo están por encima de las tolerancias por una fracción del tiempo de muestreo. Por otra parte, la Tabla 4 muestra que la implementación propuesta para el Shimmer™ supera ampliamente al algoritmo propuesto en [MAO+04] en términos de desviación estándar, con el valor añadido de que la implementación propuesta se ejecuta en tiempo real y preservando una alta sensibilidad y valor predictivo positivo.

0.4.2.2. Delineación single-lead basada en MMD

Este método usa operadores morfológicos (erosión y dilatación), que han demostrado una gran eficiencia en el análisis de señales de cualquier dimensión cuyos puntos significativos están localizados en máximos y mínimos locales. Los puntos más importantes del ECG (comienzo, pico y final del

Tabla 4: Precisión del algoritmo de delineación single-lead basado en WT vs. caso base [MAO$^+$04]

Método	este trabajo 16-bit int Shimmer™	Martinez [MAO$^+$04]	Tolerancias ($2s_{CSE}$)
Parámetro	Se (%) P^+_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^+_{min} (%) $m \pm \sigma$ (ms)	σ (ms)
P_{onset}	99,87 91,98 8,6 ± 11,2	98,87 91,03 2,0 ± 14,8	10,2
P_{peak}	99,87 92,46 10,1 ± 8,9	98,87 91,03 3,6 ± 13,2	-
P_{end}	99,91 91,70 0,9 ± 10,1	98,75 91,03 1,9 ± 12,8	12,7
QRS_{onset}	99,97 98,61 3,4 ± 7,0	99,97 N/A 4,6 ± 7,7	6,5
QRS_{end}	99,97 98,72 3,5 ± 8,3	99,97 N/A 0,8 ± 8,7	11,6
T_{peak}	99,97 98,91 3,7 ± 13,0	99,77 97,79 0,2 ± 13,9	-
T_{end}	99,97 98,50 −2,4 ± 16,9	99,77 97,79 −1,6 ± 18,1	30,6

complejo QRS y las ondas P y T) pueden considerarse como tales puntos significativos [MZ92a, Wit84].

El algoritmo se divide en tres pasos: filtrado, aplicación de MMD y detección de los puntos característicos. Para el primer paso se utiliza el filtro morfológico presentado en la Sección 0.4.1, que acondiciona la señal ECG original para minimizar el ruido, de forma que se pueda realizar una detección más precisa de los puntos característicos. Este paso es primordial puesto que la presencia de ruido de línea de base puede alterar significativamente los resultados de la delineación.

Para el segundo paso, y de acuerdo con [SCK05], los puntos característicos de la onda de ECG se corresponden con máximos y mínimos de la señal.

Por lo tanto, un punto característico se define como un punto en el que las derivadas a su izquierda y a su derecha existen y tienen distinto signos. Para facilitar la detección de estos puntos, la señal se transforma usando la MMD, de forma que un pico positivo de la señal ECG (la derivada es positiva a su izquierda y negativa a su derecha) se transforma en un mínimo local. En el comienzo y final de las ondas positivas, hay un aumento repentino de la derivada a su derecha con respecto a su izquierda, por lo que estos puntos se transforman en máximos locales.

Finalmente, el algoritmo detecta los puntos importantes en la señal transformada, estudiando sus máximos y mínimos locales que superan unos determinados umbrales, que se ajustan dinámicamente a la morfología de cada una de las ondas. La Figura 13 muestra una señal ECG (a), la misma señal tras realizar el filtrado morfológico (b) y la señal transformada después de aplicar la MMD (c).

La segunda columna de la Tabla 5 muestra los resultados del algoritmo propuesto ejecutándose en el Shimmer™. La desviación estándar está por debajo de las tolerancias máximas para todos los puntos, al contrario que el algoritmo original (tercera columna de la Tabla 5). Gracias a las mejoras que se han incluido en esta nueva versión del algoritmo, el error medio se ha reducido significativamente (4 ms de media). Esto significa que, en promedio, los puntos detectados por el algoritmo propuesto están 4 ms (es decir, una muestra) más cerca de los reales que los detectados por el algoritmo original.

La desviación estándar también se mejora con respecto al caso base para la onda P y el complejo QRS. En el caso de la onda T (filas T_{onset} y T_{end}), la desviación estándar obtenida por el algoritmo propuesto es mayor que en la versión original, pero esto no es un problema ya que los resultados obtenidos por nuestro algoritmo son significativamente inferiores a los límites establecidos por el CSE.

0.4.2.3. Delineación multi-lead basada en WT

En la práctica clínica, sin embargo, se usa más de una derivación (12 en los monitores estándar que se utilizan en los hospitales o 3 en monitorización ambulatoria usando dispositivos de tipo Holter). La utilización de un esquema multi-lead requiere el desarrollo de técnicas que sean capaces de explotar la información contenida en las derivaciones disponibles para mejorar la precisión, la estabilidad y la resistencia frente a ruido de la delineación, en comparación con un enfoque single-lead.

Hay dos métodos para la delineación multi-lead. El primero consiste en delinear cada una de las derivaciones por separado y aplicar reglas de selección sobre los puntos detectados para quedarse solo con los "mejor" delineados. Cuando sólo se dispone de dos derivaciones, como es el caso en este trabajo, debido a las limitaciones de memoria del Shimmer™, este enfoque no es aplicable, y debe ser reemplazado por una selección óptima en

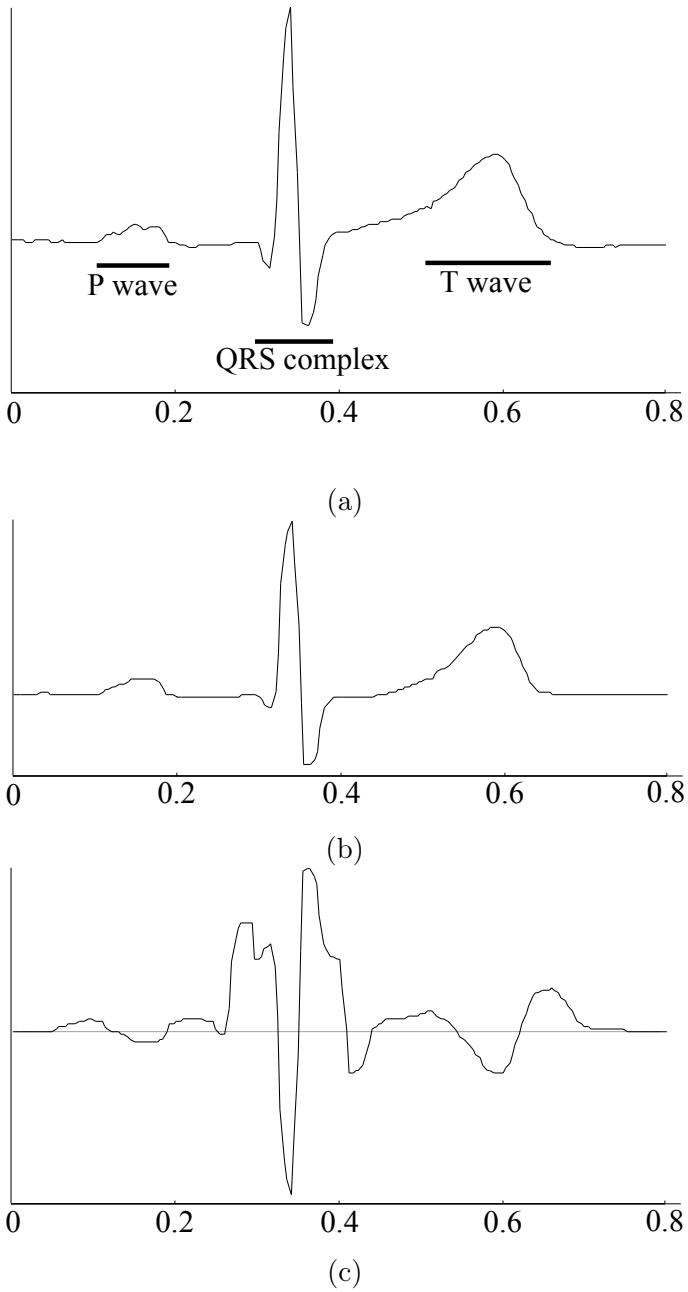

Figura 13: (a) Señal ECG original (b) ECG después de aplicar el filtrado morfológico (c) Señal resultante de aplicar la MMD a la señal filtrada

la que se elige el punto que presenta el menor error. Sin embargo, este enfoque no es realista, y solo se puede utilizar para propósitos de validación, puesto que el algoritmo de delineación no puede saber cuál es la derivación que presenta menor error para cada uno de los puntos. El segundo méto-

Tabla 5: Precisión del algoritmo de delineación single-lead basado en MMD tanto si se ejecuta en un PC como en el Shimmer™ vs. caso base [SCK05]

Método	MMD mejorado floating-point PC	MMD mejorado 16-bit int Shimmer™	MMD original	Tolerancias ($2s_{CSE}$)
Parámetro	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	σ (ms)
P_{onset}	99,75 96,66 3,6 ± 9,6	99,15 96,70 3,5 ± 10,7	97,2 - 9,0 ± 9,4	10,2
P_{peak}	99,91 96,67 2,6 ± 7,7	99,28 96,65 1,2 ± 7,8	- - -	-
P_{end}	98,56 96,98 2,2 ± 9,4	98,68 96,95 1,1 ± 10,0	94,8 - 12,8 ± 13,2	12,7
QRS_{onset}	99,37 100,0 1,0 ± 6,4	99,03 100,0 0,7 ± 6,7	100,0 - 3,5 ± 6,1	6,5
QRS_{end}	99,89 99,97 −0,4 ± 6,8	99,78 99,97 −0,2 ± 7,0	100,0 - 2,4 ± 10,3	11,6
T_{onset}	95,11 83,42 5,7 ± 26,7	96,10 83,96 8,7 ± 25,8	99,8 - 7,9 ± 15,8	-
T_{peak}	99,75 100,0 −0,3 ± 15,0	99,60 100,0 0,7 ± 13,1	- - -	-
T_{end}	95,23 97,92 −7,3 ± 19,0	95,37 97,80 −6,0 ± 17,0	99,6 - 8,3 ± 12,4	30,6

do de delineación multi-lead se basa en la combinación de las derivaciones para construir una única señal – que da una visión global del funcionamiento del corazón, independientemente de las derivaciones que se usen – sobre la que se aplica un algoritmo de delineación single-lead. En la literatura podemos encontrar diferentes técnicas para la fusión de todas las derivaciones en una única señal: (1) la combinación basada en root mean square (RMS) [IvOH06, SML06, RBB+09]; (2) una derivada obtenida usando una WT a partir de 2 o 3 derivaciones ortogonales [AMRL09]; o (3) la combinación de las derivadas de todas las derivaciones disponibles [IM10]. Dadas las limitaciones de procesamiento y memoria del nodo utilizado, se elige el enfoque basado en RMS, puesto que tienen una complejidad computacional

muy baja y una gran escalabilidad.

En consecuencia, las L derivaciones disponibles $x_l[n]$, con $l = 1, \ldots, L$, se combinan en una única señal $x_{RMS}[n]$:

$$x_{RMS}[n] = \sqrt{\frac{1}{L}\sum_{l=1}^{L} x_l^2[n]}, \tag{2}$$

sobre la cual se realiza una delineación single-lead. Para una correcta combinación de las diferentes derivaciones, es fundamental eliminar en cada una de ellas el ruido de línea de base (causado principalmente por la respiración y los cambios de impedancia de los electrodos debidos a la transpiración y los movimientos del cuerpo) antes de calcular el RMS [IvOH06, SML06]. Por lo tanto, ya que la calidad de la posterior delineación dependerá de la corrección de línea de base, este trabajo evalúa la efectividad de los dos filtros propuestos en la Sección 0.4.1: cubic splines y filtrado morfológico.

Para la delineación de la señal x_{RMS} se usa el algoritmo basado en WT propuesto en la Sección 0.4.2.1, ya que su tiempo de ejecución es mucho menor que el basado en MMD y por lo tanto implica un menor consumo de energía, ofreciendo una gran precisión en la detección de los puntos característicos del ECG.

Si comparamos la precisión de los algoritmos single-lead para la primera y la segunda derivación de la QTDB (tercera y cuarta columna de la Tabla 6, respectivamente) y la obtenida con un enfoque multi-lead (primera y secunda columna de la Tabla 6), se confirma que el uso de múltiples derivaciones puede aumentar la precisión de la delineación con respecto a la utilización de una única derivación. Los dos filtros para eliminar el ruido de línea de base ofrecen resultados muy parecidos de error medio (m) y desviación estándar (σ), por encima de las tolerancias máximas por sólo una fracción del tiempo de muestreo.

Además, la Tabla 6 muestra también que los resultados de la delineación multi-lead son muy parecidos a los de la delineación de la mejor derivación (obtenidos para la derivación 1), y superan por casi una muestra (4 ms) los de la delineación de la peor derivación (obtenidos para la derivación 2). Esto confirma que la combinación de las derivaciones disponibles mediante RMS reduce la dependencia de los resultados en determinadas derivaciones, a la vez que ofrece mayor robustez y estabilidad.

0.4.3. Diagnóstico de arritmias

Los resultados de los algoritmos de delineación previamente estudiados se pueden utilizar para realizar diagnóstico en tiempo real a bordo del nodo sensor, para proporcionar al paciente información sobre el estado de su corazón, así como reducir significativamente la cantidad de datos que se transmiten

Tabla 6: Comparativa de la precisión de los algoritmos de delineación multi-lead con los single-lead

Método	2-lead 16-bit int Cubic spline	2-lead 16-bit int Morpho.	Single-lead 16-bit int Lead 1	Single-lead 16-bit int Lead 2	Toleranc. ($2s_{CSE}$)
Parám.	Se (%) P^+_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^+_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^+_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^+_{min} (%) $m \pm \sigma$ (ms)	σ (ms)
P_{on}	94,02 95,15 4,1 ± 17,1	96,24 91,19 2,5 ± 16,4	98,84 92,66 13,4 ± 14,8	97,31 91,76 10,4 ± 19,4	10,2
P_{peak}	94,02 95,42 12,8 ± 12,1	96,24 91,38 14,7 ± 13,4	98,84 92,93 15,6 ± 12,1	97,09 92,21 6,9 ± 16,9	-
P_{end}	94,05 95,49 −2,1 ± 14,2	96,27 91,55 −1,9 ± 15,5	98,87 93,24 1,6 ± 13,2	96,81 91,43 −8,5 ± 17,9	12,7
QRS_{on}	99,67 99,23 3,2 ± 8,7	99,75 97,07 6,9 ± 8,0	99,61 99,56 5,4 ± 8,4	99,67 98,61 8,6 ± 12,6	6,5
QRS_{end}	99,67 99,20 7,3 ± 11,9	99,75 97,07 9,0 ± 9,1	99,61 99,56 1,5 ± 10,1	99,72 98,72 8,7 ± 13,1	11,6
T_{peak}	98,00 99,23 1,9 ± 17,4	98,33 98,58 5,3 ± 19,3	99,35 99,49 5,3 ± 18,4	99,35 98,90 3,5 ± 23,5	-
T_{end}	97,94 98,72 −4,1 ± 25,3	98,11 98,08 −11,4 ± 24,9	99,32 99,24 −5,3 ± 22,7	99,18 98,38 −4,6 ± 27,2	30,6

de forma inalámbrica, y por lo tanto, prolongar la vida útil del nodo. Para ilustrar esto, esta sección propone un algoritmo de diagnóstico que evalúa un conjunto de reglas que un ECG normal debe cumplir. Si alguna de las reglas no se cumple, el algoritmo genera una alarma que será enviada al coordinador WBSN.

Además, esta sección propone un enfoque más preciso, donde se adapta el algoritmo de diagnóstico para la detección de una arritmia cardíaca específica. En este caso, se propone un sistema de detección automática y en tiempo real de fibrilación auricular (AF), que también utiliza la salida de los algoritmos de delineación para llevar a cabo un análisis detallado de las irregularidades que aparecen durante los episodios de AF. Más concretamente, se estudian la variabilidad del intervalo R-R y la ausencia de la onda P. Los resultados obtenidos son muy similares a los de otras técnicas que realizan la detección de AF a posteriori sobre una señal grabada por un dispositivo de

tipo Holter, alcanzando una sensibilidad del 96 % y un 93 % de especificidad para la base de datos AFDB.

0.4.4. Consumo de energía versus rendimiento

De acuerdo con la caracterización de la plataforma Shimmer™ realizada en la Sección 0.3.3.1, esta sección propone un esudio del consumo de energía del nodo ejecutando una sencilla aplicación en la que solo se hace streaming del ECG y todos los algorimos de análisis en tiempo real propuestos en este trabajo, para evaluar el impacto del procesamiento a bordo del nodo en el consumo de energía total.

Tabla 7: Vida útil del Shimmer™ para los algoritmos de delineación

	ECG Stream.	Single lead WT	Single lead MMD	2-lead Morph. filt.	2-lead spline filt.
Ciclo de trabajo (%)	0	6,78	18,25	19,17	33,22
Paquete listo cada... (ms)	304	2250	2250	2250	2250
Beacon interval (ms)	983	7864	7864	7864	7864
Consumo de energía (mJ)	7,70	7,44	8,62	8,71	10,15
Vida útil (h)	134,6	139,2	120,2	118,9	102,1

La Tabla 7 muestra que el consumo total de energía aumenta con la carga computacional del algoritmo de delineación. Sorprendentemente, también evidencia que incluso el sencillo algoritmo de delineación single-lead basado en WT, que sólo consume 6,78 % del tiempo de CPU y reduce significativamente la cantidad de datos a transmitir (de 3,29 a 0,44 paquetes por segundo), ofrece una extensión marginal de la vida útil del nodo con respecto a streaming de ECG. Estos resultados destacan que, para las aplicaciones de WBSN, la radio no siempre es responsable de la mayor parte del consumo de energía del nodo, como se asume en otras áreas de aplicación de las WSNs. Para comprender mejor el impacto del procesamiento del ECG a bordo del nodo en su vida útil, desglosamos el consumo de energía de cada uno de los componentes del nodo y los analizamos por separado, viendo que los dos bloques con mayor consumo de energía son la radio y el microcontrolador.

Radio

El consumo de energía de la radio depende del protocolo MAC [Std03], que añade una cierta sobrecarga en la comunicación para controlar el acceso al medio compartido. La parte del consumo total de energía debido al uso del protocolo MAC sólo representa el 1,3 % para la delineación multi-lead con cubic splines, y un 14,3 % para el streaming de ECG. Por lo tanto, el

protocolo MAC basado en GTS que se ha propuesto es una buena opción para un escenario en el que se realiza una monitorización continua del ECG con envío periódico de resultados. Cabe mencionar que esta caracterización del nodo se hizo usando una configuración en la que los nodos implicados tenían visión directa entre sí. Un estudio detallado de la propagación de las ondas de radio alrededor del cuerpo se debería realizar para poder ajustar un poco mejor el consumo de la radio, sin embargo este estudio está fuera del alcance de esta tesis.

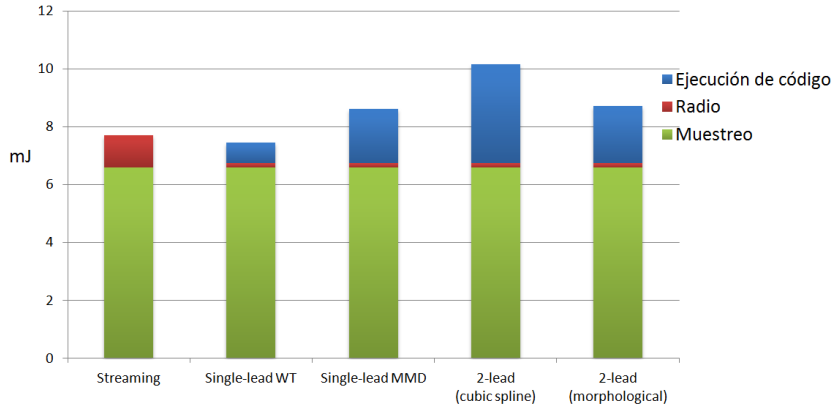

Figura 14: Desglose por componentes del consumo de energía del Shimmer™

Microcontrolador

El consumo debido a la CPU tiene un impacto significativo en la vida de nodo cuando se ejecutan los algoritmos de delineación propuestos, que requieren más potencia de cómputo que si simplemente se hace streaming del ECG. A excepción del algoritmo de delineación single-lead basado en WT, el aumento de actividad de la CPU lleva a una reducción de la vida útil del nodo, aunque el consumo de energía de la radio se reduzca 8,6 veces (con respecto a streaming). Por lo tanto, la ejecución de código puede ser muy costosa si la disipación de potencia del microcontrolador es alta o el algoritmo a ejecutar no ha sido cuidadosamente optimizado para la plataforma que se está utilizando. Por lo tanto, se debe llevar a cabo un estudio cuidadoso en el momento de elegir o diseñar el microcontrolador de cada plataforma WBSN. Este estudio es clave para tener en cuenta la disipación de potencia en modo activo, así como la posibilidad de usar una tensión de alimentación baja o modos de trabajo de bajo consumo que deshabilitan algunas partes del microcontrolador cuando no se utilizan.

Sin embargo, para el Shimmer™, el proceso dominante del microcontrolador en términos de consumo de energía no es la ejecución de código. Como se ilustra en la Figura 14, la mayor parte del consumo de energía se debe al muestreo de la señal del ECG (el 65,2 % para delineación multi-lead con

Tabla 8: Vida útil del Shimmer™ para los algoritmos de diagnóstico de arritmias

	ECG streaming	Diagnóstico general	Detección de AF
Ciclo de trabajo (%)	0	6,85	7,73
Paquete listo cada... (ms)	304	21000	21000
Beacon interval (ms)	983	62915	62915
Consumo de energía (mJ)	7,70	7,32	7,41
Vida útil (h)	134,6	141,5	139,9

cubic splines y el 85,7 % para streaming). Por lo tanto, podemos apreciar cómo el muestreo de la señal puede ser muy costoso en términos de consumo de energía en función de la implementación de hardware y la frecuencia de muestreo. En este caso particular, para la frecuencia de muestreo que se considera (250 Hz), no es posible deshabilitar el voltaje de referencia interno de los ADCs entre dos muestras consecutivas, porque el tiempo que el regulador de tensión interno tarda en alcanzar el nivel adecuado para la conversión es mayor que el tiempo entre muestras, así pues, el voltaje de referencia está siempre activo, aumentando significativamente el consumo total de energía. Por lo tanto, elegir un ADC de bajo consumo puede reducir en gran medida el consumo de energía de todo el sistema, pues la tarea de muestreo ha demostrado ser responsable de un gran consumo de energía en nuestro sistema WBSN de monitorización de ECG.

La Tabla 8 muestra el consumo de energía del Shimmer™ ejecutando una aplicación de streaming de ECG, así como de los algoritmos propuestos para diagnóstico general de arritmias y detección de AF. Ambos algoritmos sólo envían los resultados obtenidos (la clasificación de cada latido, indicando si es normal o se ha detectado alguna arritmia), por lo que solo es necesario enviar 4 bytes por latido (28 bits para indicar el momento en el que se ha detectado el latido y los restantes 4 bits para el código que contiene el resultado de la clasificación). Este código puede tomar 9 valores posibles, que indican si el latido es normal o si muestra AF o alguna otra de las 7 patologías detectadas por el algoritmo de diagnóstico general. Como un paquete puede contener un máximo de 114 bytes de datos, se pueden almacenar 28 latidos. Suponiendo una frecuencia cardiaca típica de 80 latidos por minuto, se genera un paquete cada 21 s.

De la Tabla 8 podemos concluir que a pesar de realizar procesamiento avanzado en el Shimmer™ y por tanto aumentar el consumo de energía del microcontrolador, la vida útil del nodo ejecutando el algoritmo de detección de AF en tiempo real es mayor que si sólo se hace streaming de ECG, debido a la enorme reducción en la cantidad de información que se transmite al

coordinador WBSN. Gracias a esta reducción, el algoritmo genera un paquete cada 21 s, disminuyendo el consumo de energía de la radio a sólo el 0,22 % del total de la plataforma para el caso de detección de AF. Si la frecuencia de la generación de paquetes se reduce aún más, lo cual es posible si el nodo sólo envia un paquete de alarma cuando se detecta una arritmia, el ahorro de energía será insignificante. Por ejemplo, si no se detecta ninguna alarma, el consumo de energía de la radio sería el 0,04 % del total, y la vida útil se incrementaría de 139,9 a 140,1 horas. Por lo tanto, esta observación concluye que la única manera de reducir aún más el consumo de energía de la plataforma pasa por optimizar el ADC o el microcontrolador.

0.5. Conclusiones

Como se mencionó en la introducción, las enfermedades cardiovasculares son la principal causa de mortalidad en todo el mundo, responsables del 30 % de las muertes a nivel mundial en 2005 y de un gasto económico de miles de millones de dólares [PW10]. Se prevé que estas cifras vayan en aumento debido al rápido envejecimiento de la población mundial y a su estilo de vida poco saludable. Las WBSNs están preparadas para ofrecer soluciones a gran escala y bajo coste, para disminuir el alto gasto sanitario que, de lo contrario, puede volverse insostenible en los próximos años.

En esta tesis se propone el desarrollo de un sistema de monitorización de ECG, que no sólo muestre continuamente el ECG del paciente, sino que además lo analice en tiempo real y sea capaz de dar información sobre el estado de su corazón. Para ello, en primer lugar, se ha llevado a cabo un estudio de las arquitecturas disponibles para WBSNs en la Sección 0.2, eligiéndose la plataforma Shimmer™ como caso práctico, ya que se adapta perfectamente a los requisitos de este trabajo. Esta plataforma ofrece una gran flexibilidad para aplicar las diferentes técnicas que se han propuesto, y suficiente capacidad de procesamiento y memoria para hacer frente a las aplicaciones propuestas en este trabajo. Además, su estructura hardware encaja con la arquitectura general de nodo de una WBSN, por lo tanto, todos los resultados presentados en esta tesis se pueden extrapolar a otras plataformas.

Por otra parte, este trabajo pretende mejorar la eficiencia energética de la WBSN, cumpliendo con los requisitos de rendimiento impuestos por la aplicación. Para lograr este objetivo, la Sección 0.3 presenta un protocolo MAC especialmente diseñado para cumplir con los requisitos de las WBSNs. Este protocolo MAC es una versión reducida del modo beacon-enabled del protocolo IEEE 802.15.4 que usa GTS. Se utiliza un enfoque de tipo TDMA, que ha demostrado ser ventajoso en WBSNs con topología de estrella, ya que su naturaleza centralizada elimina la aparición de colisiones y overhearing, mientras se usan muy pocos paquetes de control, reduciéndose drásticamente el idle listening. El uso de este protocolo MAC contribuye a mejorar la efi-

ciencia energética de la WBSN, ya que reduce significativamente el consumo de energía de la comunicación inalámbrica.

En la Sección 0.3 también se propone el uso de FreeRTOS, un OS muy ligero que proporciona soporte para operaciones en tiempo real y permite una gestión eficiente de los recursos hardware del nodo. Además, FreeRTOS realiza una planificación muy efectiva de las distintas tareas a ejecutar en el nodo, como algoritmos de procesamiento de señal, gestión del protocolo MAC, muestreo de la señal, así como la comunicación entre procesos. Finalmente, el uso de este OS permite desacoplar los recursos hardware y software del nodo, proporcionando una mayor flexibilidad al programador.

A continuación, la Sección 0.4 presenta varios algoritmos para realizar filtrado de ECG, delineación y diagnóstico de la arritmias en tiempo real, que dotan a los nodos de capacidad para evaluar el estado del corazón y detectar alteraciones cardiacas potencialmente peligrosas. Estos algoritmos, que han sido optimizados y portados al Shimmer™, también pueden contribuir a la reducción del consumo de energía del nodo, ya que reducen significativamente la cantidad de información que debe ser transmitida. Para lograr esta reducción del consumo de energía, hay que buscar un compromiso entre el uso de la CPU y la comunicación por radio.

En particular, la Sección 0.4 propone dos algoritmos de delineación de ECG single-lead en tiempo real, uno basado en la transformada wavelet y otro en la transformación morfológica multiescala, que ofrecen una excelente precisión en la detección de los principales puntos del ECG. Después, esta sección propone dos algoritmos de delineación multi-lead basados en la combinación de las derivaciones disponibles mediante RMS, que utilizan dos técnicas diferentes para la eliminación del ruido de línea de base: uno basado en cubic splines y otro en filtrado morfológico. La precisión de la delineación multi-lead es similar al mejor caso de delineación single-lead para la base de datos QTDB, y queda muy cerca de las tolerancias máximas propuestas por el CSE. Esto confirma que este enfoque multi-lead consigue que los resultados no estén condicionados a la elección de una determinada derivación o a la intervención de un experto que juzgue cuál es la derivación más adecuada para realizar la delineación. Además, las dos técnicas para la eliminación del ruido de línea de base muestran distintos compromisos entre predictividad positiva, por un lado, y complejidad computacional y uso de memoria por otro.

A continuación, la Sección 0.4 propone dos enfoques para el diagnóstico de arritmias en tiempo real. El primero, más general, utiliza un conjunto de reglas que tienen en cuenta la información morfológica y temporal extraída por los algoritmos de delineación de ECG propuestos. Para cada latido, el algoritmo evalúa si es normal o presenta alguna irregularidad que pueda indicar un comportamiento anormal del corazón. El segundo enfoque, mucho más preciso ya que está especialmente adaptado para detectar una arritmia

particular, consiste en un algoritmo de detección de AF que se basa en la observación de dos irregularidades características de los episodios de AF que se manifiestan en la señal ECG: la variabilidad del intervalo R-R y la ausencia de la onda P. Los resultados obtenidos por este sistema de detección de AF en tiempo real son muy similares a los de otras técnicas que realizan la detección de AF a posteriori sobre una señal grabada por un dispositivo de tipo Holter, alcanzando una sensibilidad del 96 % y un 93 % de especificidad para la AFDB.

La Sección 0.4 también presenta una evaluación del consumo de energía Shimmer™ para los distintos algoritmos de delineación y detección de arritmias propuestos, con respecto al caso base que simplemente realiza streaming del ECG. El desglose del consumo de energía de los bloques hardware del nodo demuestra lo siguiente:

1. Aunque el nodo realice procesamento avanzado de la señal ECG en tiempo real y por tanto el consumo de energia del microcontrolador aumente, la vida útil del sistema puede verse incrementada respecto al caso base que realiza streaming del ECG, ya que la cantidad de información que debe ser enviada al coordinador de la WBSN se reduce drásticamente.

2. Dado un protocolo MAC optimizado para WBSNs como el propuesto en la Sección 0.3 y una gestión eficiente de los distintos modos de trabajo ofrecidos por la radio, la delineación single-lead proporciona una mejora prácticamente despreciable de la vida útil del nodo con respecto a hacer streaming del ECG.

3. La vida útil del nodo se ve reducida si se usa alguna de las técnicas de delineación multi-lead propuestas, que conllevan un aumento de hasta 5 veces del uso de la CPU.

4. El muestreo del ECG es la tarea que demanda un mayor consumo de energía para la frecuencia de muestreo considerada.

En consecuencia, la exploración arquitectónica realizada en este trabajo sugiere que hay que hacer una elección cuidadosa de los componentes de la plataforma WBSN, teniendo en cuenta la aplicación final durante la fase de diseño. Esto puede permitir una gran disminución del consumo de energía del sistema, por lo tanto, se puede aumentar la vida útil del nodo siempre que:

1. Se use un conversor A/D de bajo consumo que permita racionalizar el consumo del muestreo del ECG.

2. El nodo esté equipado con un microcontrolador de bajo consumo que permita realizar procesamiento avanzado de la señal ECG sin que el

impacto en el consumo de energía total del nodo sea demasiado significativo.

3. Por último, pero no por ello menos importante, es fundamental realizar una correcta optimización de los algoritmos ejecutados en el nodo que tenga en cuenta los limitados recursos de procesamiento y memoria de los que este dispone.

0.5.1. Publicaciones

A continuación, he aquí una lista con las publicaciones producidas durante estos años de investigación:

Revistas

(1) F. Rincón, A. Dogan, N. Khaled, D. Atienza, A. Burg. "Energy-Aware Design of Biomedical Wireless Body Sensor Networks". *Integration, the VLSI Journal*. 2012, to appear.

(2) F. Rincón, J. Recas, N. Khaled, D. Atienza. "Development and Evaluation of Multi-Lead Wavelet-Based ECG Delineation Algorithms for Embedded Wireless Sensor Nodes". *IEEE Transactions on Transactions on Information Technology in BioMedicine (TITB)*. 2011. vol. 15, pp. 854-865.

Congresos

(1) I. Beretta, F. Rincón, N. Khaled, P.R. Grassi, V. Rana, D. Atienza. "Design Exploration of Energy-Performance Trade-Offs for Wireless Sensor Networks". In Proc. *IEEE/ACM Design Automation Conference (DAC)*. 2012.

(2) I. Beretta, F. Rincón, N. Khaled, P.R. Grassi, V. Rana, D. Atienza, D. Sciuto. "Model-Based Design for Wireless Body Sensor Network Nodes". In Proc. *IEEE Latin American Test Workshop (LATW)*. 2012.

(3) I. Bogdanova, F. Rincón, D. Atienza. "A Multi-lead ECG Classification Based on Random Projection Features". In Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2012.

(4) F. Rincón, N. Boichat, V. Barbero, N. Khaled, D. Atienza. "Multi-Lead Wavelet-Based ECG Delineation on a Wearable Embedded Sensor Platform". In Proc. *Computers in Cardiology (CinC)*. 2009.

(5) F. Rincón, N. Boichat, N. Khaled, D. Atienza. "ECG Delineation on a Commercial Wearable Embedded Sensor Platform". In Proc. *International Workshop on Wearable Micro and Nano Technologies for Personalized Health (pHealth)*. 2009.

(6) N. Boichat, N. Khaled, F. Rincón, and D. Atienza. "Wavelet-Based ECG Delineation on a Wearable Embedded Sensor Platform". In Proc. *IEEE International Workshop on Body Sensor Networks (BSN)*. 2009.

(7) F. Rincón, L. Gutiérrez, M. Jiménez, V. Díaz, N. Khaled, D. Atienza, M. Sánchez-Élez, J. Recas, G. de Micheli. "Implementation of an automated ECG-based diagnosis for a wireless body sensor platform". In Proc. *International Conference on Biomedical Electronics and Devices (BIODEVICES)*. 2009.

(8) F. Rincón, M. Paselli, J. Recas, Q. Zhao, M. Sánchez-Élez, D. Atienza, J. Penders, G. de Micheli. "OS-Based Sensor Node Platform and Energy Estimation Model for Health-Care Wireless Sensor Networks". In Proc. *Design, Automation and Test in Europe (DATE)*. 2008.

(9) F. Rincón, A. Susu, M. Sánchez-Élez, D. Atienza, G. de Micheli. "A Simulation Model for Wireless Sensor Networks Based on TOSSIM". In Proc. *Conference on Design of Circuits and Integrated Systems (DCIS)*. 2007.

(10) F. Rincón, F. Rivera, M. Sánchez-Élez, R. Hermida. "2D mesh NoC simulation tool using an FPGA based environment". In Proc. *Conference on Design of Circuits and Integrated Systems (DCIS)*. 2006.

(11) F. Rincón, F. Rivera, M. Sánchez-Élez, R. Hermida. "Herramienta de simulación de una NoC con topología de malla 2D usando un entorno basado en FPGA". In Proc. *Jornadas de Paralelismo*. 2006.

Chapter 1

Introduction

Know thyself.
Ancient Greek aphorism

Wireless Sensor Networks [CES04] (WSN) are wireless networks consisting of spatially distributed autonomous sensors to cooperatively monitor physical magnitudes. Their utility ranges from environmental measurements at different locations to medical monitoring. The WSN main workload entails sensing relevant information by the nodes, collecting that information, interpreting it, recording it and communicating it to a base station. In order to perform those activities each sensor node is made up for some sensors, a radio transceiver and a microcontroller. As power source they usually have a battery, but energy harvesting modules (like solar cells or thermoelectric converters) can also be included, increasing the life of the node and therefore minimizing its maintenance [PS05].

Figure 1.1 shows the architecture of a typical WSN. This network is composed of many *sensor nodes* that sense different parameters and send the collected information to a *mesh node*. Each sensor node can access one or more mesh nodes either directly or through other sensor nodes of the network following a multi-hop approach. Mesh nodes can be connected to each other to develop a higher level network, which can be accessed from a *management station* using a direct wireless connection to one of the mesh nodes or via Internet.

The type of sensors that can be used in a node is not restricted, so WSN applications are many and varied. They are used in commercial and industrial applications to monitor data that would be difficult or expensive to monitor using wired sensors. They could be deployed in wilderness areas, where they would remain for many years (monitoring some environmental variables) without the need to recharge/replace their power supplies. Some of the important applications of WSNs are:

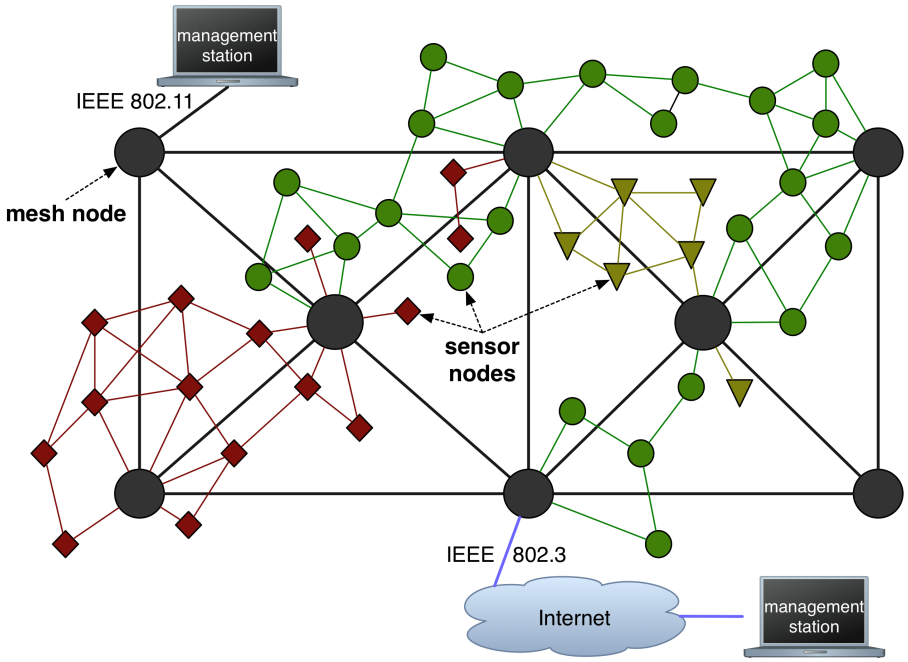

Figure 1.1: Typical WSN architecture

- Biologic research: non-invasive habitat monitoring and study of wildlife species/populations.
- Environmental monitoring: pollution (air, water, soil), early alarms for forest fires, volcanoes, earthquakes, and other natural hazards.
- Disaster response: emergency support for disaster recovery scenarios to identify risks and hazards and locating people/survivors.
- Agriculture: livestock and crops monitoring, microclimate management for dairy production or wines and soil fertility analysis.
- Industrial sector: manufacture process automation, monitoring and control, equipment failure prediction and production quality assurance.
- Retail sector: stock management, product tracking and quality monitoring.
- Architecture: smart office spaces, home automation, intrusion detection.
- Transportation: monitoring of internal systems in cars, ships and aircrafts.
- Local authorities: road traffic reporting, analysis and coordination.

- Military: surveillance, enemy identification, target acquisition and tracking and logistic operations support.
- Healthcare: staff and patient tracking, on-the-body vital signs monitoring and drug administration control.

This work is focused on the study of a subset of WSNs for healthcare applications, i.e., Wireless Body Sensor Networks (WBSNs), which have a tremendous potential to transform how people interact with and benefit from information technology [HJB+09]. As a case study, and without loss of generality, this work is focused on the analysis of electrocardiogram (ECG) signals as an example of very relevant bio-signal to study a plethora of possible cardiovascular diseases. In the next sections, the potential advantages of WBSN technology and its application to ECG monitoring are discussed more in detail.

1.1. Wireless Body Sensor Networks

An aging population and sedentary lifestyle are fueling the prevalence of chronic diseases such as cardiovascular diseases, hypertension, and diabetes. According to the World Health Organization, cardiovascular disease causes 30 % of all deaths in the world (about 17.5 million people in 2005). Diabetes currently affects 180 million people worldwide and is expected to affect around 360 million by 2030. More than 2.3 billion people will be overweight by 2015. A rapid rise in debilitating neuro-degenerative diseases such as Parkinson's and Alzheimer's is threatening millions more [PW10]. In summary, non-communicable diseases are today involved in 63 % of all deaths worldwide, and are predicted to cost 75 % of the current GDP by 2030.

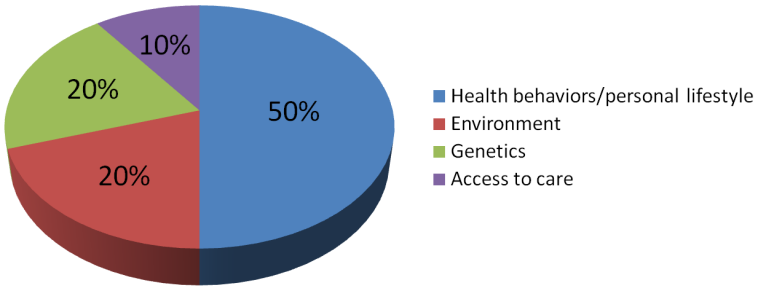

Figure 1.2: Determinants of health (source: Institute for the future, Center for disease control and prevention, 2006)

Burgeoning healthcare needs are exerting an enormous strain on the fragile healthcare delivery infrastructure. Moreover, a shortage of skilled staff, overload, and tightening of budgets have aggravated the impending healthcare crisis. These economical, social, and demographic trends highlight the

need to exploit technological breakthroughs to bring affordable and efficient solutions to people that will improve their quality of life. Moreover, as depicted in Figure 1.4, 50 % of the determinants of health are related to health behaviors and personal lifestyle. These changes call for a paradigm shift in healthcare delivery, where wearable personal health IT systems are key for the advent of personalized preventive healthcare. These systems are able to collect unprecedented longitudinal patient data and improve the effectiveness of therapeutic intervention and the science of health promotion. They can also support clinical decision making and provide a convenient, cost-effective care delivery channel.

WBSNs for health monitoring and diagnosis, as well as emergency detection, are gaining popularity and will deeply change healthcare delivery in the next years. The use of these networks enables continuous biomedical monitoring and care, helping to the prevention and early diagnosis of diseases, while enhancing the patient's autonomy with respect to more invasive and bulky state-of-the-art technologies for healthcare monitoring. However, the adoption of WBSN technology must overcome formidable technical and social challenges (e.g., form factor, battery life, reliability, safety, security, privacy, interoperability, ease of use, etc.). Although WBSNs share some of these challenges with general WSNs, many WBSN-specific research and design questions that require new lines of inquiry have emerged. Therefore this kind of networks deserve a separate analysis, since the nature of the sensed signals, the type of signal processing algorithms and the inherent limitations of healthcare applications make them very different from the rest of WSNs.

Unlike generic WSNs, which are composed of a high number of nodes spread over a big area and performing the same task, WBSNs have a few nodes to cover a human's body (very often less than 10), with each node dedicated to a specific task. For example, a sensor node monitoring acceleration at the ankle for gait analysis clearly cannot also measure the electroencephalogram (EEG) since both the location and sensing hardware are so different.

From the point of view of the data acquisition, the signals commonly used in WBSN applications such as EEG, ECG or acceleration data have quite high-frequency components, which entails the use of high sampling rates to acquire them, typically in the range from 100 Hz to 1 kHz. This leads to the generation of a huge amount of information in each node compared to generic WSNs, that needs to be stored, processed or transmitted to a WBSN coordinator.

To achieve a widespread adoption, WBSN nodes must be extremely non-invasive, which means that the nodes must have a small form factor that makes them comfortable to use and not impair normal activities. Smaller nodes imply smaller batteries, creating strict tradeoffs between WBSN node energy consumption and the fidelity, throughput, and latency requirements of WBSN applications. This thesis studies these tradeoffs to improve the

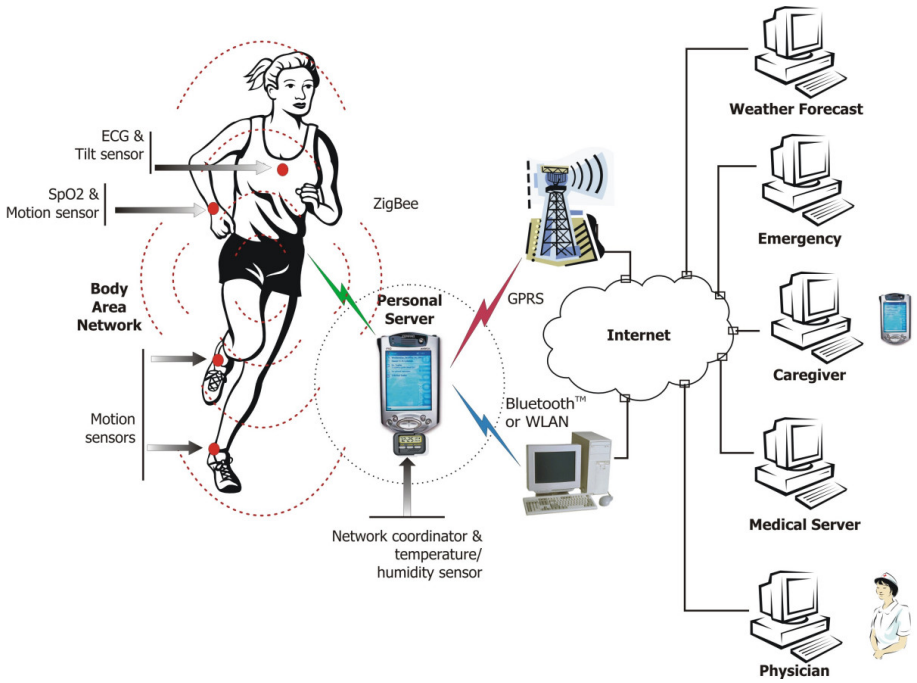

Figure 1.3: WBSN architecture

energy efficiency of the WBSN, and therefore improve its lifetime, while exploiting the limited processing capabilities and memory resources of the nodes to achieve smart WBSNs that are able to provide heath information of the patient in real time.

WBSN technology can be used to monitor many heterogeneous vital signs, giving raise to a wide application domain. In the literature we can find a great variety of WBSN applications that use one or more of these vital signs (multi-parametric monitoring) to track every facet of life, from sleep to pain. For example, Grundlehner et al. [GBPG09] introduce a WBSN for real-time arousal monitoring based on the ECG, respiration, skin conductance and skin temperature. These vital signs are collected by a WBSN and transmitted to a central PC, that online processes the information to provide real-time arousal evaluation. Abbate et al. [AAC+11] present a WBSN system for fall detection based on accelerometry data. They propose an approach that filters "activities of daily-living" from the data sensed by the accelerometers, which significantly improves the performance of the fall detection, achieving 100 % sensitivity and specificity.

This thesis considers as a case study a WBSN for cardiac monitoring that performs an automatic analysis of the ECG signal. Traditionally, this analysis was either taking place online on bulky, high-performance bedside cardiac monitors, or performed offline during a postprocessing stage after ambulatory ECG recording using wearable, yet obtrusive, ECG data loggers

(Holter devices). Recently, a significant industrial and academic effort has been dedicated to build miniature, wearable and wireless ECG monitors as an enabler of next-generation mobile cardiology systems. However, these systems are mainly limited to streaming the raw ECG data, leading to a high energy consumption that reduces their autonomy to few days. The aim of this work is to design a miniaturized long-lived wireless Holter monitor with embedded real-time diagnosis capability.

1.2. Cardiac monitoring

The choice of ECG is mainly motivated by the deep and detailed knowledge of the medical community about this signal, which reflects cardiac disturbances that can be diagnoses using many of the algorithms already available in the literature. Moreover, as previously mentioned, cardiovascular diseases are the major cause of mortality worldwide, responsible for 30 % of deaths worldwide in 2005 and economic fallout in billions of Dollars [PW10]. Their burden is only expected to rise due to, on the one hand, the fast aging of the world population, and on the other hand, the increasing prevalence of unhealthy lifestyles. WBSN technologies are poised to offer large-scale and cost-effective solutions to decrease the high healthcare costs and medical management needs of ECG analysis and diagnosis that, otherwise, are expected to be unsustainable for traditional healthcare delivery systems in the next years. Nevertheless, the results obtained in this thesis using ECG monitoring as a case study can be extended without loss of generality to other applications that use different biometric signals, since they all have very similar requirements.

Regarding ECG, a plethora of research projects started in recent years: MyHeart [LSD+06], Health@Home [SSS+10], MobiHealth [vHBW+04], CodeBlue [KKG07], etc. These projects propose WBSNs as a solution to monitor real-time patients' vital activities and collecting data for further processing. In addition, the development of two commercial products and a research prototype can be highlighted: Toumaz's Sensium Life Pebble [Tec09], a CE-certified ultra-small and ultra-low-power monitor for single-lead ECG, heart rate (HR), physical activity and skin temperature measurements with a reported autonomy of 5 days on a hearing aid battery; Corventis's PiiX [Cor09], a CE and FDA-cleared lead-less band-aid-like ECG sensor able to perform continuous arrhythmia detection based on HR measurements; and finally IMEC's prototype of a single-lead bipolar ECG patch [YTP+09] for ambulatory HR monitoring with a claimed 10-day autonomy on a 160mAh Li-ion battery. Accordingly, state-of-the-art unobtrusive wireless mobile/ambulatory ECG monitors are single lead and limited to streaming of the whole ECG signal to a coordinator or simple embedded HR measurement and analysis. However, this thesis advocates for an advanced real-time automated multi-

lead ECG diagnosis system. This long-lived smart wireless Holter requires a major breakthrough in the energy efficiency of ECG monitors. To this end, this thesis proposes to reduce the amount of streamed data by means of embedding the automated ECG signal analysis without compromising the system lifetime.

The ECG signal that the proposed smart Holter monitor captures and analyzes describes the electrical activity of the heart recorded by electrodes placed on the body surface [SL05]. The voltage variations measured by the electrodes are caused by the action potentials of the excitable cardiac cells as they make the cells contract. The resulting heartbeat in the ECG is manifested by a series of waves whose morphology and timing convey information which is used for diagnosing diseases that are reflected by disturbances of the heart's electrical activity. The time pattern that characterizes the occurrence of successive heartbeats is also very important.

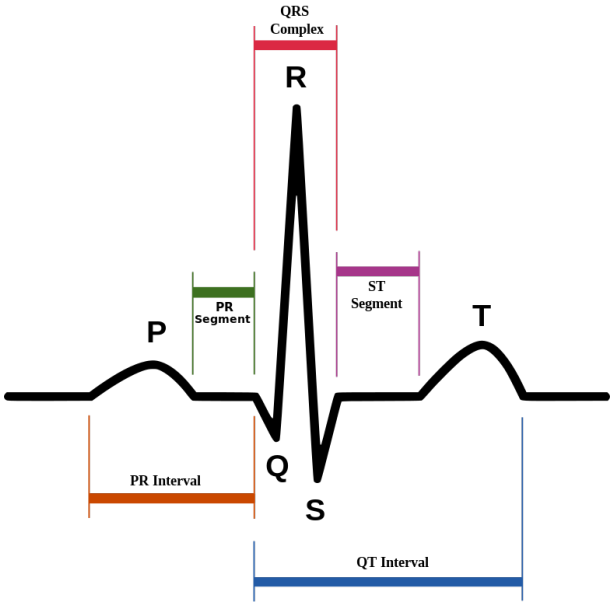

Figure 1.4: ECG beat showing the QRS complex, and the P and T waves.

The ECG is roughly a periodic signal, and each beat is composed of a QRS complex, preceded by a P wave, and followed by a T wave (see Figure 1.4), which correspond with the sequence of mechanical events that defines a cardiac cycle. This cardiac cycle can be assumed to start in the right atrium where blood is collected from all the veins in the body except those of the lungs. When the right atrium is triggered to contract, it forces blood into the right ventricle. When the right ventricle has been filled, it contracts and forces blood into the lungs, where the excess carbon dioxide is replaced by oxygen. The pulmonary veins return the oxygenated blood to the left atrium which in turn empties into the left ventricle. In its capacity as

a high-pressure pump, the left ventricle forces blood to all of the body organs and tissues (except the lungs) through the arterial vessels which evolve into capillaries and, finally, into the venous return system of the heart.

Each cardiac cycle is composed of two phases, activation and recovery, which are referred to in electrical terms as depolarization and repolarization and in mechanical terms as contraction and relaxation. Depolarization is manifested by a rapid change in the membrane potential of the cell (from -90 to 20 mV in approximately 1 ms) and constitutes the initial phase of the cardiac action potential. The rapid change in voltage causes neighboring cells to depolarize, and, as a result, an electrical impulse spreads from cell to cell throughout the myocardium. Depolarization is immediately followed by repolarization during which the membrane potential of the cells gradually returns to its resting state. Atrial depolarization is reflected by the P wave, and ventricular depolarization is reflected by the QRS complex, whereas the T wave reflects ventricular repolarization. Atrial repolarization cannot usually be discerned from the ECG, since it coincides with the much larger QRS complex.

The P wave reflects the sequential depolarization of the right and left atria. In most leads, the P wave has positive polarity and a smooth, monophasic morphology. Its amplitude is normally less than 300 μV, and its duration is less than 120 ms. An absent P wave may, for example, suggest that the rhythm has its origin in the ventricles, causing atrial depolarization to coincide with ventricular depolarization.

The QRS complex reflects depolarization of the right and left ventricles which in the normal heart lasts for about 70-110 ms. It has the largest amplitude of the ECG waveforms, sometimes reaching 2-3 mV. The first negative deflection of the QRS complex is denoted the Q wave, and the first positive is denoted the R wave, while the negative deflection subsequent to the R wave is denoted the S wave (Figure 1.4). Although the QRS complex may be composed of less than three individual waves, it is nevertheless referred to as a QRS complex.

The ST segment is not really a wave, but represents the interval during which the ventricles remain in an active, depolarized state. The ST segment begins at the end of the S wave (the J point) from where it proceeds nearly horizontally until it curves into the T wave (Figure 1.4). Changes in the ST segment, which make it either more elevated, depressed, or more steeply sloped, often indicate various underlying cardiac conditions.

The T wave reflects ventricular repolarization and extends about 300 ms after the QRS complex. The position of the T wave is strongly dependent on heart rate, becoming narrower and closer to the QRS complex at rapid rates; this "contraction" property does not apply to the P wave or the QRS complex. The normal T wave has a smooth, rounded morphology which, in most leads, is associated with a single positive peak.

The R-R interval represents the length of a ventricular cardiac cycle, measured between two successive R waves, and serves as an indicator of ventricular rate. The R-R interval is the fundamental rhythm quantity in any type of ECG interpretation and is used to characterize different arrhythmias as well as to study heart rate variability.

The PR interval is the time interval from the onset of atrial depolarization to the onset of ventricular depolarization. Accordingly, the PR interval reflects the time required for the electrical impulse to propagate from its origin, where the cardiac cycle is initialized, to the ventricles. The length of the PR interval is weakly dependent on heart rate.

The QT interval represents the time from the onset of ventricular depolarization to the completion of ventricular repolarization. This interval normally varies with heart rate and becomes shorter at more rapid rates. Prolongation of the QT interval has been observed in various cardiac disorders associated with increased risk of sudden death.

1.3. Motivation and challenges

In 2009, the prestigious Wired magazine published an article about how wireless sensor technology can help us to track and analyze our vital signs, exercise, nutrition and health to optimize our lives. A fragment of the article says: "a data-driven health revolution promises to make us better, faster, and stronger". WBSNs are poised to be technology to enable this revolution, since they can provide long-term non-invasive monitoring of our vital signs and habits to give us information about our health status. Then, the first challenge of this thesis is to develop an ECG monitoring system that not only continuously acquires and displays the ECG signal of the patient, but it also features on-board real-time analysis and diagnosis, instantly providing information about the operation of the heart. This information is accessible to the patient through a hand-held device, such as a mobile phone, and it can also be sent to the medical staff in real-time. If an abnormal event happens, the system will automatically detect it and the patient and the doctors will be immediately informed, making possible a very quick reaction in case of an emergency. To achieve the development of such system, several algorithms are proposed and optimized to perform on-board real-time ECG signal filtering, characteristic points detection and arrhythmia classification.

The second main challenge of this thesis is the improvement of the WBSN energy efficiency, while meeting the data fidelity requirements of the application. The battery-size versus -life tradeoff plays a major role in defining any WBSN system, and applying design techniques to reduce energy consumption can improve both size and lifetime. If the energy consumption can be reduced far enough, perpetual operation on harvested energy becomes a possibility. Thus, WBSN node sensing, processing, storage, and wireless

transmission must all be done in a way that reliably delivers the important data but with the lowest possible energy consumption, thus minimizing battery size (which dominates WBSN node form factor) and maximizing time between battery recharge (which is a key factor in wearability), both of which can impact the performance and practicality of possible applications. Hence, to accomplish an improvement in the energy efficiency of the proposed real-time ECG monitoring system, several solutions are studied at the level of the medium access control (MAC) protocol and the operating system (OS).

1.4. Thesis outline

The rest of the thesis is organized as follows: chapter 2 reviews some of the available state-of-the-art WBSN architectures and provides a detailed description of the one used in this thesis, the Shimmer™ platform. Chapter 3 presents optimizations at the level of the MAC protocol and the OS to improve the energy efficiency of the platform. Besides, it provides a characterization of the Shimmer™'s energy consumption using the proposed low-level optimizations. Chapter 4 describes a set of application-level optimizations that consist of several algorithms for ECG filtering, delineation and arrhythmia detection. In addition to this, it shows a detailed study of the accuracy, execution time and memory usage, as well as a comprehensive evaluation of the energy consumption entailed by the considered algorithms. Our concluding remarks will be given in chapter 5, as well as the contributions of this work and future lines of work.

Chapter 2

Case study architecture evaluation

Nothing in life is to be feared,
it is only to be understood.
Now is the time to understand more,
so that we may fear less.

Marie Curie

2.1. WBSN architectures: state of the art

This work considers the typical WBSN architecture, where a software application for data processing is executed on a microcontroller, and the remaining services are delegated to an operating system. Figure 2.1 provides a structural outlook of this class of nodes. We assume that a node transmits its data to a central coordinator through the typical star topology used in WBSNs. Each node in the network generates a constant traffic, thus avoiding data bursts that may interfere with the periodic transmission of the other nodes in the network.

The *sensor* component describes the hardware that conditions and samples the raw signal at a frequency that depends on the signal and on constraints such as the Nyquist-Shannon theorem. Then, the samples are quantized by an A/D converter using a number of bits that depends on its resolution.

The *applications* components comprise all the software programs used to process the sensed data, including filtering, feature extraction, compression and aggregation.

The *operating system* (OS) provides services, such as, the interaction with the hardware, the software-level management of the sensing, the memory and the radio transmission. The OS also manages a set of queues, including the

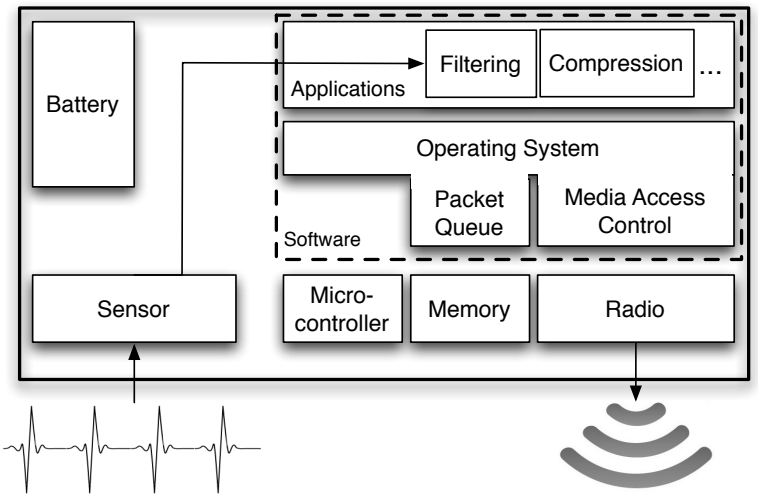

Figure 2.1: Block diagram of the reference node architecture

ones for interprocess communication, and the one containing the packets to be transmitted. Furthermore, the OS implements a MAC protocol to share the access to the wireless medium among the nodes in the star-topology network.

The *microcontroller* (μC) is the component of the platform in charge of executing the OS and the software applications. Depending on the hardware, techniques such as dynamic voltage scaling [GC97] might be available to allow the microcontroller to be active for a limited time (or duty cycle), and to switch to a low-power state when there is no task to be executed.

The *memory* bank is used to store the data required by the applications and the OS. Although a larger memory exhibits a higher energy consumption [KOKE06], a limited size may affect the capacity of the internal queues, and hence the performance of the applications and the throughput of the transmission.

Finally, the *radio* component describes the hardware used to modulate and transmit the data through the wireless channel. Depending on the characteristics of the platform, the wireless transmission power and the modulation scheme can be adjusted, even dynamically [SRS03], to determine the distance that is covered with a predetermined packet error rate (PER).

With the proliferation of WBSN applications, many platforms whose architecture corresponds to the one depicted in Figure 2.1 have appeared during the last years. The first critical choice when we face the development of a WBSN is the selection of the most appropriate platform, taking into account the requirements of the application. In this work, the different platforms are going to be analyzed from the point of view of their flexibility and efficiency. In this sense, we can divide them into two main groups: *application-specific*

integrated circuits (ASICs) and *commercial off-the-shelf (COTS)* platforms.

2.1.1. Application-Specific Integrated Circuits

The use of ASICs is always more efficient in terms of energy consumption and size, since they are especially designed for a very specific application. However, these devices involve a great design effort and offer a very limited flexibility. They can only perform a predefined specific sequence of operations due to their hardwired nature, since they are tailored for a particular application.

In the literature, many energy-efficient ASIC implementations are available for digital signal processing in health monitoring systems, performing heart rate monitoring and extraction of ECG characteristics. For instance, Massagram et al. [MHM+10] introduced an ASIC designed for digital heart rate variability (HRV) parameter monitoring and assessment. This ASIC chip measures R-R intervals and stores HRV parameters into its internal memory in real time. The prototype chip was fabricated in a 0.5 μm complementary metal-oxide semiconductor technology on a 3×3 die area, with a measured dynamic power consumption of 10 μW and measured leakage current of 2.62 nA. The HRV monitoring system including this HRV ASIC, an analog-to-digital converter, and a low complexity microcontroller was estimated to consume 32.5 μW, which is seven times lower power than a stand-alone microcontroller from Microchip performing the same functions.

Similarly, in another study, Hui-Min et al. [HMYLH+10] introduced an ASIC platform to extract QRS complexes from ECG signals. The detection is based upon digital analysis of slope, amplitude and width, reaching 99.36 % of accuracy. This ASIC features only 2.21 μW power consumption and 0.68 mm^2 core area.

In [WC04], a Fast Fourier Transform (FFT) processor with programmable FFT length and bit precision was proposed. The FFT processor, fabricated in a standard 0.18 μm CMOS process, operates in sub-threshold voltage region. At the minimum supply voltage (180 mV), where the circuit is still functional, it dissipates only 90 nW.

Moreover, [KSR03] presents an ultra-low-power, delayed least mean square (DLMS) adaptive filter operating in the subthreshold region for hearing aid applications. Simulation results show that the DLMS adaptive filter can operate at 22 kHz using a 400 mV supply voltage. To validate the robust operation of subthreshold logics, a 0.35 μm, 23.1 kHz, 21.4 nW, 8×8 carry save array multiplier test chip was fabricated. The test chip showed stable operation at a supply voltage of 0.30 V.

2.1.2. Commercial Off-The-Shelf platforms

Even though ASIC platforms achieve significant energy efficiency, their usage are very limited due to their rigid, not reprogrammable nature. COTS are in general less energy efficient and have a bigger size, but on the other hand they provide more flexibility, enabling the execution of more heterogeneous tasks. Many times, COTS are a good solution since they can provide reasonable size, performance and lifetime for many WBSN applications, while having a very low production cost.

COTS platorms normally use general purpose processors (GPPs), which offer high flexibility for embedded biomedical signal processing. For instance, Jocke et al. [JBW+09] present a 0.13 μm CMOS sub-threshold mixed-signal system-on-chip (SoC) that acquires and processes an ECG signal for wireless ECG monitoring. This SoC uses a sub-threshold digital microcontroller to process the acquired ECG data, proving that biomedical signal processing can be performed on WBSN nodes by relatively simple general purpose low power microcontrollers. Moreover, many commercially available WBSN nodes include a well-established general purpose low-power processor as a digital processing platform on the node. For example, TelosB [Tecb], Shimmer™ [BGM+10] and Tinynode [SA] are equipped with the MSP430 microcontroller from Texas Instruments [Insc] and MICAz [Teca] includes the ATmega128L low-power microcontroller from Atmel corporation. However, these commercially available low-power microcontrollers do not exploit very sophisticated energy-efficiency techniques. A custom implementation of these well-established instruction set architectures (ISAs) can provide a higher energy efficiency compared to commercially available low-power microcontrollers. For example, Zhang et al. [ZSB+11] show that a custom implementation of a well-established ISA achieves 100x higher energy efficiency per instruction with respect to the MSP430. However, custom implementations generally lack of built-in units such as clock generator, voltage regulator as well as compatible peripherals that reduce their flexibility compared to commercially available microcontrollers.

2.1.3. Application-Specific Instruction-set Processors

A balanced choice between hardware flexibility and energy efficiency for online biomedical signal processing is an *application-specific instruction-set processor (ASIP)*. ASIP systems offer hardware flexibility for a specific application group while achieving a higher energy efficiency compared to general purpose low-power microcontrollers, as their instruction set is reduced and customized for the corresponding application group.

The reduced and customized instruction set leads to a significant energy saving. For instance, in [YKH+09], a general purpose processor architecture was modified to have several customized instructions for a cardiac beat de-

tector algorithm. These modifications include a combined instruction which executes two load and one multiply-accumulate operation in parallel, leading to a 81 % cycle count reduction and 55 % energy saving compared to the general purpose processor for the beat detector algorithm. In addition to instruction set modifications, having dedicated hardwares for common operations of targeted application groups can improve the energy efficiency. These dedicated hardware blocks can be implemented either as an internal functional unit of an ASIP processor like an arithmetic logic unit (ALU) or as a peripheral attached to the processor. In [KC11], a biomedical signal processing platform with a GPP and several programmable hardware accelerators, attached for common operations of biomedical signal processing algorithms such as FIR and median filters, a coordinate rotation digital computer engine and a FFT module, was introduced. The authors show that programmable accelerators perform signal processing tasks with up to 215x lower energy consumption compared to the GPP. Moreover, the GPP with the accelerators achieves up to 91 % energy savings compared the GPP-only implementation.

2.2. The Shimmer™ platform

Considering the previously discussed advantages and drawbacks of each type of architecture available for WBSN nodes, I decided to use the Shimmer™ platform as representative target architecture case study of WBSN. The choice of this platform was mainly motivated by the following reasons:

- High flexibility, due to the use of the MSP430 general purpose microcontroller. This work is not only focused on the biomedical signal processing, but also analyzes and proposes optimizations in other levels of the system's architecture (OS, communication protocol, etc.). For this reason ASIC and ASIP implementations were discarded, seeking to have an instruction set that is a flexible as possible.

- The computational and memory resources of this platform, as well as its energy supply, offer a representative set of resources of WBSNs for the considered applications in this work. Moreover, the use of COTS components reduce the cost of the platform compared to ASICs or ASIPs.

- Software and firmware support, as though the MSP430 is not as low-power as other architectures, it is a very spread microcontroller used in a heterogeneous application domains. Therefore, there is a lot of support for it in terms of tools or ported code (OSs, MAC protocols, applications, etc.).

- Broad target application range, since there is a great number of accessories for the Shimmer™ platform, making it usable non only for ECG, but also for other biomedical (EEG, electromyogram, galvanic skin response, etc.) and non-biomedical (temperature, humidity, light, etc.) applications.

- It is a good representative of a WBSN platform. Therefore, all the results provided in this thesis, where the Shimmer™ has been used, are similarly applicable to other WBSN platforms.

2.2.1. Overall Shimmer™ architecture

Our target WBSN node is the Shimmer™ [BGM+10], a small wireless sensor platform designed by Intel for wearable health-sensing applications. This platform has the long term goal of facilitating research in independent living technologies, since it can be mounted in an MP3 player armband and the durable enclosure has a size of $54 \times 35 \times 18$ mm. Shimmer™ is equipped with the ultra-low-power TI MSP430F1611 [Insc] microcontroller, two radio chips (Bluetooth and IEEE 802.15.4-compliant), a 3-axis accelerometer and an expansion port used to connect a daughter board equipped with additional sensors (ECG, gyroscopes, etc.). It also provides a MicroSD™ Flash socket for up to 2 GB of additional storage, four light-emitting diodes (LED) for display purposes and an integrated 280 mAh Li-ion battery.

This sensor platform has been designed to be programed with the TinyOS [LMP+04] operating system and its specific drivers. Unfortunately, TinyOS shows several drawbacks, which will be discussed in Section 3.3.1.1, that make it unsuitable for the applications presented in this work. Due to these limitations, we ported FreeRTOS [biba], which is a portable, open source, hard real-time mini kernel that includes support for the TI MSP430 microcontroller and the CC2420 IEEE 802.15.4-compliant radio chip used by Shimmer™. Finally, I have used the open-source GCC 3.2.3 toolchain for the MSP430 [bibb] and Code Composer Essentials (CCE) v.3.1 [Insb], which use all the hardware resources of the target microcontroller, to compile the applications for the node.

2.2.2. Texas Instruments MSP430 microcontroller

The Texas Instruments MSP430F1611 [Insc] is one of the core components of the Shimmer™ baseboard and its primary advantages are its extremely low power during periods of inactivity and its proven history for medical sensing applications. The MSP430 is based on the 16-bit RISC CPU, peripherals and an adaptable clocking mechanism connected via a von-Neumann memory address bus (MAB) and memory data bus (MDB). The MSP430 offers a maximum frequency of 8 MHz and is optimized for high-level pro-

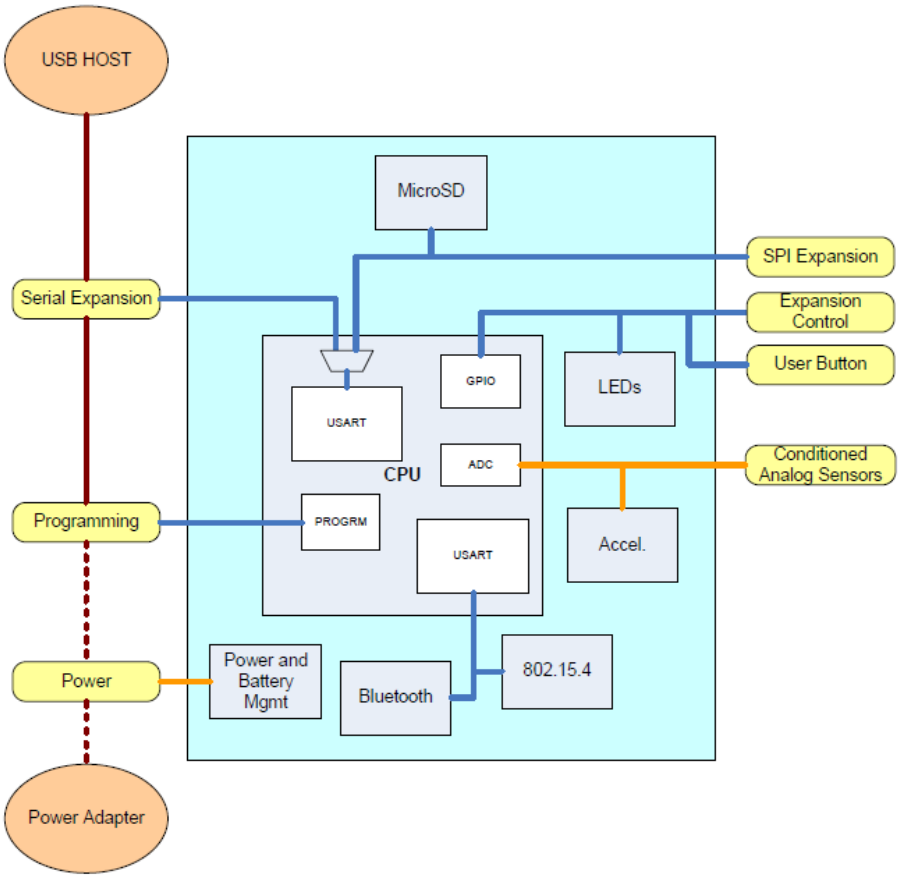

Figure 2.2: Simplified Shimmer™ system interconnections

gramming with only twenty seven core instructions and seven addressing modes. Figure 2.4 illustrates an architectural overview of the CPU, which shall be discussed under the following headings: clocking, memory, timers, USART and ADC.

2.2.2.1. Clocking

The MSP430 clocking mechanism is designed to support a battery supply with low power consumption by allowing the user to select from three inherent clock signals to gain optimum correlation between performance selection and power usage. The three available clocks are:

- Auxiliary Clock (ACLK), which is a low-frequency clock supplied by a 32 kHz crystal, and utilized for self wake-up and self-timing operation and low-power mode.
- Master Clock (MCLK), a high-speed clock which is fed from an internal

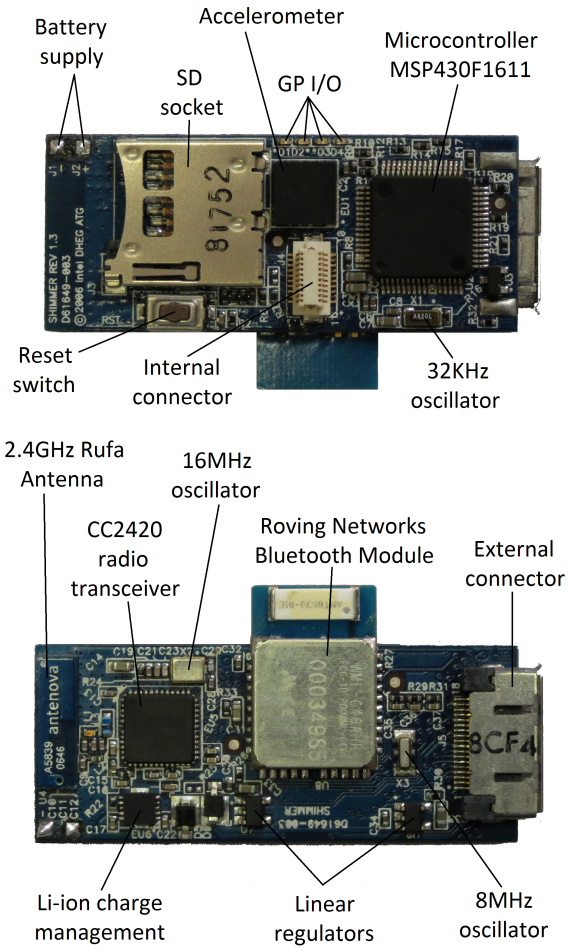

Figure 2.3: Photograph of Shimmer™ baseboard

digitally controlled oscillator (DCO), selected using interrupt requests. This is used for high-speed signal CPU and peripheral applications. The DCO, which may operate up to 8 MHz, can be activated from sleep mode in 6 μs and allows the CPU to be operated in very quick bursts.

- Sub-Main Clock (SMCLK), which is a clock designed to utilize the DCO or standard crystal oscillators as necessary for peripheral operation.

2.2.2.2. Memory

The MSP430 contains 10 kB of RAM, 48 kB of Flash and 128 B of information storage. The programming of the flash memory can be completed

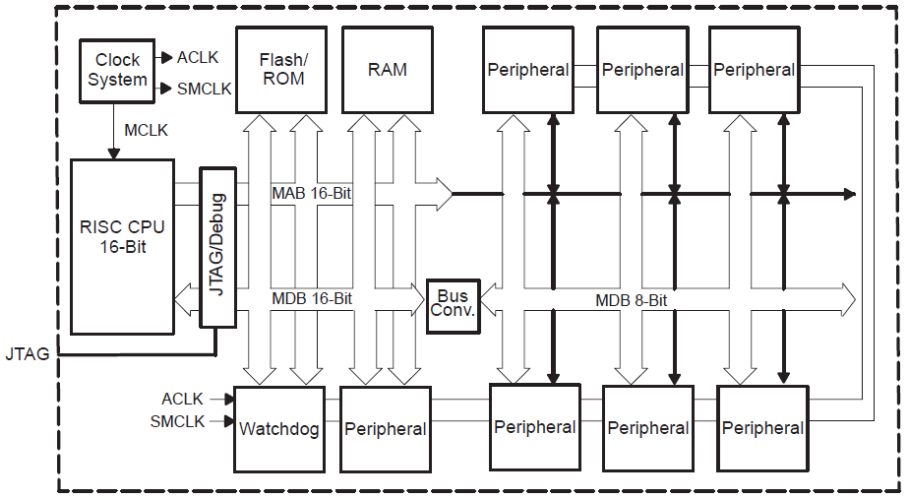

Figure 2.4: MSP430 architectural layout

in system using the CPU or via the bootstrap loader or JTAG port. The flash memory can be utilized for storing both the data and the code and the CPU has the ability to write to the flash memory via single-bytes and single-words.

2.2.2.3. Timers

The MSP430 microcontroller is equipped with two timers, Timer_A and Timer_B. Timer_A is a 16-bit asynchronous timer/counter which contains three capture/compare registers which allow for pulse-width modulation (PWM) outputs, interval timing and multiple capture/comparing. Timer_A also includes asynchronous input and output latching and the ability to generate interrupts depending on overflow conditions of the counter and capture/compare registers.

Timer_B is identical to Timer_A with the additional functionality of having a programmable length (8, 10, 12, or 16 bits), double-buffered and group-able registers and the ability to force the outputs into a high-impedance state. Timer_B however does not contain the synchronized capture/compare input bit function.

2.2.2.4. USART

The MSP430 contains two hardware-universal asynchronous/synchronous receive/transmit serial communication USARTs (USART0 and USART1), as illustrated in Figure 2.4, which enable SPI (3 or 4 pin) and asynchronous UART functionality. The USART0 also allows I2C transmissions and has two specific DMA channels to ensure maximum throughput with data rates up to 400 kbps. The signal transmission and reception is possible using

double-buffered channels.

Within the Shimmer™ platform, the MicroSD™ Flash Socket and 802.15.4 radio are connected to the microcontroller using the SPI mode, where serial data is transmitted using a shared master clock. This is activated when the I2C bit is cleared and the SYNC bit is set. The flow of the data transfer is possible using three or four pins; the slave in, master out (SIMO), slave out, master in (SOMI), USART SPI clock (UCLK) and slave transmit enable (STE) formats. In addition to master and slave modes 7-bit and 10-bit addressing are also supported. For low-power operation, receiver start-edge detection is incorporated for auto wake-up.

2.2.2.5. ADC

The MSP430 has 8 ADC channels for 12-bit A/D conversions using a 16-word conversion-and-control buffer which enables data to be read and stored without the need for CPU intervention. For the Shimmer™ platform the external ports are utilized for reading data from the XYZ accelerometer (3 channels), the internal expansion connector (3 channels) and the external expansion connector (2 channels). The ADC internal ports can be used to read data from the internal temperature sensor or read in the battery voltage. The maximum conversion rate of the ADC is greater than 200 ksps and the sampling period can be controlled by software or one of the internal timers. To maintain the low-power usage capabilities of Shimmer™ the MSP430 ADC core is disabled when not in use and re-enabled when necessary.

2.2.3. CC2420 radio transceiver

One of the key functions of the Shimmer™ board is its ability to communicate as a wireless platform. Although Shimmer™ has the dual functionality of having both 802.15.4 and Bluetooth radio modules, in this work we only consider the CC2420 [Insa] 802.15.4-compliant radio chip due to its low-power capabilities and flexibility that allow the implementation of power-saving techniques.

The CC2420 radio chip is suitable for WBSN, as it is designed for low-power and low-current applications. The radio may also be turned off by the MSP430 for low-power operation. The CC2420 radio has an inherent direct-sequence spread-spectrum modem and a theoretical data rate of 250 kbps. The CC2420 is controlled by an SPI connection over the USART1 and with the CC2420 having support for applications such as packet handling, data transmissions, data encryption, received signal strength, link quality and packet timing, the work load on the MSP430 microcontroller is reduced. The CC2420 requires only a few extra additional components such as a reference crystal oscillator and does not need any external filters.

2.2.4. ECG daughter board

The Shimmer™ platform provides a 20-pin internal expansion connector that allows the use of other add-on modules with additional sensors, such as ECG, galvanic skin response, electromyogram, gyroscope, magnetometer, GPS, etc. Within this work we only consider a daughter board that is able to acquire and condition 3-lead ECGs.

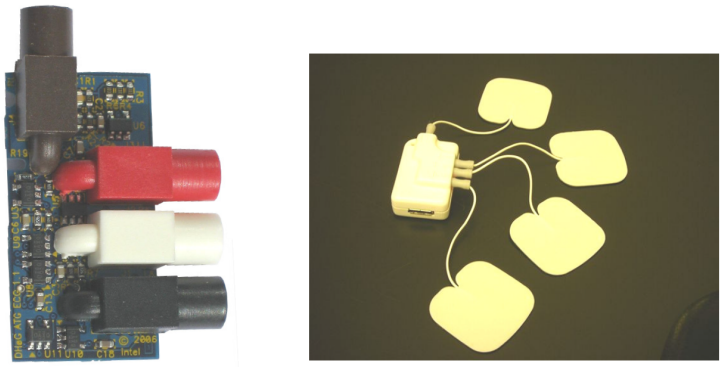

Figure 2.5: ECG add-on board and Shimmer™ with connecting sensor pads

This expansion board, as in Figure 2.5, has four connectors for four electrodes, which have to be placed in the right arm (RA), left arm (LA), right leg (RL) and left leg (LL) of the patient. The design uses CMOS operational amplifiers and produces two derivations, obtained by reading the voltage difference RA - LL (lead II) and LA - LL (lead III). The remaining standard derivation (lead I, RA - LA) can be calculated on the host CPU from leads II and III. RL electrode is actively driven with the average of the other three, to improve signal-to-noise, this type of connection is called a "Wilson's Central Terminal". This creates a negative-feedback loop where common-mode voltage on the body is driven to zero by the RL electrode.

In addition, lead inputs have weak pull ups to detect floating electrodes. The ECG front end was designed to provide fast baseline restoration for ambulatory applications and a gain of 175. It implements the following three filters:

- First-order low-pass filter at 724 Hz, at lead input.
- First-order low-pass filter at 159 Hz, at final gain stage.
- AC-coupling high-pass filter at 0.05 Hz to reduce motion artifacts.

2.2.5. Energy consumption estimation

In order to further justify the choice of the Shimmer™ platform and motivate the techniques for energy efficiency proposed in this thesis, this

section presents energy consumption results of a WBSN node in two different application scenarios: ECG streaming (a 2-channel ECG signal is acquired by a node placed on the body and transmitted to the WBSN coordinator) and Rpeak (application that detects when a heart beat occurs and only transmits the location of the R peaks to the WBSN coordinator).

These results [RPR+08] were obtained using a WBSN platform developed at IMEC Netherlands [PGdVT07], whose hardware architecture is very similar to the Shimmer™ node. This platform is capable of monitoring up to 24 EEG channels and 1 ECG channel by means of an ultra-low-power ASIC that features a constant power consumption 10.5 mW at 3.0 V. The processing capability is provided by a TI MSP430F149 low-power microcontroller. The wireless communication is based on the Nordic nRF2401, a low-power single-chip 2.4 GHz transceiver. This component consumes 10.5 mA at an output power of -5 dBm and 18 mA in receive mode.

Regarding the software, the node includes an embedded operating system, TinyOS, on top of which the entire software architecture is built, including the MAC protocol. This study proposes the use of a Time Division Multiple Access (TDMA) protocol, which allows several users to share the same frequency channel by dividing the timeline into different time-slots. Each node has a slot assigned, and can use it to exchange data with the WBSN coordinator. The timing of the protocol is regulated by the coordinator, which periodically broadcasts synchronization signals (beacons) to all the nodes.

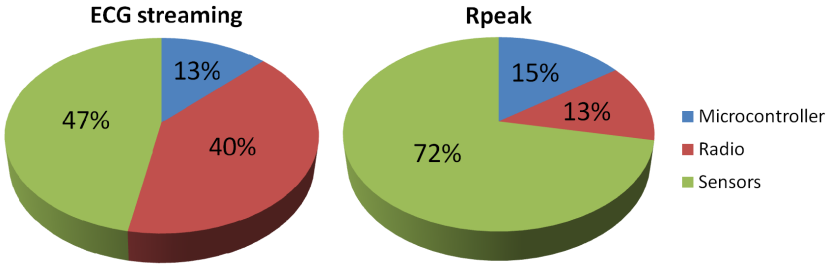

Figure 2.6: Energy consumption of the main components of a WBSN platform in two different application scenarios: ECG streaming and Rpeak

Using this hardware and software architectures, Figure 2.6 presents the energy consumption by hardware component for two different scenarios (ECG streaming and Rpeak application), assuming a sampling frequency of 200 Hz. This figure shows that the sensing part features a very high energy consumption, which can be decreased with the use of the Shimmer™ platform, since the power consumption is reduced to 6.6 mW. The radio is also an energy-hungry component, specially in the ECG delineation scenario, but its consumption can be decreased by using a more optimized MAC protocol and a more efficient OS. Moreover, Figure 2.6 also proves that the energy

consumption of the radio can be significantly reduced if the microcontroller of the node is used to perform ECG signal processing to decrease the amount of data that needs to be transmitted to the WBSN coordinator.

For these reasons, this thesis proposes several design techniques for smart and energy-efficient WBSNs, that will be detailed in the following chapters. The final goal is to design a long-lived smart wireless Holter with embedded real-time diagnosis capability.

Chapter 3

Communication-related optimizations for WBSNs

Everything should be made as simple as possible, but not simpler.

Albert Einstein

A WBSN node is equipped with a radio, which allows the device to send the collected and processed information to the WBSN coordinator or receive commands from it. The medium access control (MAC) protocol manages the access to the wireless channel and has to deal with several communication issues. The most important ones are the following:

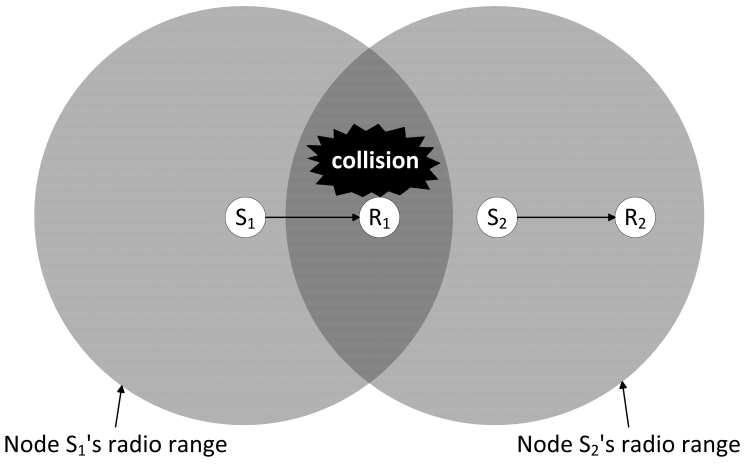

Figure 3.1: A collision happens when nodes S_1 and S_2 transmit a packet at the same time

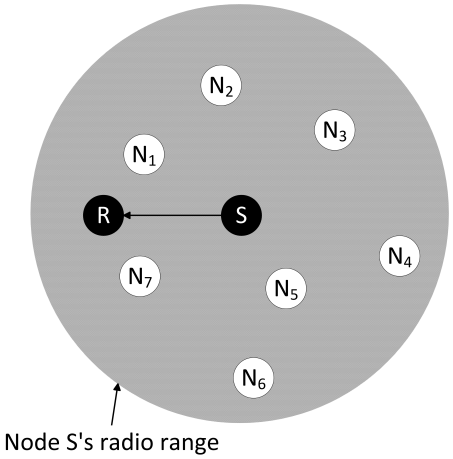

Figure 3.2: Overhearing: node S sends data to node R, but nodes N_1, N_2, ..., N_7 all receive such packet

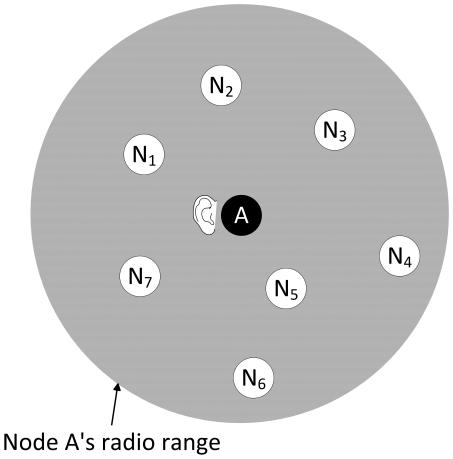

Figure 3.3: Idle listening: node A does not know when it will be receiver of a message from N_1, N_2, ..., or N_7

- Collisions. They happen when two or more devices access the wireless channel to send information at the same time (see Figure 3.1). This causes the corruption of the packets that are being sent, which have to be transmitted again.
- Overhearing. It occurs when a node A is sending data to a node B, and there is a node C that receives this packet (see Figure 3.2).
- Idle listening. It appears when a node A does not know when another node B of the network is going to send a packet to it (see Figure 3.3).

This causes that node A needs to keep the radio in receiving mode for a period of time when it is not receiving any information.

All these problems lead to an important increase in the energy consumption, which can be due to retransmission of the packets, reception of information that is not for the node or listening to the channel when no information is being sent to the node. A MAC protocol for WBSN devices has to consider all these issues and try to minimize their occurrence to reduce the energy consumption of the node as much as possible. Moreover, the MAC protocol involves the transmission of control packets like synchronization messages, acknowledgments to indicate that a packet has been correctly received, etc. whose transmission should also be minimized in order to further reduce the energy consumption of the radio.

3.1. State of the art: Medium Access Control protocols

Since the MAC protocol significantly influences the energy consumption of the WBSN, different solutions presented in the literature are going to be described in this section.

Centralized schedule-based MAC protocols, such as time division multiple access (TDMA), perfectly fit the requirements of WBSN networks, since typically all the nodes of the network transmit their data to a central coordinator through a star topology. The use of these protocols avoid collisions, since the central coordinator synchronizes the nodes of the network so that only one of them is accessing the wireless channel at a time. They also avoid overhearing, because a node switches its radio to receiving mode only when it is expecting some information. For example, during the time slot the WBSN coordinator uses to send a synchronization packet to all the nodes of the network or after sending a packet, in order to receive the acknowledgement the coordinator will send if the packet has been correctly received. For the same reason, idle listening does not occur either in this kind of protocols.

In this regard, H-MAC [LT10] aims to improve WBSNs energy efficiency by exploiting heartbeat rhythm information, instead of using periodic synchronization beacons, to perform time synchronization. Biosensors in a WBSN can extract the heartbeat rhythm from their own sensory data by detecting waveform peaks. All the rhythms represented by peak sequences are naturally synchronized since they are driven by a same source, the heartbeat. Following the rhythm, biosensors can achieve time synchronization without having to turn on their radio to receive periodic timing information from a central controller, so that energy cost for time synchronization can be completely avoided and the lifetime of network can be prolonged.

The reservation-based dynamic TDMA (DTDMA) [LLK09] focuses on

the dependability and power efficiency. In DTDMA, the slots are allocated by the WBSN coordinator only to the devices which have buffered packets and released to other devices after the current allocation. Through the adaptive allocation of the slots in a DTDMA frame, the coordinator adjusts the duty cycle adaptively with the traffic load. Comparing with the IEEE 802.15.4 MAC protocol, the DTDMA provides more dependability in terms of lower packet dropping rate and less energy consumption.

BodyMAC [FD09], which also shows superior performance than IEEE 802.15.4, uses flexible bandwidth allocation to improve node energy efficiency by reducing the possibility of packet collisions and by reducing radio transmission times, idle listening and control packets overhead. BodyMAC is based on a downlink and uplink scheme in which the contention free part in the uplink subframe is completely collision free. Three types of bandwidth allocation mechanisms allow for flexible and efficient data and control communications. An efficient sleep mode is introduced to reduce the idle listening duration, especially for low duty cycle nodes in the network.

Recently, the IEEE 802.15.6 task group [Net] has approved a draft of a standard for Body Area Network (BAN) technologies with the scope of short range, wireless communication in the vicinity of, or inside, a human body (but not limited to humans). The aim of the group is to create a protocol that can use existing ISM bands as well as frequency bands approved by national medical and/or regulatory authorities. The protocol needs to also provide support for Quality of Service (QoS), extremely low power, and data rates up to 10 Mbps while simultaneously complying with strict non-interference guidelines. This standard considers effects on portable antennas due to the presence of a person (i.e., varying with male, female, skinny, heavy, etc.), radiation pattern shaping to minimize the specific absorption rate (SAR) into the body, and changes in characteristics as a result of the user motions.

3.2. Proposed MAC protocol

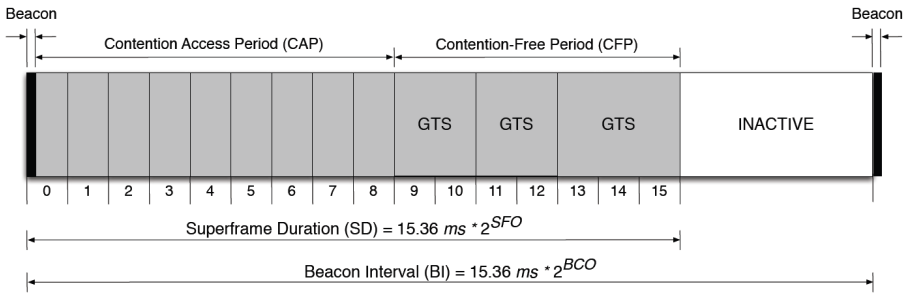

Figure 3.4: Structure of the IEEE 802.15.4 superframe

In order to fully exploit the low-power capabilities of this radio chip, and therefore reduce the energy consumption of the overall system, this work proposes a MAC protocol specially optimized to fit the requirements of the herein considered WBSN applications. Our target node implements a simple TCP/IP stack for the CC2420 radio instead of the IEEE 802.15.4 [Std03]. To optimize the energy consumption of the WBSN platform, we decided to implement a reduced version of the beacon-enabled mode of the IEEE 802.15.4 protocol using Guaranteed Time Slots (GTS) [Std03] over a star network (i.e., each sensor node can only send data to a designated WBSN coordinator/sink).

In this mode, the coordinator broadcasts periodic beacons, which are used by the rest of the nodes to synchronize their transmissions with their exclusively allocated GTS. These beacons have to be sent periodically every beacon interval (BI) = $15.36 \cdot 2^{BO}$ ms, where BO is the beacon order, $0 \leq BO \leq 14$. Following the beacon reception, an active period called superframe, whose structure is shown in Figure 3.4, is defined. The duration of the superframe (SD) is equal to $15.36 \cdot 2^{SO}$ ms, where SO is the superframe order, that has to be less or equal to BO. The SO has been set to 4, that leads to a SD of 245.76 ms divided in 16 equal time slots. In order to allow multiple sensor nodes in the network, a fixed two time slots per GTS have been assigned, which means 30.2 ms per GTS. In this configuration, up to 8 nodes can be simultaneously in the WBSN. Let us observe that, although the superframe is divided into a contention access period (CAP) and a contention-free period (CFP), the proposed protocol only uses the latter, since it pursues collision-free operation.

The size of the packets used in this protocol is 127 bytes (11 for the header, 114 of payload and 2 for the checksum). Then, each node can send up to four full packets in its GTS, because 7.36 ms are needed for filling the CC2420 transmission buffer and sending the packet over the air, as it will be shown in Section 3.4.1.

The proposed protocol uses a TDMA-like approach, which has been proven to be advantageous in star-topology WBSNs, since its schedule-based nature eliminates the possibility of having collisions and overhearing. Moreover, this protocol keeps the control packet overhead very low. The only control packets that are sent are the beacons to synchronize the nodes with their GTS and the ACKs that the WBSN coordinator sends to the nodes to confirm the reception of the transmitted packets. In addition, idle-listening is dramatically reduced, since the nodes know quite precisely when they have to be in receiving mode to listen to the beacon and the ACKs sent by the WBSN coordinator.

Very tight timing constraints are required for the protocol to keep all the nodes of the WBSN synchronized while reducing the idle-listening periods as much as possible, since an accurate timing will allow the nodes to adjust very

precisely the instants of time when they have to switch their radio on to listen for an incoming beacon. In addition, each platform of a WBSN can sense one or several vital signs (multi-parametric monitoring). A set of algorithms are applied on each signal separately, in order to condition the signal or extract meaningful information that can subsequently be combined to perform a more analysis considering all the signals. The algorithms used to process each of the acquired signals may also change dynamically, depending on the required performance or accuracy of the analysis at each moment. The use of an operating system (OS) helps to achieve an efficient and energy-aware management of the heterogeneous hardware resources of the WBSN node. Moreover, the OS must provide an effective scheduling of the different tasks to be executed in the microcontroller, such as signal processing algorithms, MAC protocol management or sensing, as well as interprocess communication. Finally, unlike standalone applications, in which the hardware resources have to be manually managed by the programmer, the use of an OS allows to decouple the hardware and software resources of the node, providing more flexibility. For this reason, the use of an OS is introduced in the following section.

3.3. Operating system

An OS for WBSNs should provide real-time capabilities to deal with the high sampling frequencies required by some of the biopotential signals as well as with the tight timing constraints, which have been presented in the previous section, and that are imposed by the proposed centralized schedule-based MAC protocol. The OS should also implement multitasking, since the node will be executing different software tasks concurrently. These tasks comprise sensing, MAC protocol management and all the algorithms that are applied on the various signals that are acquired. Finally, the OS must be small in terms of computational overhead and memory usage, due to the limited processing and storage resources of WBSN platforms.

3.3.1. State of the art

During the last years, several OSs specially designed for WSN devices have been proposed [FK11]. Two of the most salient and spread among the WSN community are TinyOS [LMP+04] and Contiki [DGV04]. These OSs are analyzed in the following subsections, as well as FreeRTOS [biba], which is not specifically designed for WSNs, but fulfills all the requirements to be used by WBSN devices.

3.3.1.1. TinyOS

TinyOS [LMP+04] is an open source event-driven OS designed for low-power wireless devices, such as those used in sensor networks, ubiquitous computing, personal area networks, smart buildings, and smart meters. TinyOS can support concurrent programs with very low memory requirements. The OS has a footprint that fits in 400 bytes. The TinyOS component library includes network protocols, distributed services, sensor drivers, and data acquisition tools. Unfortunately, although it is possible to use TinyOS for some specific real-time applications due to its event-driven capabilities, this is not enough for the applications considered in this thesis. Sometimes, serving a new event, as for example a packet reception, requires a long execution time. In these cases, TinyOS recommends to launch a new task, in order to keep the code associated to the event as short as possible and therefore not disabling other events during long periods. Anyway, due to the non-preemptive nature of the TinyOS task scheduler, tasks should also be short, in order to yield the CPU often enough so that the rest of tasks can accomplish their deadlines. For complex systems, that execute many task, this forces the programmer to generate very specific and complicated code for each application.

Summarizing, TinyOS does not feature hard real-time capabilities, which are required by many WBSN applications due to the high sampling frequency that is required, the complexity of the algorithms and, especially, the wireless communication, since most of the MAC protocols for WBSNs follow a centralized approach that involves very tight timing constraints to maintain the synchronization among all the devices of the network.

3.3.1.2. Contiki OS

Contiki [DGV04] is a small, open source, highly portable, multitasking computer OS developed for use on a number of memory-constrained networked systems ranging from 8-bit computers to embedded systems on microcontrollers, including sensor network platforms. Contiki is designed for embedded systems with small amounts of memory. It consists of an event-driven kernel on top of which application programs are dynamically loaded and unloaded at runtime. The processes of this OS use light-weight protothreads that provide a linear, thread-like programming style on top of the event-driven kernel. It also supports per-process optional preemptive multi-threading, interprocess communication using message passing through events, as well as an optional GUI subsystem with either direct graphic support for locally connected terminals or networked virtual display with VNC or over Telnet. Contiki runs on a variety of platform ranging from embedded microcontrollers such as the MSP430 and the AVR to old home PCs. Code footprint is on the order of kilobytes and memory usage can be configured

to be as low as tens of bytes. However, as TinyOS and most of the OSs especially designed for WSNs, Contiki does not have real-time capabilities, which makes it impossible to use in some WBSNs.

3.3.1.3. FreeRTOS

FreeRTOS [biba] is a portable, open source, hard real-time OS for embedded devices. The main advantages of this operating systems are:

- Supports a full preemptive task scheduler (although it also includes cooperative and hybrid configuration options).
- Implements task prioritization, without software restrictions on the number of priorities that can be used.
- Different memory models (from a very light one to a fully featured version).
- Queues, binary semaphores, counting semaphores, recursive semaphores and mutexes for communication and synchronisation between tasks, or between tasks and interrupts.
- Mutexes with priority inheritance.
- Supports efficient software timers.
- Powerful execution trace functionality.
- Stack overflow detection options.
- Free development tools for many supported architectures.

FreeRTOS is designed to be small and simple. The kernel itself consists of only three or four C files. Moreover, all the previously enumerated features can be selected separately to be included or not, allowing to create a fully customized version that only includes the needed functionalities to reduce the program memory usage. This possibility leads to the creation of very light prioritized multitask real-time applications. For these reasons, FreeRTOS was ported to the Shimmer™ platform, which implements the following tasks:

- High priority: manage the MAC protocol.
- Medium priority: sample ADC and store the value.
- Low priority: signal processing algorithm, including packing in frames the information to be sent via radio.

When none of these tasks is being executed, the scheduler puts the microcontroller in low-power mode. This enables an efficient and energy-aware management of the resources of the platform while having a very low overhead associated with the OS.

FreeRTOS has been ported to several microcontrollers. In particular, it includes support for the TI MSP430 microcontroller and the CC2420 IEEE 802.15.4-compliant radio chip included in the Shimmer™ platform.

3.4. Characterization of the Shimmer™'s energy consumption

This section presents a characterization of the timing and power consumption of all the components of the Shimmer™ platform in their different modes of operation. Once this information, which is not fully available in the datasheets of the components, is obtained, a very accurate estimation of the Shimmer™'s energy consumption (and therefore lifetime) under controlled circumstances can be calculated.

3.4.1. Experimental framework

This work is based on the Shimmer™ platform, as a representative state-of-the-art wearable platform, equipped with an ultra-low-power microcontroller (Texas Instruments MSP430) and a low-power low-rate IEEE 802.15.4-compliant radio (CC2420). Anyway, the results obtained for this platform can be generalized to any other WBSN platform and application with similar requirements. For all the following, we consider the use of FreeRTOS and the MAC protocol proposed in Section 3.2.

To measure the power consumption during operation, a 10.3 Ω resistor is placed in the power path of the node. The voltage is then measured using an oscilloscope, and the corresponding current and power consumption are calculated. Using this configuration, the power consumption of each single component of the node is measured by enabling it and disabling the rest of components of the platform. For each component, a very detailed profile is obtained for all its operating modes. Apart from the power consumption values for each operating mode, this profile provides timing information that shows, for example, how much time is required to transmit a packet from the microcontroller to the radio, how long it takes to send a packet through the wireless link, the required time to switch between different modes, etc.

For all the experiments, two nodes have been used. One node is in charge of sampling the ECG signal, processing it and sending the results to the other node, which works as a base station. Since GTS scheme is being used, the results obtained with this setup can be extrapolated to a network with more nodes, because additional nodes do not interfere with each other.

Figure 3.5 depicts the power consumption trace related to a sensing node running a simple streaming application, in which the raw ECG signal is forwarded to the WBSN coordinator without performing any on-board processing. Regarding the radio, three main phases can be distinguished in Figure 3.5: beacon reception, low power (LP) mode and transmission. During beacon reception, the radio first switches to reception mode before the beacon is expected, then receives it. Therefore, the beacon reception phase in Figure 3.5 has two clearly distinct parts: (1) the first part lasts for 1.39 ms, has a power consumption of 72.39 mW, and corresponds to a radio in reception mode waiting for the beacon while the microcontroller is idle; (2) the second part lasts for 0.97 ms, has a power consumption of 82.59 mW, and corresponds to a radio in reception mode and a microcontroller awake reading from the radio and analyzing the received beacon. After the beacon reception, the node radio enters a LP mode until the start of its assigned GTS. Finally, during its GTS, the node transmits the ECG signal to the coordinator. Figure 3.5 depicts the transmission of two data packets. In every packet transmission, two different zones can be clearly distinguished: (1) the first zone corresponds to the microcontroller sending the full packet to the CC2420 radio chip (in idle mode), and consumes 16.8 mW during 2.85 ms; (2) the second zone corresponds to the radio actually transmitting the packet while the microcontroller switches back to idle mode, and dissipates 51.92 mW during 4.51 ms. After each packet transmission, the radio is seen to automatically go to reception mode, while it waits for an ACK from the WBSN coordinator. The ACK reception lasts for 0.75 ms and has an associated power consumption of 72.39 mW. During the entire process, the microcontroller is continuously updating the operating system tick counter (every 0.32 ms) and the sampling task is reading one sample from the ADC every 4 ms. This process is especially noticeable during the low-power phase of the radio. Table 3.1 summarizes the power consumption of the microcontroller and the radio for the different operating modes of our platform, while Table 3.2 reports the timing information related with the aforementioned phases of the radio.

Table 3.1: Average power characterization (in mW) of Shimmer™

	CPU	Radio	Total
Radio reception (μC idle)	6.60	65.79	72.39
Radio reception (μC active)	16.80	65.79	82.59
Packet transmission (from μC to radio)	16.80	0.00	16.80
Packet transmission (actual radio transmission)	6.60	45.32	51.92
Sampling (μC and radio idle)	6.60	0.00	6.60
Sampling (μC active and radio idle)	16.80	0.00	16.80

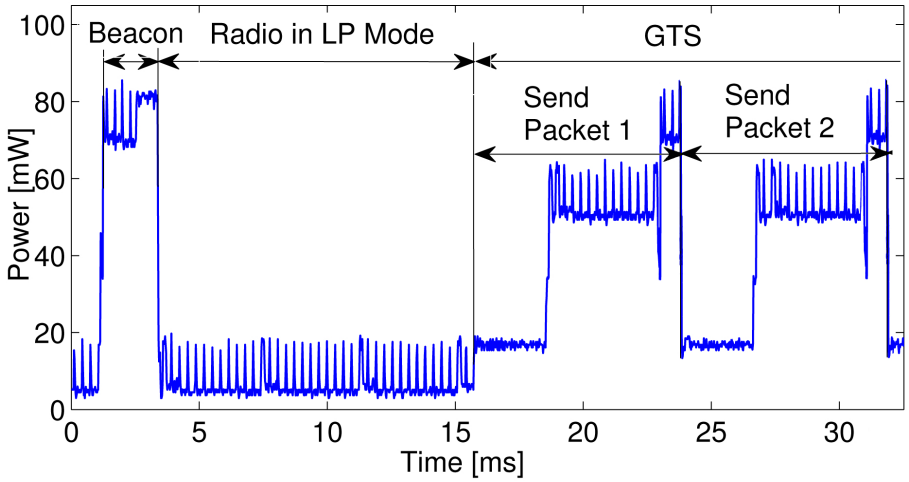

Figure 3.5: Power dissipated in Shimmer™ during sampling, processing and transmission

Table 3.2: Average timing characterization (in ms) of Shimmer™

Beacon reception (radio in reception and μC idle)	1.39
Beacon reception (radio in reception and μC active)	0.97
Packet transmission (from μC to radio)	2.85
Packet transmission (actual radio transmission)	4.51
ACK reception	0.75

3.4.2. Reference case study: ECG streaming

Although WBSN platforms are equipped with a microcontroller that provides processing capabilities to the system, in most cases this microcontroller simply acts as a control unit to manage all the hardware resources of node. Instead of performing any processing on the data acquired by the sensors, the complete raw signal is sent to the WBSN coordinator. One of the main contributions of this work consists of achieving on-board on-line diagnosis to be able to detect possible arrhythmias in real time (traditionally this processing has been performed following an off-line approach). Hence, the next chapter proposes several techniques to exploit the capabilities of the microcontroller to reduce the amount of information that needs to be sent by the nodes, and therefore reduce the energy consumption of the radio, leading to a longer lifetime of the sensor platform. Thus, this section presents a configuration that performs ECG streaming as a reference case study to evaluate the possible savings or penalization in terms of energy consumption of the

different techniques to perform advanced on-board processing proposed in the next chapter.

Based on the above power characterization, Table 3.3 shows the energy consumption of a Shimmer™ node running a simple ECG streaming application that simply sends the raw ECG signal sensed by the platform to the WBSN coordinator without any preprocessing. In this case, if any information wants to be extracted from the ECG, additional software needs to be used on the WBSN coordinator to process the signal (eg.: Holter).

The first row reports the duty cycle in percentage of ECG signal acquisition time. Since the application does not carry out any signal processing, the duty cycle in this case is 0. The second row indicates the time interval between the generation of two consecutive data packets. Given the facts that the payload of a packet is 114 bytes, a sample is represented using 1.5 bytes and the considered sampling frequency is 250 Hz, the streaming application generates one packet every 304 ms. The third row shows the beacon interval, which is adjusted according to the time interval between consecutive packets. Taking into account that the length of the queue provided by FreeRTOS to store the data to be transmitted is limited to 4 packets, BCO should be lower or equal to 6. The highest value is chosen, to reduce as much as possible the energy consumption in the radio, therefore the beacon interval is 983 ms. The fourth row reports the total energy consumption per second, which is 7.70 mJ in this case. Finally, the last row is directly calculated from the fourth one, assuming the energy supply of the Shimmer™ platform is a 280 mAh Li-ion battery at 3.7 V, and shows the lifetime of the system, which is 134.6 hours.

Table 3.3: Node lifetime for the ECG streaming application

Duty cycle (%)	0
Packet ready every... (ms)	304
Beacon interval (ms)	983
Energy consumption (mJ)	7.70
Lifetime (h)	134.6

Figure 3.6 shows the percentage of the total energy consumption of the Shimmer™ due to sensing and radio communication for the ECG streaming application. The microcontroller values are not included in the figure, since this application does not require any kind of processing. The energy consumption (due to sensing) is constant, and cannot be reduced unless a different ADC is used. Therefore, the only part of the energy consumption that can be reduced is the one due to the radio communication. Hence, the next chapter proposes several algorithms that extract only the relevant information of the ECG signal in order to significantly reduce the amount of data that needs

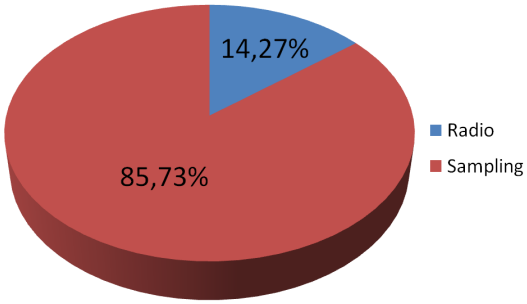

Figure 3.6: Energy consumption of the Shimmer™ due to sensing and radio communication for the ECG streaming application

to be transmitted, and therefore decrease the radio's energy consumption. If the computational overhead of these algorithms is low enough, the lifetime of the platform would not be affected (or it can even be increased), while achieving smart real-time ECG diagnosis.

Chapter 4

Application-level optimizations

The function of good software is to make
the complex appear to be simple.
Grady Booch

In Chapter 2, we have seen that WBSN platforms are equipped with microcontrollers, which supply processing power and memory resources. The capabilities of the node's microcontroller can be exploited to apply different real-time signal processing algorithms on the RAW ECG data sensed by the node, such as filtering, compression, feature extraction, arrhythmia detection, etc. These algorithms can significantly reduce the amount of information to be transmitted by the sensor platform, therefore reducing its energy consumption due to data transmission. Apart from the possible reduction of the total energy consumption of the node, the onboard execution of applications helps the patients and healthcare givers providing real-time monitoring and diagnosis of the health conditions of the patient. Thus, a sensible development and optimization of these algorithms for its real-time execution on a WBSN platform is needed, in order to reduce the computational complexity and memory usage requirements of the algorithms and therefore keep the impact in the energy consumption of the microcontroller as low as possible.

Figure 4.1 depicts the software application architecture of the wireless ECG monitor proposed in this work. The ECG signal is first acquired and conditioned to reduce the noise and artifacts (ECG filtering). Then, the peak and boundaries of the characteristic ECG waves are detected (ECG delineation). Finally, this information is used to perform arrhythmia diagnosis. The results obtained after the ECG delineation and arrhythmia diagnosis phases (and optionally the RAW ECG signal) are wirelessly transmitted to a WBSN coordinator, where they can be displayed or stored.

In this chapter, I first introduce several algorithms to perform ECG signal filtering, delineation and arrhythmia detection. Then, for each algorithm,

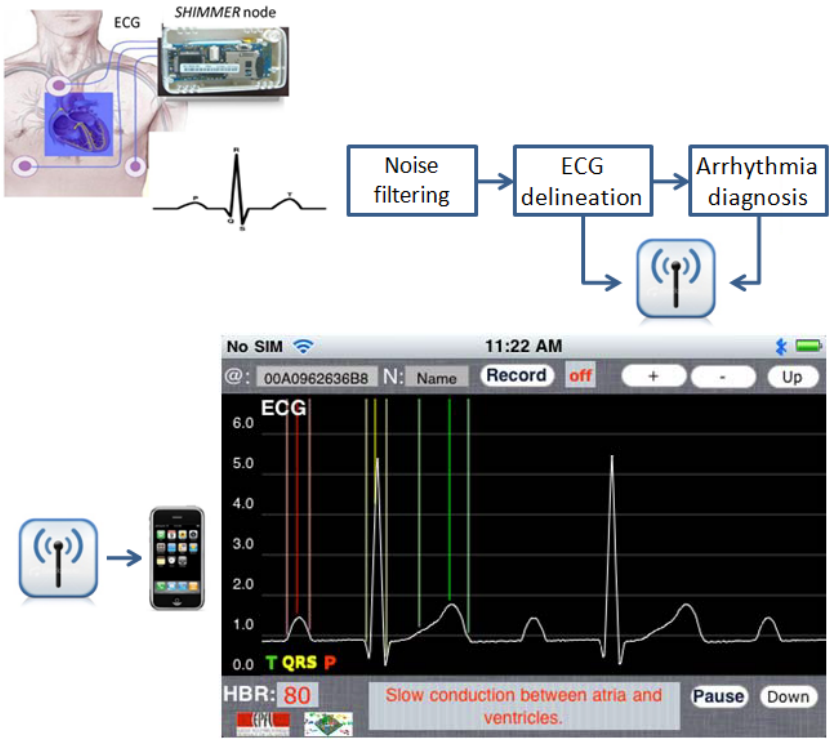

Figure 4.1: Software application architecture of our wireless ECG monitor

the required optimizations for their execution on the Shimmer™ platform are also tackled. Finally, I provide a performance evaluation as well as a comparative study of its energy consumption (and corresponding sensor node lifetime).

4.1. ECG filtering

In order to improve the quality of the ECG signal and be able to perform a more accurate diagnosis, it is crucial to reduce as much as possible the presence of noise and artifacts, such as baseline wander and electromyographic noise. The baseline wander is an abnormal, low-frequency activity in the ECG which may interfere with the signal analysis, making the clinical interpretation inaccurate. Baseline wander is often exercise-induced and may have its origin in a variety of sources, including perspiration, respiration, body movements and poor electrode contact. The spectral content of the baseline wander is usually in the range between 0.05 and 1 Hz [fQEWP85] but, during strenuous exercise, it may contain higher frequencies.

The electromyographic noise is caused by the electrical activity of skeletal muscles during periods of contraction, commonly found in ECGs recorded

during ambulatory monitoring exercise. This kind of noise can either be intermittent in nature, due to a sudden body movement, or have more stationary noise properties. The frequency components of this noise considerably overlap those of the QRS complex.

In this section, two techniques for ECG filtering are proposed, namely, cubic spline baseline estimation and morphological filering. The first one only performs baseline wander removal, while the second one also includes a noise suppression phase to reduce the high-frequency noise such as electromyographic noise or power-line interference.

4.1.1. Cubic spline baseline estimation

4.1.1.1. Method

This technique, based on [MK77], uses a third-order polynomial to approximate the baseline wander, which is then subtracted from the original signal. To this end, a representative sample (or knot) is chosen for each beat from the silent isoelectric line, which is represented by the PR segment in most heart rhythms. The polynomial is then fitted by requiring it to pass through successive triplets of knots.

This idea comes from a previous research [MPL$^+$77] that tried to adapt a straight-line to the segments connecting the pre-P-wave period and the post-T-wave period of each beat as successive baseline estimates. While this solution preserves low-frequency heart activity and leads to a small computational cost, such a first-order estimator can only accurately track baselines of very low frequencies. Furthermore, the resulting baseline estimate does not adapt properly to the variations and, what is worse, its derivatives at the knots are discontinuous.

Increasing the order of the polynomial and selecting one knot per beat through which the baseline estimation must pass is the method used to remove higher-frequency baseline noise and preserve low-frequency heart information, which is useful for other processes to apply after the baseline wander removal. By using higher-order polynomials, the likelihood of producing an accurate baseline estimate increases, although it is obviously linked to an increased computational complexity.

Instead of letting the order increase as the number of knots does, third-order polynomial fitting to successive triplets of knots represents a popular approach [MK77, BMT91] and leads to good results in terms of baseline removal. This technique requires the QRS complexes to be detected and the corresponding PR segments to be accurately determined. An averaged point in each PR segment of the ECG signal is chosen as sample of the baseline. This segment is used because of the ease and accuracy in locating it. At each PR segment, there is a knot through which the baseline noise estimator must pass. By fitting a third-order polynomial through these knots in the ECG

signal, we get the estimation for the baseline (see Figure 4.2). The polynomial is fitted in such a way that, once it is subtracted from the original signal, these knots have a value of 0.

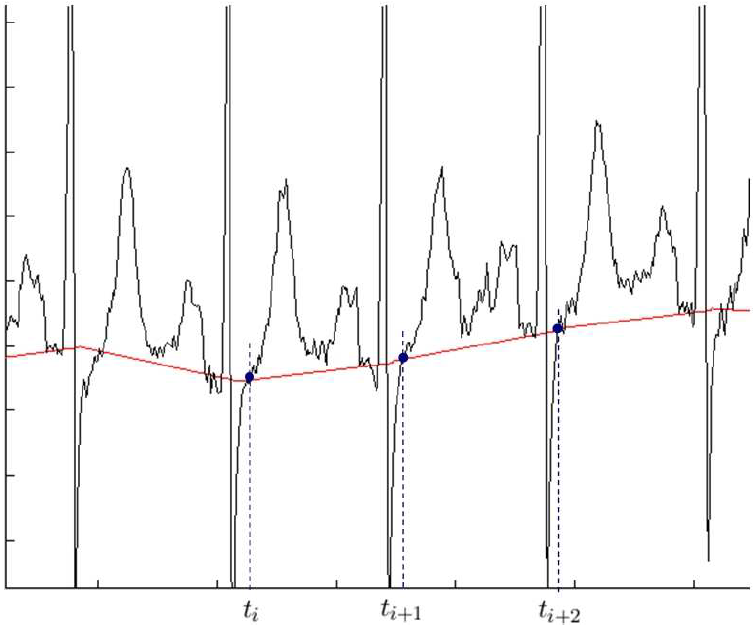

Figure 4.2: ECG signal with three knots and the cubic spline baseline wander estimation $y(t)$

The performance of the cubic spline technique critically depends on the accuracy of the knot detection. The segment is relatively easy to delimit in ECG signals recorded during resting conditions but it may be very difficult to find it in recordings with muscle noise or when certain types of arrhythmias are present, such as ventricular tachycardia, which severely distorts the ECG signal and makes the location process quite difficult. On the other hand, this approach to baseline wander removal results in a time-variable cut-off frequency linear filtering since the baseline estimate tracks rapid baseline wander when a fast heart rate is encountered. More knots become available at faster heart rates, so the segment between beats is shorter and the cubic spline can adapt itself better to the located knots. According to this, polynomial fitting performs poorly when the available knots are too far from each other, since the interval between t_i and t_{i+1} is too long to achieve a proper estimation.

In order to accurately detect the PR segment, a reduced version of the single-lead WT-based delineator that will be presented in Section 4.2.2.1 is used. This version of the delineator detects the end of the P wave and the beginning of the QRS complex. Then, one of the points between these two is selected as knot of the heart beat under analysis. Once the knots of three

consecutive beats of the input signal have been located, we can proceed with the cubic spline baseline estimation algorithm. The knots of the successive heart beats are denoted for the signal $x(t)$ as

$$x(t_i), i = 0, 1, 2, ..., \tag{4.1}$$

The baseline estimate $y(t)$ is computed for the interval $[t_i, t_{i+1}]$ by incorporating the three knots $x(t_i)$, $x(t_{i+1})$, $x(t_{i+2})$ into the Taylor series expanded around t_i.

$$y_\infty(t) = \sum_{l=0}^{\infty} \frac{(t-t_i)^l}{l!} y_\infty^{(l)}(t_i) \tag{4.2}$$

For a third-order polynomial description, this series is truncated to

$$y(t) = y(t_i) + (t-t_i)y'(t_i) + \frac{(t-t_i)^2}{2}y''(t_i) + \frac{(t-t_i)^3}{6}y'''(t_i) \tag{4.3}$$

And the series expansion for the first derivative $y'(t)$ is

$$y'(t) = y'(t_i) + (t-t_i)y''(t_i) + \frac{(t-t_i)^2}{2}y'''(t_i) \tag{4.4}$$

At $t = 0$ we assume, to get this technique working, that

$$y(0) = x(0) \tag{4.5}$$

We must approximate the first derivative $y'(t_i)$ at t_i by the slope between $x(t_{i+1})$ and $x(t_i)$

$$y'(t_i) = \frac{x(t_{i+1}) - x(t_i)}{t_{i+1} - t_i} \tag{4.6}$$

As shown in [Gre69], classical splines of order three and higher, in which only the highest derivative is discontinuous, suffer stability problems during computation so we define both $y(t)$ and $y'(t)$ at each knot to arrive at a stable solution.

At the next beat, and to keep the cubic spline adapted to pass through all the considered knots, we must approximate, once more,

$$y'(t_{i+1}) = \frac{x(t_{i+2}) - x(t_i)}{t_{i+2} - t_i} \tag{4.7}$$

To find the remaining two variables $y''(t_i)$ and $y'''(t_i)$ in $y(t)$, the Taylor series for $y(t)$ and $y'(t)$ is studied for $t = t_{i+1}$:

$$y(t_{i+1}) = y(t_i) + y'(t_i)(t_{i+1} - t_i) + y''(t_i)\frac{(t_{i+1} - t_i)^2}{2} + y'''(t_i)\frac{(t_{i+1} - t_i)^3}{6} \tag{4.8}$$

$$y'(t_{i+1}) = y'(t_i) + y''(t_i)(t_{i+1} - t_i) + y'''(t_i)\frac{(t_{i+1} - t_i)^2}{2} \tag{4.9}$$

To get the cubic spline to pass through this knot,

$$y(t_{i+1}) = x(t_{i+1}) \tag{4.10}$$

Inserting these values of $y(t_{i+1})$ and $y'(t_{i+1})$ into the previous equations we get:

$$y''(t_i) = \frac{6(y(t_{i+1}) - y(t_i))}{(t_{i+1} - t_i)^2} - \frac{2(2y'(t_i) + \frac{y(t_{i+2}) - y(t_i)}{(t_{i+2} - t_i)})}{(t_{i+1} - t_i)} \tag{4.11}$$

$$y'''(t_i) = -\frac{12(y(t_{i+1}) - y(t_i))}{(t_{i+1} - t_i)^3} + \frac{6(y'(t_i) + \frac{y(t_{i+2}) - y(t_i)}{(t_{i+1} - t_i)})}{(t_{i+1} - t_i)^2} \tag{4.12}$$

where, as we know,

$$y(t_{i+2}) = x(t_{i+2}) \tag{4.13}$$

We have then the baseline estimate $y(t)$ completely specified to be computed in the interval $[t_i, t_{i+1}]$. To get the signal without baseline wander we have to subtract the baseline estimate $y(t)$ from the original ECG signal in that interval. Then, this procedure has to be repeated for the next interval $[t_{i+1}, t_{i+2}]$ using the knots x_{i+1}, x_{i+2} and so on.

4.1.1.2. Real-time embedded implementation

The use of this technique in an embedded platform has to consider the fact that the PR segment needs to be accurately defined in each beat. To perform this task, a reduced version of the single-lead WT-based delineator that will be presented in Section 4.2.2.1 is used. Then, we experimentally choose as a knot the point that is 28 ms (7 samples) before the beginning of the QRS complex. We use a *struct* to store all the information related to each of the three consecutive beats needed by the algorithm. This information includes how many data samples the beat has, the ordinal value of its knot, a pointer to a buffer where all its data samples are stored and how many bits are needed to represent these values.

When a new knot is located, it also means that the previous beat has just finished. The length of the previous beat is then stored in its associated *struct* and, by using a variable to keep track of the maximum sample value in that beat, we can get how many bits are needed to represent all the values in that beat. We can then restart the length counter and reinitialize the variable for the search of the maximum value in the new beat. Subsequently, it is time to proceed with the cubic spline filtering in case it is at least the third beat that has taken place. Otherwise, we have to wait until we have detected three beats, with their associated *structs* completely filled and all the samples values stored in memory.

By using the amount of bits needed to represent all the values of the beat, we calculate how many bits we can shift the samples to the left. In this way

we optimize the use of the available bits for the coding of the intermediate operands and therefore we get the best accuracy in all the operations. For a beat i, and following the equations explained in the previous section, we need to get the ordinal value of its knot and the knots of the two following beats. These values are arithmetically left-shifted according to the maximum number of bits needed for their coding. Furthermore, we need the information about how long the beats i and $i+1$ are. With these data, we can obtain $y'(t_i)$, $y''(t_i)$ and $y'''(t_i)$. These variables use the length of the beat to the power of 3 and to the power of 2. Once we get the value of these three variables, which will remain constant during the processing of the whole beat, we use a new *struct* to facilitate the computation of the result, that stores the values of $y'(t_i)$, $y''(t_i)$ and $y'''(t_i)$, the ordinal value of the knot of the beat, the length of the beat, a pointer to the buffer with all its samples and how many bits the variables in the calculation have been shifted. This new *struct* uses a queue scheme: we will have a pointer to a *struct* with the information about the beat in which the computation is taking place and this *struct* will also have a pointer to the following *struct* of a beat ready to be processed, if it is the case. When the new *struct* is ready, it is stored in the process queue. Finally, once the new input sample is processed, we pay attention to the queue to check whether there is a new beat ready to be processed or not.

If a new beat is available, every time a new sample is read (every 4 ms) we provide as output a new value of the first beat in the queue. With the values stored in its *struct* inserted into the equation for $y(t)$, we get the baseline estimation for the current output value. The, by subtracting the estimation from the original input sample, we get a new value of the filtered output signal. After moving indexes for the circular buffer, and increasing the counter for the length of the current one, the algorithm is ready to receive a new input sample.

A first implementation of this algorithm was developed for a PC, which used floating-point operands for the computation of $y'(t_i)$, $y''(t_i)$ and $y'''(t_i)$, but since the MSP430 microcontroller does not include a floating-point unit, we had to adapt the calculations to use only integer variables. However, we realized that a simple 16-bit integer implementation is not accurate enough for the calculations needed by this technique. Actually, for a 16-bit integer implementation, the result is a straight horizontal line which passes through the knot of each beat. So we decided to use 32-bit integers in the required cases to arithmetically left-shift as much as possible all the data needed in the computation of those variables to get a better accuracy.

In terms of calculation, the MSP430 microcontroller meets the real-time and accuracy requirements of this technique except for the computation of $y'''(t_i)$. As defined in its equation, this variable uses the power of 3 of the length of the beat. Again, since the MSP430 uses by default 16-bit operands,

only 40 to the power of 3 can be operated. Beats longer than 40 samples, which is obviously the commonest case since 40 samples between each beat means a heart rate of 375 beats per minute (bpm), lead to an overflow in the register. Therefore, the calculation of $y'''(t_i)$ has to be performed using 32-bit operands, supported by the MSP430 but more expensive in terms of computation.

Another factor we have considered is that we do not know how long a beat is going to be. We can try to estimate the duration of each beat once we have detected the first beats in the signal but this duration cannot be considered as constant. Therefore, we have to decide the length of the buffer in which all the data samples of a beat is going to be stored. In our tests, we consider a beat length of 500 samples. This results in a 30 bpm minimum heart rate for this technique to work.

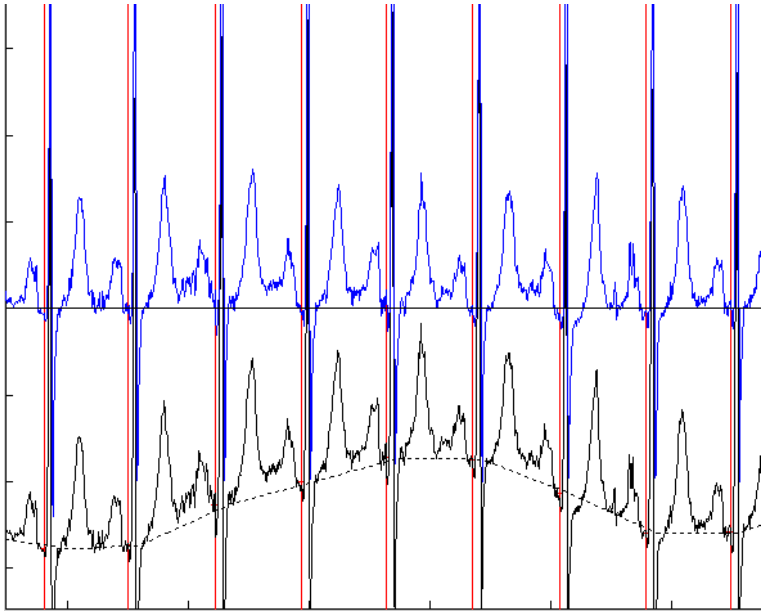

Figure 4.3: Input ECG signal, baseline estimation and output signal

To conclude with the implementation of this technique, Figure 4.3 shows how it works for a given input with baseline wander. The vertical red lines mark the position of the knot in each detected beat. The black ECG signal is the input signal whereas the dotted line is the estimation of the baseline wander. Finally, the blue ECG signal is the result of the baseline wander removal from the original input signal. The horizontal solid line shows the zero value.

4.1.2. Morphological filtering

4.1.2.1. Method

This method, first introduced in [SCK02], applies several morphological operations to the original ECG signal to estimate the baseline wander. Morphological operators have been widely used in the signal and image processing fields due to their robustness and adaptive performance in extracting information from the shape of the signals, as well as to their simple and quick computation. Mathematical morphology, based on sets operations, provides an approach to the development of non-linear signal processing methods, in which the shape information of a signal is incorporated. In these operations, the result of a data set transformed by another set depends on the shapes of the two sets involved. A structuring element has to be designed depending on the shape characteristics of the signal that is to be extracted.

Two basic morphological operators exist: erosion ($\ominus$) and dilation ($\oplus$). Using erosion and dilation we can define derived operators: opening ($\circ$) and closing ($\bullet$). We consider $f(n), \{n = 0, 1, ..., N-1\}$ as a discrete signal consisting of N points and $B(m), \{m = 0, 1, ..., M-1\}$ a symmetric structuring element of M points.

Erosion ($\ominus$) is a shrinking operator in which the values of $f \ominus B$ are always less than those of f.

$$(f \ominus B)(n) = min_{m=0,...,M-1}\left\{ f(n - \frac{M-1}{2} + m) - B(m) \right\} \tag{4.14}$$

Dilation ($\oplus$) is an expansion operator in which the values of $f \oplus B$ are always greater than those of f.

$$(f \oplus B)(n) = max_{m=0,...,M-1}\left\{ f(n - \frac{M-1}{2} + m) + B(m) \right\} \tag{4.15}$$
$$\text{for } n = \left\{ \frac{M-1}{2}, ..., N - \frac{M+1}{2} \right\}$$

The opening of a data sequence can be interpreted as sliding a structuring element along the data sequence from beneath and the result is the highest points reached by any part of the structuring element. Opening is used to suppress peaks and is defined as: $f \circ B = f \ominus B \oplus B$. The closing of a data sequence can be interpreted as sliding a flipped-over version of the structuring element along the data sequence from above, and the result is the set of lowest points reached by any part of the structuring element. Closing is often used to suppress pits and is defined as: $f \bullet B = f \oplus B \ominus B$.

This technique uses a sequence of opening and closing operations to perform a baseline wander removal and electromyographic noise suppression.

Based on the different characteristics of the baseline drift and the noise contamination in the ECG signals, different structuring elements and different morphological operators are used.

The baseline wander removal is performed by removing the drift in background from the original ECG signal, following the method presented in [CD89]. To get the baseline estimation we use $f_b = f_o \circ B_o \bullet B_c$ and then this baseline drift is subtracted from the original input signal to get the filtered output signal. The signal is first opened by a structuring element B_o for removing peaks in the signal. Then, the resultant waveforms with pits are removed by a closing operation using the other structuring element B_c. B_o and B_c are defined as two horizontal line segments of zero amplitude but with different lengths. The result of this compound operation is then an estimate of the baseline drift f_b and, therefore, the correction of the baseline is then done by subtracting f_b from the original signal f_o.

Different lengths in B_c and B_o are used because the construction of the structuring element for baseline correction depends on the duration of the characteristic wave and the sample frequency of the ECG signal F_s Hz. If the width of a characteristic wave is $T_w(s)$, the number of samples of that wave is $T_w F_s$ so the structuring element B_o should have a length larger than $T_w F_s$. The subsequent closing operation, which uses B_c, takes place to remove the pit left by the opening operation, so the length of the structuring element B_c must be longer than the length of B_o. The width of the characteristic waves of an ECG signal, such as the P wave, the T wave, and the QRS complex, is generally less than 0.2 s.Hence, L_o, the length of B_o, is selected as $0.2F_s$, and L_c, the length of B_c, is typically selected to be longer than B_o, at about $1.5L_o$. Since we are using $F_s = 250$ Hz as sampling frequency, we get $L_o = 0.2F_s = 0.2 \times 250 = 50$ and $L_c = 1.5L_o = 1.5 \times 50 = 75$.

After baseline correction, noise suppression is performed by processing the data through an opening and a closing operation concurrently, and then the results are averaged. The opening and closing operations for noise suppression use a structuring element pair, B_{pair} defined as $B_{pair} = \{B_1, B_2\}$ with B_1 and B_2 different in shape but equal in length. The sequence of B_1 and B_2 corresponds to the order of dilation and erosion in the opening and closing operations. The process of signal conditioning for noise suppression is described by the following equation:

$$\begin{aligned} f &= \frac{1}{2}(f_{bc} \bullet B_{pair} + f_{bc} \circ B_{pair}) \\ &= \frac{1}{2}(f_{bc} \oplus B_1 \ominus B_2 + f_{bc} \ominus B_1 \oplus B_2) \end{aligned} \tag{4.16}$$

where f is the resultant signal after noise suppression and f_{bc} the signal after baseline correction. The B_{pair} is selected by considering the purpose of analysis and the morphological properties of the ECG signal. B_1 is selected to be a triangular shape, used to retain the peaks and valleys of the charac-

teristic waves, such as the QRS complex. To minimize the distortion to the ECG signal, the length of B1 is chosen to be the same as that of B_2. The length of both structuring elements is related to the bandwidth of the ECG signal and the sampling rate. Since the sampling frequency is fixed, a shorter structuring element can be used to reduce the distortion of the waveform, so $B_1 = (0, 1, 5, 1, 0)$. B_2 is chosen to be a line segment of zero value and the same length as B_1, so $B_2 = (0, 0, 0, 0, 0)$.

Using the proposed structuring element pair, noise can be suppressed while reducing the smoothing of the significant peaks and valleys in the ECG signal, which are essential to subsequent reliable detection of the characteristic waves of the input signal.

4.1.2.2. Real-time embedded implementation

This morphological filtering technique has clearly been designed to work offline. According to the definition of the morphological operators, we need all the input data to perform the first erosion and this is impossible to implement on a real-time embedded platform. Even worse, in the proposed technique different morphological operations are compound. The subsequent dilation needs the result of the previous erosion so we need to keep track of a big amount of data in memory to get this technique working.

However, our implementation uses a constant and low amount of memory and supports the baseline wander removal and the noise suppression operations provided by this technique. We use several circular buffers to operate and store the partial results needed in each morphological operation achieving a very optimized final implementation that involves a very reduced execution time due to an improvement in the use of buffer indexes and peaks/pits comparison analysis that is presented in the following of this section.

All the buffers used in this implementation are circular buffers of 16-bit integers, the basic datatype of the MSP430 platform. The calculation of the baseline drift $f_b = f_o \circ B_o \bullet B_c$, is made through a sequence of erosion and dilation operations. The first operation, $f_1 = f_o \ominus B_o$ is an erosion that uses the input signal f_o and the structuring element B_o, which is a horizontal line segment of zero amplitude. According to the definition of this technique, B_o is 50 elements long. If we pay attention to what the erosion operation means, each sample $f(n)$, with $n = \{\frac{M-1}{2}, ..., N - \frac{M+1}{2}\}$, is going to be used in an operation where each element of B_o will be subtracted to this sample. The minimum value obtained in all the subtractions will be stored in $(f \ominus B_o)(n)$ (see Figure 4.4). For $n = \{0, ..., \frac{M-1}{2} - 1\}$, there is no definition of $(f \ominus B_o)(n)$, thus we use $(f \ominus B_o)(n) = f(n)$. In addition, since it is a real-time implementation and it is not known how long the input data will be, there is no definition for N.

Thanks to the fact that the structuring element B_o is a horizontal line

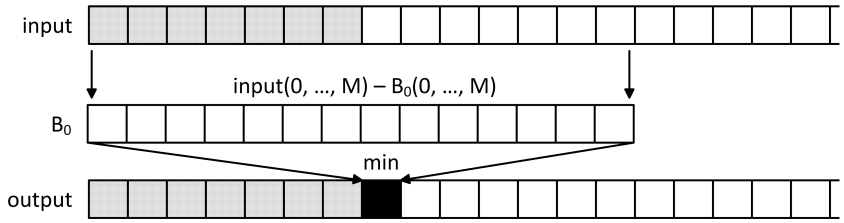

Figure 4.4: Representation of the erosion operation

segment of zero amplitude, the first erosion can be implemented as a simple search of the minimum value in a vector with the same length as B_o (50 elements in this case). Hence, the first circular buffer we use to support the erosion operation is 50 elements long. In this buffer, each new input sample is hosted in a position. When the buffer is completely filled for the first time, each new input sample will overwrite the sample stored 50 samples before. In order to implement the erosion operation, we use an index that points to the minimum value available in the buffer.

The first 50 input data samples are only used to fill this circular buffer. When a new sample is received in this initialization period, its value is compared to the previous minimum value received and stored in the buffer. If the new sample is lower or equal than the previous minimum, the correspondent index is updated and it is ready to receive a new sample. The idea of updating the index when the new value is equal to the previous minimum is made to keep the minimum index alive as much as possible and therefore reduce the chances of this index to be disabled by an overwriting operation in the position it is pointing.

While the circular buffer is being initialized, the result of the erosion operation is the same value received as input. Once the buffer is completely filled for the first time, the operation is different. Before storing the new input sample, we check whether it is going to overwrite the previous minimum found in the buffer. If it happens but the new value is equal or even lower than the previous minimum, there is no action to take: the minimum index will stay pointing to the current position and the operation will work properly. If the new value is higher than the previous minimum and it is going to overwrite it, we need to find the new minimum in the buffer. This is made by storing the new input sample in the buffer and then performing a linear search. On the other hand, if the new sample is not going to overwrite the previous minimum, we have to check whether its value is lower or equal to the previous minimum. If it is, the minimum index is updated to point to the current position. Otherwise, no change is made to any index.

We use this implementation because it provides a low number of indexes changes and comparisons. However, it has a flaw in terms of efficiency and computational cost, due to the linear search that needs to be done in the

circular buffer when a new input sample overwrites the previous minimum value. This buffer is 50 elements long so such a linear search would be likely to be avoided. A possible solution to avoid it consists in the use of a second index to point to the second minimum value. In order to evaluate if the use of a second index is worthy, since it involves more comparisons and conditional statements in the code, we analyzed how many linear searches are done in average for a heterogeneous set of ECG signals. For test purposes, we use several 15-minute ECG recordings sampled at 250 Hz (225000 samples per recording). Without the use of a second pointer, 25983 linear searches per recording were performed, in average, in this first erosion operation. This corresponds to the 11.55 % of the input samples.

The use of a second index changes the behavior of the implementation of the operation and increases the amount of comparisons. In this case, during the initialization period, if the new input sample is not lower or equal than the previous minimum found, we have to check whether it is lower or equal to the value pointed by the second index. As before, updating this pointer in case the new sample is equal to the pointed value allows this index to remain alive as long as possible so that the final number of linear searches is reduced. The second index can be disabled by storing the value -1.

When the circular buffer is filled, the operation with two indexes is more complex. If the new sample is going to overwrite the minimum in the buffer, we check whether it is lower or equal to that value. If it is, there is no change to make and the buffer is ready. Otherwise, we pay attention to the second index. If this second index is disabled (its value is -1), we have to perform a linear search to update the minimum and the second minimum indexes. If the new input sample is not lower or equal to the value pointed by the second index, we make the minimum index point to that value in the buffer and disable the second index.

If the new sample does not overwrite the position of the minimum value but it is lower or equal to this minimum, the minimum index will now point to the current position and the second minimum index will point to the previous minimum position. If the new sample is going to overwrite the value in the position pointed by the second minimum pointer, we disable this pointer in case the new value is greater than the previous second minimum. Otherwise, the second minimum pointer is still valid. Finally, if this second minimum index is not disabled and the new input sample is lower or equal to the value pointed by it, we change the index to point to the current position.

Once implemented and tested with the same input signal as the case in which we use a single index, only 12536 linear searches were performed, which corresponds to the 5.57 % of the input samples. This means a reduction of 13447 searches thanks to the use of two indexes instead of one.

Algorithm 4.1 shows the operation of the circular buffer B, the minimum index min, the second minimum index $secmin$ and the new position to be

Algorithm 4.1 Operation of the circular buffer and indexes

```
if cur = min then
  if newSample > B[min] then
    if secmin = −1 then
      Perform a linear search
    else if newSample > B[secmin] then
      min = secmin
      secmin = −1
    end if
  end if
else if newSample ≤ B[min] then
  secmin = min
  min = cur
else if cur = secmin AND newSample > B[secmin] then
  secmin = −1
else if secmin ≠ −1 AND newSample ≤ B[secmin] then
  secmin = cur
end if
```

written *cur*.

The next operation to be performed is the dilation that corresponds to the first opening with the input signal and the structuring element. This dilation takes as input the result of the previous erosion and uses the same structuring element B_o. Since B_o is a horizontal line segment of zero amplitude, this operation is similar to what has been explained before, with the only difference of searching for the maximum value in the vector, instead of the minimum one. As before, we need a 50-element circular buffer, corresponding to the length of the structuring element B_o.

Since there is no definition for the result of the first values of the input signal when $n < \frac{M-1}{2}$, this buffer is initialized with the input data until its half is reached. Once the buffer is half full, it will have to wait for the first output values of the first erosion operation. At the beginning we must fill the first circular buffer (erosion) and, then, we can fill this second circular buffer (dilation) with the results of the first one. The dilation operation is based on additions and a search for a maximum element, so in this case we need to keep two indexes to the maximum value and the second maximum value of the buffer. The behavior of these indexes is the same as in the previous case except for the fact that the maximum value is taken under consideration instead of the minimum one. As before, during the initialization of this buffer, we pay attention to the new values to keep track of where the maximum values are stored.

As in the previous case, we performed tests using one or two indexes to keep track of the maximum values. In the implementation that uses only one

index, for the same test input of 225000 samples, 77507 linear searches were made, which corresponds to a 34,45 % of the input data. Such an amount of searches justifies the additional comparison process needed by the use of a second pointer. As a result of this improvement, only 10255 linear searches were made, which corresponds to a 4,56 % of the input data. It is truly significant that for a higher amount of searches than in the case of the erosion operation, the use of a second pointer reduces even more the final number of searches. This is explained by the shape of the data provided as input to this dilation operation. At first, more linear searches are made but, with a second index, we get an greater reduction because of the lower number of invalidations of this second index.

This buffer cannot return any result until it is completely full. The circular buffer of the erosion had a specific initialization process which guarantees that after 50 input samples, it could be possible to get results from it. To get this dilation buffer working we have to wait for 75 input samples: 25 of them are going to initialize both the erosion and the dilation buffers, the following 25 will complete the erosion buffer and the last 25 samples will produce 25 output values after the erosion, which will be enough to completely fill the dilation buffer. Once the dilation buffer is full, each new input value will produce an output value after the erosion operation. This output will be at the same time an input for the dilation operation, and will produce an output value after the dilation (the one in the position pointed by the maximum index). This value is the result of the first opening operation $f_o \circ B_o$.

The next operation to be performed is a closing, which consists of a dilation and an erosion, and uses as input the result of the previous opening operation and the structuring element B_c, which is a horizontal line segment of zero amplitude with a length of 75 elements. After this closing operation we will get the baseline drift f_b to be subtracted from the input signal. As before, we use a circular buffer to store the result of the first dilation operation. To initialize this buffer, the first 25 samples are taken from the input signal, since there is no definition for them in the morphological operators. This buffer has to wait until the two previous buffers are filled and start producing results. Our implementation uses a control variable to check whether the previous dilation buffer is full and, then, start copying its output data to this new dilation buffer. While it receives the first raw input data it keeps track of the values to update the index pointing to the maximum stored value. When a new value is received, the procedure is equal to the previous cases. First, we have to check whether the new value is going to overwrite a value pointed by any index. Then, we check if the new value is greater than the maxima pointed by the indexes and finally store the new sample. According to my tests, the dilation operation triggers more linear searches than the erosion operation. In this buffer, using the input signal of 225000 samples, 67492 linear searches were made in the implementation with a single index, which

corresponds to a 30 % of the input data. However, by using a second index, the amount of linear searches is amazingly reduced to only 3419, a 1.52 % of the input data. This reduction is explained by several factors. Firstly, the signal this operation receives as input has been previously processed by an opening operation so its shape is not as rich as the shape of the raw input. Secondly, the structuring element, B_c is 50 % longer than the element B_o used in the previous opening operation, which reduces the possibility of overwriting an index.

This dilation operation starts returning output values when its associated circular buffer is full. Then, the value in the position pointed by the maximum index is the result of the dilation operation. The baseline drift computation finishes with the last erosion operation, which also completes the closing operation with the structuring element B_c. This last operation is implemented by a new 75-element circular buffer. The initialization of this buffer is more complex than for the previous buffers. Once more, the first 25 elements of the input dataset are stored directly at the beginning of this buffer. The dilation buffer of the closing operation has to wait until the dilation buffer of the opening operation was full. Then, it received values after the opening operation. The erosion buffer of the closing operation also starts receiving values from the dilation buffer of the opening operation, since there are values before the half of the previous buffer that are going to be copied directly (see Figure 4.5).

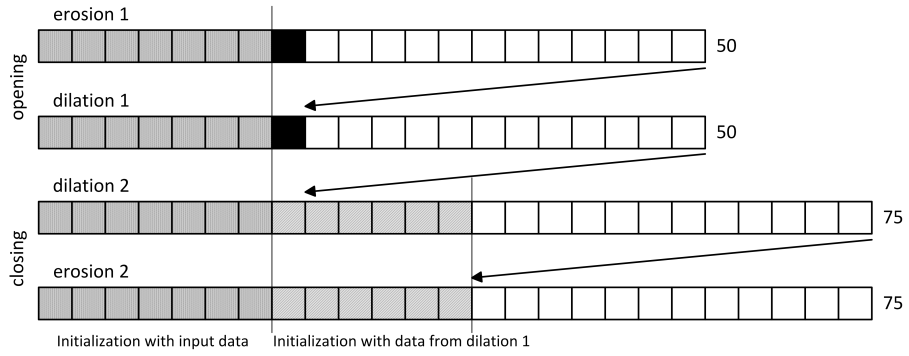

Figure 4.5: Scheme of the use of circular buffers for the implementation of the opening and closing operations and their initialization process

The behavior of this last erosion buffer is similar to the previous cases, since the structuring element B_c is a horizontal line segment of zero amplitude. The erosion operation returns as a result the minimum element in the array. In the implementation with a single index, and for the same input signal of 225000 samples as we used in the previous cases, 34416 linear searches were performed, which corresponds to a 15.3 % of the input dataset. By using a second minimum index, the number of linear searches is reduced

to 3062, which stands for only 1.36 % of the input signal length. Again, this huge reduction is due to the longer structuring element, which means that the chances of disabling an index are lower and, thus, the use of a second index is completely justified.

When this last circular buffer is full, we are able to get values for the baseline drift estimate f_b. After processing a new input sample, performing the opening and closing operations, the value in the position pointed by the minimum index of this last erosion buffer is the new value of the baseline drift. This value has to be subtracted from the input signal to get a signal without baseline wander. However, we have to remind that there is a delay between the input signal and the result of the filter because of the initialization process of the different buffers. While each buffer is being filled, no outcome is returned so the delay is increasing until all the buffers are full. An additional buffer is used to store the input signal and therefore overcome the delay between the input and the output. This is also a circular buffer with two indexes: one to control the next position to be written with a new input sample and another one to control which position has the next value to be used for the final output. Each output value produced by the last erosion operation has to be subtracted from the correspondent input data sample which is stored in this buffer. When the subtraction is performed, we have the input signal with baseline correction $f_{bc} = f - f_b$.

After the baseline correction is performed, we apply noise suppression. This is done by a concurrent closing and opening operation based on the use of two different structuring elements. These elements are 5-element long so, as we did in the previous cases, we will use a buffer of equal length. The first concurrent operation to be performed is an erosion and a dilation using $B_1 = (0, 1, 5, 1, 0)$. The approach followed for the implementation of this calculation is different to what has been carried out before. At first, we use a 5-element circular buffer to store the output of the baseline correction f_{bc}. To carry the first erosion and dilation out, we use a static array of elements to store the offset of B_1 for its use with the circular buffer. Since we can find the first element in any of the five positions of the circular buffer, we control that the structuring element matches the right element in the buffer (see Figure 4.6). This is done by unfolding all the combinations of B1, according to its starting point, in a bigger array of 25 elements: (**0**, 1, 5, 1, 0, 0, **0**, 1, 5, 1, ...).

Then, to calculate the first operations we only have to add (dilation) or subtract (erosion) each value in the structuring element to/from each value in the circular buffer. In the case of the erosion operation, we use a variable to store the minimum value obtained after the subtraction and, likewise, we use another variable to store the maximum value obtained after the addition. By accessing these values once the buffer has been covered, we get the result of this new erosion and dilation operations.

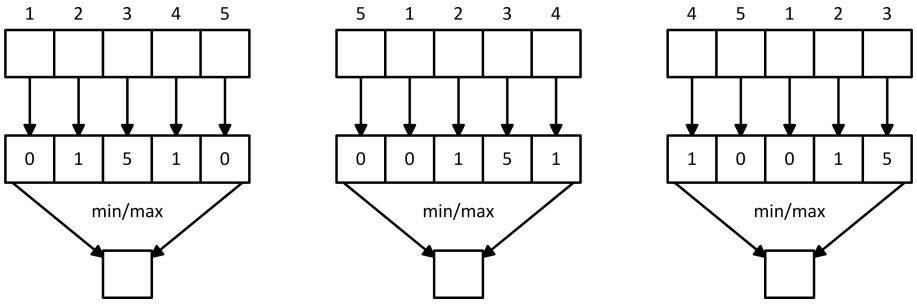

Figure 4.6: Matching of the circular buffer and the corresponding instance of the structuring element B_1 in its array

We need another variable to control in which position of the structuring element we have to start to calculate. When a new value is stored in the circular buffer, we have to get the maximum and minimum values obtained after the calculation with the structuring element. This is done by going through the circular buffer and adding or subtracting each of its values to the corresponding value of B_1. We use a chain of comparisons to get the final maximum and minimum values and, to optimize the amount of changes in each control variable, an analysis was made to know in which relative positions the maximum and minimum values were found.

The most straightforward implementation of this erosion and dilation operations, from the programmer's point of view, would be to start by the first element in B_1 and calculate the result with the corresponding first element in the circular buffer. Then, process the second and so on. However, since we are interested in an efficient implementation in terms of memory and computational power, we performed an analysis to check whether the maximum and the minimum values of each operation can be found with a higher probability in a certain position, so that we could reduce the amount of comparisons and changes of variables. Just by going through the array in a linear sequence, we cannot be sure of whether we are following the best path in the comparison chain.

The first implementation used a linear sequence to calculate with the structuring element by `buffer[0] + B1[j], buffer[1] + B1[j+1], ...` At first, a solution to reduce the amount of changes of the value of the variables would be to analyze in which positions the maximum and the minimum are found more often. It should obviously depend on the value of j, the variable that controls in which instance of the structuring element B_1 we are working on in the buffer `B1`. However, performing such an analysis for each instance and implementing this variable access method would lead to an extra cost of addressing, since the access to the array B_1 would depend, for optimization purposes, on another static array in which the best access chain for each instance would be stored: `buffer[0] + B1[opt[j]]`. This op-

timization would not be efficient since the extra addressing operation would lead to an higher cost in terms of computation and memory accesses.

The solution to avoid this extra cost while trying to perform such an optimization is not to consider in which instance of B_1 we are working to calculate the array `B1`. At first, we could consider that the maximum and minimum values are stored in whatever position so that it might be difficult to find a pattern for their locations. Thanks to the use of control variables, we discovered that, for the input test of 225000 samples, 295232 changes in the variable for the dilation operation were made and 271059 changes took place in the variable for the erosion. This amount of changes corresponds to 1.31 changes per sample in the dilation and 1.20 changes per sample in the erosion.

To try to reduce these figures, we performed an analysis using all the records of the QT database. The structuring element has five elements so we only had to check in which of these positions the maximum and the minimum were found while doing the calculation. Counting the changes for all the signals, for the sequence of accesses we tested, the sequence (3, 4, 0, 1, 2) resulted in 57596223 changes for the variable of the dilation operation and the sequence (2, 3, 4, 0, 1) gave as a result 57262531 changes for the variable of the erosion operation. We did not test all the possible combinations since the first element of the sequence can be used to prune the search space for the best combination.

The best access sequence for the dilation operation is (0, 2, 4, 1, 3), which reduces the amount of changes to 53035250 and, for the erosion operation, the best sequence is (0, 2, 3, 1, 4), which reduces the amount of changes for all the input dataset to 53075361. Applying these sequences to the implementation and using the same input test, the result was that only 261362 changes were made (a 11.47 % reduction and only 1.16 changes per sample) in the dilation operation. In the erosion operation, only 254809 changes took place (a 6.00 % reduction and 1.13 changes per sample).

This optimization does not involve extra computational cost since the implementation has to go through the hole array for the structuring element. Although the number of comparisons cannot be reduced, because we have to compare all the values, the number of changes in the variables of the morphological operations is reduced, leading to a better performance.

These concurrent erosion and dilation operations are needed in three different parts of the implementation of the algorithm. The structuring element is only 5 elements long, so it is completely filled with the first five input samples. While the first buffer for the erosion operation at the beginning of the implementation is being filled, those values are also stored in the first buffer for noise suppression. Since its length is of only five elements, the noise suppression code starts producing results before the baseline correction is done.

While the buffer for the second erosion in the baseline correction is being filled, those samples also have to be processed by the noise suppression part. They do nor produce any output for the baseline correction but they are needed in this filtering process. Finally, once all the buffers are full, the code for this first concurrent operation is ready to produce right results.

The second concurrent operation corresponding to the final erosion and dilation in the concurrent opening and closing is based on the use of the structuring element B_2, which is a horizontal line segment of zero amplitude with a length of five elements. The implementation of this second concurrent operation follows the same approach as the baseline wander removal part. We use two new circular buffers of five elements to support these operations and, since the structuring element has a constant zero value, we do not have to go through the buffer as we had to do when using B_1. The small size of the buffer (i.e., only five elements) can lead us to think that the use of only one index to control where the maximum and minimum values are stored is enough, since a linear search in such a small buffer is rapid. In order to analyze this, we performed experiments using one and two indexes. If we only use one index, 71090 linear searches where triggered for the dilation operation, which corresponds to the 31.6 % of the input test data. For the erosion operation, 97650 linear searches where performed, which stands for a 43.4 % of the input data. This high number of linear searches is due to the small size of the structuring element. With the use of a second index, the number of linear searches was reduced to 31844 (14.15 % of the input data) in the dilation operation and to 45061 (20.03 % of the input data) in the erosion operation. These new circular buffers have to be properly initialized by waiting until the previous buffers are full and start to produce output values.

Table 4.1: Reduction of linear searches in the circular buffers due to the use of a second index for a test input of 225000 samples

Buffer	# searches	# searches with 2nd index	Reduction
erosion 1	25983 (11.55 %)	12536 (5.57 %)	51.75 %
dilation 1	77507 (34.45 %)	10255 (4.56 %)	86.77 %
dilation 2	67492 (30.00 %)	3419 (1.52 %)	94.93 %
erosion 2	34416 (15.30 %)	3062 (1.36 %)	91.10 %
dilation 4	71090 (31.60 %)	31844 (14.15 %)	55.21 %
erosion 4	97650 (43.40 %)	45061 (20.03 %)	53.85 %

Once the concurrent operations are performed, we get the result of $f_{bc} \oplus B_1 \ominus B_2$ and $f_{bc} \ominus B_1 \oplus B_2$ by accessing the values pointed by the maximum and the minimum indexes, respectively, in the last circular buffers. To get the output of the filter after the noise suppression part, we have only to add

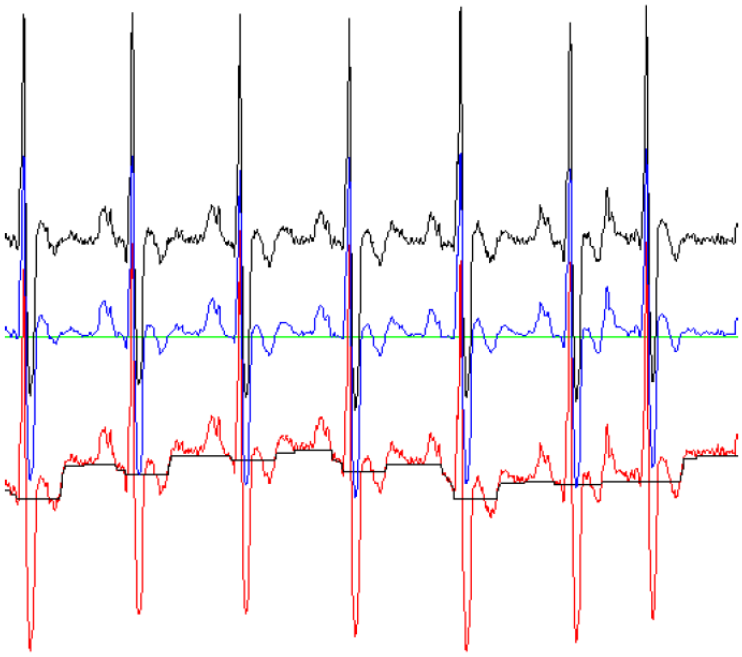

Figure 4.7: Baseline correction and noise suppression output (centered, in blue), baseline correction output (in black, at the top) and input signal and baseline estimation (below, in red and black)

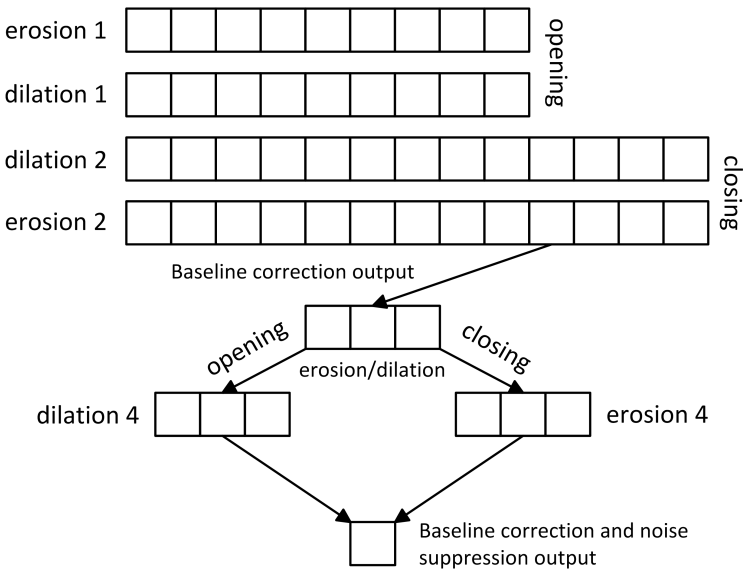

Figure 4.8: Scheme of the circular buffers and the performed operations

those values and average them: $f = \frac{1}{2}(f_{bc} \oplus B_1 \ominus B_2 + f_{bc} \ominus B_1 \oplus B_2)$. The result of each part of this filter is depicted in Figure 4.7.

In contrast to the cubic spline technique, this morphological filtering provides more stability. There is no estimator to use and, in fact, all the technique works the same for any point in any characteristic wave of the ECG signal. We have to keep in mind that the use of morphological operators are based on structuring elements chosen according to the shape of the signal we are trying to process. Thanks to the fact that the structuring elements B_o, B_c and B_2 are horizontal line segments of zero amplitude, the algorithm can be efficiently implemented with no extra cost for calculation for each element in those structuring arrays. Then, the main goal is to reduce the amount of linear searches. This technique is not artifact-dependent since an artifact in the input signal cannot trigger any special condition in the filtering process.

4.2. ECG delineation

A significant amount of research effort has been devoted to the automated analysis of ECG signals, and the underlying detection of the major ECG characteristic waves, namely the QRS complex, P and T waves [SL05], described in Section 1.2. In fact, the performance of an automated ECG analysis system critically depends on the reliable detection of these fiducial waves, so-called ECG delineation. This has motivated the rich variety of state-of-the-art ECG delineation approaches.

In this section, two of the most salient techniques for automated ECG delineation are analyzed. The first one is based on the wavelet transform (WT) [LZT95, MAO+04], and the second one on the multiscale morphological derivative (MMD) [SCK05]. Using these two approaches, several algorithms for single and multi-lead delineation are developed and optimized for their execution on the Shimmer™ platform.

The algorithmic transformations and software optimizations necessary to enable embedded ECG delineation notwithstanding the limited processing and storage resources of the target platform are described, and the performance of the resulting implementations are analyzed in terms of delineation accuracy, execution time and memory usage. Finally, a comprehensive evaluation of the energy consumption entailed by the considered algorithms is proposed.

4.2.1. Database and performance metrics

In this section, the database and metrics used for the performance evaluation of the subsequently presented algorithms are described. We used the validation procedure detailed in [MAO+04], in order to be able to compare the results. This procedure uses the 105 records of the QT database

(QTDB) [LMGM97]. The QTDB is a free-access database, which consists of 15-minute excerpts of two-lead ECG recordings sampled at 250 Hz, with manual annotations of the QRS complexes, P and T waves performed by expert cardiologists. These annotations span a total of 3600 beats, and feature a wide variety of ECG morphologies.

To assess the performance of the investigated algorithms, the following rules are applied. An automatic annotation is considered to be related to a manual one if the time interval between them is shorter than 320 ms. This value has been empirically chosen to compare our results to the ones of [MAO$^+$04] as fairly as possible, since part of the herein proposed algorithms are based on this work. This value allows us to obtain the closest results of sensitivity and positive predictivity to [MAO$^+$04], and therefore be able to directly compare the mean error and standard deviation values. Hence, this pair of annotations is a true positive. A manual annotation that has no corresponding automatic annotation is counted as a false negative, and automatic annotations without manual annotations count as false positive. Accordingly, the sensitivity (Se) and positive predictivity ($P+$) of a delineation algorithm are defined as [EC508]:

$$Se = \frac{TP}{TP + FN}, \quad P+ = \frac{TP}{TP + FP} \tag{4.17}$$

where TP is the number of true positive detections, FN is the number of false negative detections and FP is the number of false positive detections. The reported error statistics correspond to the computation of (automatic-manual) marks. In addition, due to the QTDB format, it is impossible to know if the lack of an annotation means that the wave is not present or that the cardiologist did not annotate the point with confidence for some reason. In the first case, an automatic detection would count as a false positive, but in the second case, it would not necessarily mean that the detection is erroneous. As a result of this overestimation of false positives, we show in the tables the value of P^+_{min}, which is a lower bound on the real value of P^+.

Moreover, the mean (m) and the standard deviation of the delineation error (σ) are also computed. The mean error (m), which is computed as the average of the errors across all the records, shows how close the results of the algorithm are to those that have been manually annotated in the database. Finally, the standard deviation of the error (σ), which is defined as the average of the standard deviation of each record, provides us information about the stability of the detections.

In order to compare the results of a single-lead delineation algorithm with the manual annotations of the database, which are performed using 2 leads, the algorithm is executed in both leads separately. Then, if 2 automatic annotations on different leads fall within the 320 ms interval, the closest one to the manual annotation is chosen to compute the error, and the other one is not counted as a false positive.

4.2.2. Single-lead WT-based ECG delineation

The first ECG delineation approach under analysis is based on the wavelet transform (WT) [LZT95, MAO+04], mainly because this transform provides a description in the time-scale domain that allows the suitable representation of signals having multiresolution characteristics such as the ECG signal. Indeed, the ECG signal is characterized by a cyclic occurrence of patterns at different frequency content (QRS complex, P and T wave).

4.2.2.1. State-of-the-art single-lead offline delineation algorithm

Each of the three waves of the ECG signal has different frequency content, as seen in figure 4.9. The QRS complex is characterized by relatively high frequencies, while the P and T waves are composed of low frequencies.

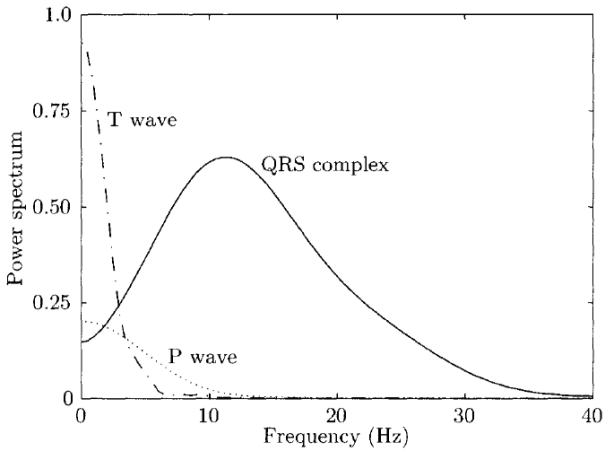

Figure 4.9: Frequency components of P, T and QRS waves (see [SL05], p. 428).

This kind of signal can be well decomposed using a dyadic discrete wavelet transform (DWT) [LZT95, MAO+04], which provides us with multiple outputs, called scales. Each scale matches different frequency bands of the original ECG signal, which will allow us to perform a multi-scale analysis to detect the ECG waves.

A dyadic DWT is composed of cascaded identical high-pass (G) and low-pass (H) filters. Mallat [Mal89] provides an implementation using decimation, which, as in [MAO+04], we will not consider here, as it introduces time-variance. We will instead use an implementation without decimation (*Algorithme à trous* implementation [CK96], as shown in Figure 4.10).

In the z-domain, the output $W_{2^k}X(z)$ of the DWT at scale 2^k can be

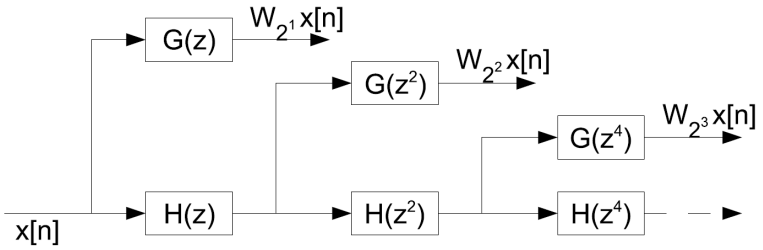

Figure 4.10: *"algorithme à trous"* implementation, as shown in [MAO+04], Figure 1.b

expressed as follows:

$$W_{2^k}X(z) = \begin{cases} G(z) & \text{if } k = 1 \\ G(z^{2^{k-1}}) \cdot \prod_{i=0}^{k-2} H(z^{2^i}) & \text{if } k \geq 2 \end{cases}$$

This work uses the prototype wavelet $\psi(t)$ proposed by [MZ92b]:

$$\Psi(\Omega) = j\Omega \left(\frac{sin(\frac{\Omega}{4})}{\frac{\Omega}{4}} \right)^4 \tag{4.18}$$

For this prototype wavelet, the filters $H(z)$ and $G(z)$ can be computed, then their finite impulse response in the time domain $h[n]$ and $g[n]$ can be obtained [MAO+04]:

$$h[n] = \frac{1}{8} \cdot (\delta[n+2] + 3\delta[n+1] + 3\delta[n] + \delta[n-1]) \tag{4.19}$$

$$g[n] = 2 \cdot (\delta[n+1] + \delta[n]) \tag{4.20}$$

These filters have the frequency responses shown in Figure 4.11, for an input sampling rate of 250 Hz. It is also important to mention that the output at each scale 2^k corresponds to the derivative of the filtered version of the input $x[n]$.

Five different scales are used in this work: from scale 2^1 matching high frequencies, thus very sensitive to noise, to scale 2^5 matching only low frequencies, causing it to smoothen the original signal, as seen in Figure 4.11.

In order to implement the filters, we first have to make them causal:

$$h[n] = \frac{1}{8} \cdot (\delta[n] + 3\delta[n-1] + 3\delta[n-2] + \delta[n-3]) \tag{4.21}$$

$$g[n] = 2 \cdot (\delta[n] + \delta[n-1]) \tag{4.22}$$

Since the implementation without decimation is used, the filters have to be rescaled at each step, where $g_k[n]$ and $h_k[n]$ are the time-domain equiva-

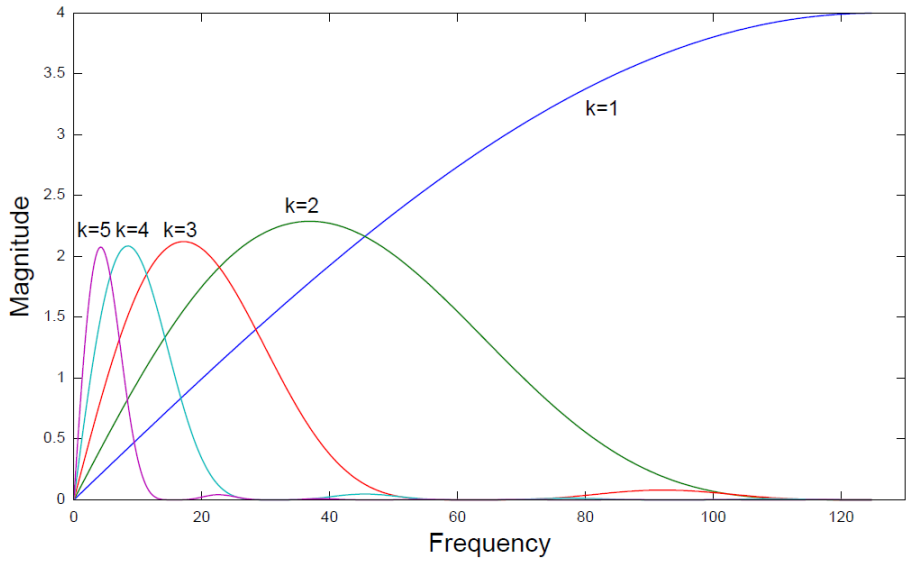

Figure 4.11: Frequency response of the DWT at scales 2^1 to 2^5

lent of respectively $G(z^{2^k})$ and $H(z^{2^k})$.

$$\left.\begin{array}{ll} h_k[n] & = \frac{1}{8} \cdot (\delta[n] + 3\delta[n-2^k] + 3\delta[n-2\cdot 2^k] + \delta[n-3\cdot 2^k]) \\ g_k[n] & = 2 \cdot (\delta[n] + \delta[n-2^k]) \end{array}\right\} \text{with } k = 0,1,2,3,4 \tag{4.23}$$

As the filters have linear phase, their outputs can be realigned for the purpose of obtaining no delay with respect to the ECG input signal, allowing us to perform analysis over multiple scales.

This delay is given by the mid-point of the non-zero portion of the impulse response:

$$\left.\begin{array}{l} \text{delay}(g_k[n]) = 0.5 \cdot 2^k \\ \text{delay}(h_k[n]) = 1.5 \cdot 2^k \end{array}\right\} \text{with } k = 0,1,2,3,4 \tag{4.24}$$

Thus, the delay of each filter, after rounding, is:

$$\text{delay}(W_{2^1}x[n]) = \lfloor \text{delay}(g_0[n]) \rfloor = \lfloor 0.5 \cdot 2^0 \rfloor = 0 \tag{4.25}$$

$$\begin{aligned} \text{delay}(W_{2^k}x[n]) &= \left\lfloor \text{delay}(g_{k-1}[n]) + \sum_{j=0}^{k-2} \text{delay}(h_j[n]) \right\rfloor = \left\lfloor 0.5 \cdot 2^{k-1} + 1.5 \cdot \sum_{j=0}^{k-2} 2^j \right\rfloor \\ &= \left\lfloor 0.5 \cdot 2^{k-1} + 1.5 \cdot \frac{1-2^{k-1}}{1-2} \right\rfloor = \left\lfloor 2^k - 1.5 \right\rfloor, \text{ with } k = 2,3,4,5 \end{aligned} \tag{4.26}$$

The numerical values are reported in Table 4.2.

Table 4.2: Position we need to look at for each filter when considering the input at time n

Input	2^1	2^2	2^3	2^4	2^5
n	n	$n+2$	$n+6$	$n+14$	$n+30$

Once the outputs of the discrete wavelet transform are computed, a set of rules are applied to extract the fiducial points of the ECG signal. The first detected point is the main peak of the QRS complex, then the QRS complex is delineated, to find its onset and end. Then the P and T waves are detected, along with their onsets and ends. A detailed description of these rules, that had to be adapted to allow a real-time implementation on the Shimmer™, are given in the following section.

4.2.2.2. Real-time embedded WT-based single-lead delineation

The algorithm is based on [MAO+04] and [LZT95], however, several modifications had to be introduced to allow an implementation on an embedded platform. Some new rules that were not explicitly mentioned in both publications have also been included.

QRS detection

QRS detection is performed as in [MAO+04]. First, we search for maximum moduli sets at scales 2^4 to 2^1 exceeding some thresholds ϵ^4_{QRS} to ϵ^1_{QRS}. Since the scales are derivatives of the original signal, this will correspond to slopes of the QRS complex. We start by looking for a maximum modulus at scale 2^4, then for a related one at scale 2^3, and so on for lower scales.

A maximum modulus at scale 2^k is considered to be related to another one at scale 2^{k-1} if the time interval between them is less than 40 ms. We simplified the rule proposed by [LZT95] in case more than one maximum lies within this interval, the closest one is selected.

As proposed by [LZT95], a QRS complex is characterized by a pair of maximum moduli sets with different signs at scale 2^3, whose time interval is lower than 120 ms. If two minima, min_1 and min_2, are found within 120 ms of a maximum, we call their absolute values A_1 and A_2, and their distance to the maximum L_1 and L_2, and the following rules are applied:

1. If $\frac{A_1}{L_1} > 1.2 \cdot \frac{A_2}{L_2}$, min_2 is considered to be redundant
2. If $\frac{A_2}{L_2} > 1.2 \cdot \frac{A_1}{L_1}$, min_1 is considered to be redundant
3. If both are on the same side of the maximum, the minimum with the largest distance to the maximum is considered as redundant.

4. If they are on different sides, the minimum following the maximum is considered as redundant.

A symmetric procedure is applied if two minima are found in the neighborhood of a maximum.

Once a pair of sets is found (due to the rules above, the maximum moduli at scale 2^3 must have different signs), the R peak is found by looking for the zero-crossing between the 2 maximum moduli at scale 2^1. Again, as the scales are derivatives of the original signal, the zero-crossing corresponds to a peak, i.e. the R peak.

Thresholds

[MAO+04] computes the thresholds ϵ^k_{QRS} by taking the root mean square (RMS) over 2^{16} samples of scale 2^k (these results are then divided by 2 for scale 2^4). This is fine for an offline implementation, but as we want our algorithm to run in real-time, only a limited set of samples can be considered.

In the proposed algorithm, the RMS is first computed on each block of 512 samples, which corresponds to the size of our circular buffer (this size was chosen to support the limited memory of our embedded system, see Section 2.2.2.2), and the average over the last 8 blocks is considered. As in [MAO+04], the RMS value needs to be divided by 2 for ϵ^4_{QRS}.

A strong advantage of using a threshold computed over a limited amount of time is that the locality of the signal is considered, which is important for an implementation meant to be used in ambulatory conditions, as the noise, heart beat rate and other parameters may change quickly compared to records of the QTDB, which are usually rather stationary.

R-R interval

For some of the following rules, the current R-R interval (i.e. the interval between 2 consecutive R waves, denoted rr in the rest of this document), needs to be estimated. A moving average is used to perform this estimation (rr_i being the current estimate, and rr_{i-1} the previous one):

$$rr_i = \frac{3}{4} \cdot rr_{i-1} + \frac{1}{4} \cdot \max(\frac{rr_{i-1}}{2}, \min(rr_{i-1} \cdot 2, RR_{measured})) \qquad (4.27)$$

This allows the estimate to be sensitive to a change in the heart beat rate, without being too vulnerable to false detections, which would in turn produce an avalanche of other false detections due to the search back rule.

Additionally, the heart beat rate cannot take a value above 300 beats per minutes ([SL05], p. 415).

Search back

If no QRS complex is detected for $1.5 \cdot rr$, we look backwards and try to find a

pair of maximum moduli at scale 2^3, using the same threshold ϵ^3_{QRS}. I then directly look for the zero-crossing point at scale 2^1, without searching for maximum moduli at scales 2^2 and 2^1. This is somehow similar to [LZT95], except they lower the threshold ϵ^3_{QRS} to half its original value, while we do not perform this step, and keep the same ϵ^3_{QRS} values. This is however less sensitive than the normal rule above, since we do not ask for maximum moduli at scales 2^1, 2^2 and 2^4.

QRS delineation

Individual waves

Once the main wave of the QRS complex has been detected, I look for secondary waves of the QRS complex. We search for maximum moduli at scale 2^2 exceeding a threshold γ_{QRS} in a 160 ms search window centered on the main wave.

$$\gamma_{QRS} = 0.06 \cdot \max(|W_{2^2}x[n]|), \text{ with } n \text{ in the search window defined above.} \tag{4.28}$$

[MAO+04] uses 2 different thresholds to detect waves before and after the main peak. The proposed embedded implementation uses only one since it does not change the results, and allows to simplify the algorithm.

As for the main wave of the QRS complex, a secondary wave is characterized by 2 maximum moduli at scale 2^3 with different signs. And the peak is marked as the zero-crossing at scale 2^1 between these 2 maximum moduli.

[MAO+04] suggests to take only 3 waves into account, but does not say what to do if more are detected. In this implementation, if more than 3 significant waves are detected (i.e. the main wave plus 2 secondary waves), the closest to the main peak are selected.

Onset and end

The QRS onset and end are detected at scale 2^4. A QRS complex, at scale 2^4, is normally composed of a single pair of maximum moduli, while it may be composed of many maximum moduli at lower scales, because the signal gets smoothed down at higher scales, due to the lower cut-off frequency.

Given this, we look for the maximum moduli at scale 2^4 around the main QRS peak, and their values are used as thresholds:

$$\xi_{QRS_{onset}} = 0.25 \cdot \max(|W_{2^4}x[n]|), n \in [R - 80ms;\ R - 4ms] \tag{4.29}$$

$$\xi_{QRS_{end}} = 0.3 \cdot \max(|W_{2^4}x[n]|), n \in [R + 4ms;\ R + 80ms] \tag{4.30}$$

n_{max_on}, n_{max_end} are then defined, as the samples where the maximum moduli above occur.

The first and last maximum moduli at scale 2^2 associated with any of the QRS individual waves are also used. These values are named n_{first} and n_{last}, like [MAO+04].

To detect the QRS onset, we start from $\min(n_{max\_on}, n_{first} - 8 \text{ ms})$ and we go backwards until we find a value at scale 2^4 smaller than $\xi_{QRS_{onset}}$.

A similar rule is used to detect the QRS end, starting from $\max(n_{max\_end}, n_{last} + 8 \text{ ms})$, then going forward using $\xi_{QRS_{end}}$ as a threshold.

P and T waves detection

P and T waves detection and delineation follow the same procedure, except that their detection search windows and thresholds are different. As suggested in [MAO+04], we first look for all local maximum moduli at scale 2^4 within a search window exceeding a threshold $\epsilon^4_{P/T}$. We then only consider maximum moduli with amplitude greater than a threshold $\gamma_{P/T} = 0.125 \cdot \max(|W_{2^4} x[n]|)$, with n within the search window. If more than 1 pair of maximum moduli is detected, different rules are used for the P and T waves to determine which pair to take into account.

[MAO+04] takes multiple pairs into account, in order to be able to detect biphasic P and T waves. My experience with records in the QTDB ([LMGM97]) shows that using only the main pair is sufficient to delineate satisfactorily the wave peak, onset and end, even when the wave morphology is biphasic. Also, taking only one pair into account has some advantages when the P wave is faint (it is not uncommon to detect up to 4 pairs in that case), or when both T and U waves are present. If no pair is found at scale 2^4, the process is repeated at scale 2^5, using new thresholds ($\epsilon^5_{P/T}$).

Once the pair of maximum moduli is found at scale 2^k ($k = 4$ or 5), the zero-crossing at scale 2^3 between them is selected as the wave peak. If no zero-crossing is found at that scale, then the zero-crossing at scale 2^k is used.

Thresholds

[MAO+04] uses some thresholds $\epsilon^4_{P/T}$ computed over the last R-R detection to find if a modulus maximum is significant. Our observations suggest that, most of the time, these thresholds were very close to zero, and better results were obtained by simply considering every positive local maximum and negative local minimum.

This is logically equivalent to take:

$$\epsilon^4_P = \epsilon^4_T = 0 \tag{4.31}$$

The same reasoning was followed for $\epsilon^5_{P/T}$:

$$\epsilon^5_P = \epsilon^5_T = 0 \tag{4.32}$$

Onset and end

The onset and end of the P and T waves are detected on the scale 2^k where the relevant pair of maximum moduli above is found [MAO+04]. To detect the onset, we start from the first element of the pair of maximum moduli (n_{first}), and go backwards until we find a local minimum modulus, or until the value goes below a threshold $\xi_{P/T_{onset}}$, whichever comes earlier.

For the end, we start from the second element (n_{last}), and go forward using the same rules, but with other thresholds $\xi_{P/T_{end}}$.

These thresholds are defined as follows [MAO+04]:

$$\xi_{P_{onset}} = 0.5 \cdot W_{2^4}x[n_{first}] \tag{4.33}$$

$$\xi_{P_{end}} = 0.9 \cdot W_{2^4}x[n_{last}] \tag{4.34}$$

$$\xi_{T_{onset}} = 0.25 \cdot W_{2^4}x[n_{first}] \tag{4.35}$$

$$\xi_{T_{end}} = 0.4 \cdot W_{2^4}x[n_{last}] \tag{4.36}$$

P detection

The search window for the P wave relatively to the QRS complex was not mentioned in [MAO+04]. We used as a base search window (BSW):

$$BSW = [\mathrm{QRS}_{\mathrm{onset}} - \min(340\ \mathrm{ms}, rr/2);\ \mathrm{QRS}_{\mathrm{onset}} - 8\ \mathrm{ms}] \tag{4.37}$$

When a P wave is detected, the positions of the surrounding maximum moduli, relatively to the R peak, are recorded in pr_1 and pr_2. A moving average with parameters 0.75/0.25 is used to smooth down the values, allowing us to be flexible but resilient to detection errors. For the next searches, we look for the P wave at scale 2^4 using a narrowed search window (NSW):

$$NSW = BSW \cap [\mathrm{R}_{\mathrm{peak}} - 1.5 \cdot pr_1;\ \mathrm{R}_{\mathrm{peak}} - 0.6 \cdot pr_2] \tag{4.38}$$

If the P wave is not found at scale 2^4, we look for it again at scale 2^5, using the same window. If the P wave is still not found, we search again using a wider window (NSW'):

$$NSW' = BW \cap [\mathrm{R}_{\mathrm{peak}} - 2.0 \cdot pr_1;\ \mathrm{R}_{\mathrm{peak}} - 0.5 \cdot pr_2] \tag{4.39}$$

These rules improves the results significantly when the P waves are hard to distinguish.

If multiple maximum moduli pairs are found within the search window, the one with the maximum slope is selected

$$|(W_{2^k}x[n_1] - W_{2^k}x[n_2])/(n_1 - n_2)|, \tag{4.40}$$

with n_1 and n_2 being the 2 maximum moduli of the pair. The maximum slope usually corresponds to the most important wave in terms of amplitude.

T detection

As for the P wave, the search window was not stated in [MAO+04]. We decided to use a search window SW defined as follows:

$$SW = \Big[\max\left(\text{QRS}_{\text{end}} + 60 \text{ ms}, \text{QRS}_{\text{onset}} + (0.5 \cdot \text{QTc}_{\text{min}} \cdot \sqrt{rr})\right), \\ \text{QRS}_{\text{onset}} + (1.5 \cdot \text{QTc}_{\text{max}} \cdot \sqrt{rr})\Big] \tag{4.41}$$

The values for QTc_{min}, QTc_{max} can be found in Table 4.3, and are based on normal values for the QT coefficient in Bazet's formula[RGJ+09]:

$$QTc = \frac{\text{time interval from } QRS_{onset} \text{ to } T_{end}}{\sqrt{\text{previous R-R interval}}} \tag{4.42}$$

Table 4.3: Normal values of the QTc, as found in [RGJ+09]. HBR = heart beat rate (per minute), RR = R-R interval (in seconds). Values in the last two columns correspond to normal maximum and minimum values of the QT coefficient

HBR	RR	QTc_{min}	QTc_{max}
40	1.5	0.41	0.51
50	1.2	0.38	0.46
60	1.0	0.35	0.43
70	0.86	0.33	0.41
80	0.75	0.32	0.39
90	0.67	0.30	0.36
100	0.60	0.28	0.34
120	0.50	0.26	0.32
150	0.40	0.23	0.28
180	0.33	0.21	0.25
200	0.30	0.20	0.24

Ascending or descending T waves do not cause a pair of maximum moduli, but just a single maximum modulus, generally well above the threshold. Thus, the following rule was added: if there is only one maximum modulus in the search window, and this maximum modulus is greater than $2 \cdot \epsilon_T^k$, that point is considered to be an ascending or descending T wave, and we look for the onset and the end of the wave from there.

If within the search window, at scale 2^4, there are more than one maximum modulus pair, and if the ratio between the slopes is less than 2, the pair with the biggest modulus maximum at scale 2^5 is chosen.

This rule enables us to distinguish the T wave properly when both T and U waves are present, the U wave being a second wave, after the T wave, with

similar characteristics and frequency content that can be visible under some conditions. In theory, this rule would prevent us from detecting biphasic T waves, but in practice, the end of the T wave is still detected correctly thanks to the rules concerning the end point.

Porting on the Shimmer™ platform

In addition to the previously described algorithmic adaptations, some optimizations to adapt the algorithm to the limited resources or the Shimmer™ platform are performed.

Considering the amount of memory available (10 kB), the maximum buffer size for the input data and each of the 5 wavelet transforms is 512 samples. Since a 16-bit integer needs 2 bytes, these data use 6 kB of RAM. If we take into account the rest of variables used by the code, the total amount of memory required by the algorithm is 7.2 kB bytes (72 % of the available memory).

Regarding the arithmetic operations, they have to use integers only, since there is no hardware floating-point unit in this node's CPU (MSP430). Using any kind of floating-point operations would lead the compiler to generate slow emulation code.

The most time-consuming parts of the algorithm are the low-pass and high-pass filters, since they have to be called 4 and 5 times respectively for each data sample. However, these filters are very simple, since they do not even require any division or hardware multiplications, and can be performed efficiently using only shift and addition operations. These time-critical functions were coded in assembly, which provides a gain of more than 20 % in terms of processing time, as seen in Table 4.4.

Table 4.4: Processing time needed to delineate lead 1 of the `sel302` record from the QTDB, i.e. 37 seconds of data, using either C or assembly filters.

Filter implementation	Time	Gain
C	2863 ms	-
MSP430 assembly	2244 ms	22 %

The compiler automatically uses the hardware multiplier to execute multiplications by factors other than powers of 2 (which can be replaced by simpler shift operations). Since no hardware divisor is available, the compiler provides software emulation for division (again, divisions by powers of 2 are replaced by shift operations).

Finally, the ϵ thresholds are computed using the RMS of data blocks, which requires a square root to be performed. I had to write the code for this operation, since there is no C function for integer square root. We could have used the function `sqrt`, which uses floating points, but this would require

many slow software floating point operations. A simple algorithm to perform the operation $\sqrt{n}$ is the Longhand square root algorithm [Rol87], using the following recursive formula:

$$x_{k+1} = \frac{1}{2}\left(x_k + \frac{n}{x_k}\right), k \geq 0, x_0 > 0 \tag{4.43}$$

until it converges to a number. In addition, division is expensive on the MSP430, as it has to be performed by software. To avoid unnecessary steps, the starting value x_0 is first roughly estimated using the order of magnitude of the input number n.

4.2.2.3. Validation and experimental results

For this first set of experiments, I run the improved online single-lead delineation algorithm on the Shimmer™ platform and compute the four relevant performance metrics described in Section 4.2.1. Table 4.5 reports the obtained results, together with those of the original offline single-lead algorithm [MAO+04]. The last column indicates the standard deviation tolerances by the CSE working party [fQEWP85]. Note that for the sack of consistency with the validation procedure of [MAO+04], the results of our online algorithm in Table 4.5 correspond to performing online single-lead delineation on the two channels of each QTDB excerpt, then selecting the automatic annotation on the channel producing less error, for each manual annotation of the database. The results obtained using this validation procedure represent the optimal single-lead delineation that can be achieved using a 2-channel ECG input signal (*"genie"* selection).

Table 4.5 illustrates that our implementation performs within the standard deviation tolerances for all points, except the QRS_{onset} and P_{onset}, for which we are above the agreed tolerances by only a fraction of the sample duration. Furthermore, Table 4.5 shows that this implementation of our proposed algorithm consistently outperforms the state-of-the-art offline algorithm proposed in [MAO+04] in terms of standard deviation, while running in real time and preserving a high sensitivity and predictivity.

In addition, we measured the real-time performance of our algorithm implementation on the Shimmer™ platform. To this end, in order to emulate the behavior of the real system, in which the ECG signal is read by the sensors, all the data of the QTDB is sequentially sent from a PC to the Shimmer™ through the serial port in small blocks of 512 kB. The signal is processed in the platform and the results are sent back to the PC using also the serial port to enable a visual verification of the performed automatic ECG delination. In addition, we used an internal timer of the MSP430 to compute the total processing time. This timer is suspended when data transfers are performed through the serial port, and resumed when the data is being processed, as if it was in real time. Thus, we can extract a ratio

Table 4.5: Single-lead delineation performance of our online optimized implementation vs. the baseline offline algorithm [MAO+04]

Method	this work 16-bit int Shimmer™	Martinez [MAO+04]	Tolerances ($2s_{CSE}$)
Parameter	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	σ (ms)
P_{onset}	99.87 91.98 8.6 ± 11.2	98.87 91.03 2.0 ± 14.8	10.2
P_{peak}	99.87 92.46 10.1 ± 8.9	98.87 91.03 3.6 ± 13.2	-
P_{end}	99.91 91.70 0.9 ± 10.1	98.75 91.03 1.9 ± 12.8	12.7
QRS_{onset}	99.97 98.61 3.4 ± 7.0	99.97 N/A 4.6 ± 7.7	6.5
QRS_{end}	99.97 98.72 3.5 ± 8.3	99.97 N/A 0.8 ± 8.7	11.6
T_{peak}	99.97 98.91 3.7 ± 13.0	99.77 97.79 0.2 ± 13.9	-
T_{end}	99.97 98.50 −2.4 ± 16.9	99.77 97.79 −1.6 ± 18.1	30.6

between the amount of data to process and the actual time needed to process the data of the ECG waves of the QDTB. Interestingly, the obtained processing ratio results using the proposed algorithm and embedded implementation prove that we can perform real-time delineation while keeping the MSP430 microcontroller in sleep mode about 93 % of the time on average. Thus enabling a very low-power embedded ECG wave delineation system.

4.2.3. Single-lead MMD-based ECG delineation

A second widely used approach for ECG delineation is based on the multiscale morphological derivative (MMD) [SCK05]. As a nonlinear filtering

technique, it has been proven that the morphological dilation and erosion operations that involve the MMD, satisfy the causality and the additive semigroup property required by multiscale analysis [PL96, JD96, BS93] for signals of any dimension with local maxima and local minima as singular points. The fiducial points in ECG signal, such as the Q wave, R peak, S wave, the onsets and ends of the P and T waves, can be regarded as such singular points [MZ92a, Wit84].

4.2.3.1. State-of-the-art single-lead offline delineation algorithm

The original algorithm for offline ECG signal delineation [SCK05] is divided in three main steps (filtering, transform and characteristic wave delineation), which are subsequently explained in detail.

Filtering
The sensed ECG signal is filtered for baseline wander suppression and noise reduction. To this end, the same morphological filter [SCK02] already presented in Section 4.1.2 is used, which conditions the input signal for a better detection of the fiducial waves with minimum signal distortion. For this delineation algorithm, baseline correction is critical, since the elimination of baseline wander can change dramatically the information extracted from the ECG signal.

Application of the MMD transform
According to [SCK05], the singular points of the ECG wave (onsets, peaks and ends of the QRS complex and P and T waves) correspond to maxima and minima of the signal. Therefore, a singular point is defined as a point whose derivatives on the left and right exist with different signs.

The derivative on the right at point x of function/signal f can be defined by the morphological sup-derivative $M_f^+(x)$, defined as:

$$M_f^+(x) = \lim_{s \to 0} \frac{(f \oplus g_s)(x) - f(x)}{s}, \tag{4.44}$$

and the derivative on the left can be similarly defined by the morphological inf-derivative $M_f^-(x)$, defined as:

$$M_f^-(x) = \lim_{s \to 0} \frac{f(x) - (f \ominus g_s)(x)}{s}, \tag{4.45}$$

where for the functions $f : D \subset \mathbb{R}^n \to \mathbb{R}$ and $g_s : G_s \subset \mathbb{R}^n \to \mathbb{R}(s > 0)$, the two fundamental operations of multiscale morphology are defined in the same way as in 4.2.4, as:

$$Dilation : (f \oplus g_s)(x) = \sup_{t \in (G_s \cap D_x)} \{f(x - t) + g_s(t)\} \tag{4.46}$$

$$Erosion : (f \ominus g_s)(x) = \underset{t \in (G_s \cap D_x)}{\text{ínf}} \{f(x+t) - g_s(t)\} \tag{4.47}$$

where D_x is the translation of $D, D_x = x + t : t \in D$, $\sup(f)$ and $\inf(f)$ refer to the supremum and infimum of f, s is scale and g_s is the scaled structuring function [DS88]. In the discrete case, where the function is a finite set of points , $\max(f)$ and $\min(f)$ are used instead of $\sup(f)$ and $\text{ínf}(f)$.

The proposed multiscale morphological derivative transform M_f^d [SCK05] is defined as the derivative on the right minus the derivative on the left:

$$\begin{aligned} M_f^d(x) &= M_f^+(x) - M_f^-(x) \\ &= \underset{s \to 0}{\text{lím}} \frac{(f \oplus g_s)(x) - f(x)}{s} - \underset{s \to 0}{\text{lím}} \frac{f(x) - (f \ominus g_s)(x)}{s} \\ &= \underset{s \to 0}{\text{lím}} \frac{(f \oplus g_s)(x) + (f \ominus g_s)(x) - 2f(x)}{s} \end{aligned} \tag{4.48}$$

Therefore, the scaled version of M_f^d at scale s, $M_f^{d_s}$ is defined as:

$$M_f^{d_s}(x) = \frac{(f \oplus g_s)(x) + (f \ominus g_s)(x) - 2f(x)}{s} \tag{4.49}$$

This means that a positive peak in the ECG signal (its left derivative is positive and its right derivative is negative) is transformed into a local minimum. At the onsets and ends of the positive waves, there is an abrupt increase of the derivative from the left to the right, hence, these points are transformed into local maxima.

For a discrete signal, if we choose a flat structuring element of amplitude 0 ($g_s(x) = 0, x \in G$, where $G = \{x : \|x\| \leq s\}$, as stated in [JD96]), $M_f^{d_s}$ can be simplified by choosing a moving window of length $2s+1$ samples and finding the maximum and minimum values in the window, as well as the value of the signal at the central point $f(x)$. By applying this, the multiscale morphological derivative transform of a discrete signal f at scale s, $M_f^{d_s}$, is defined as:

$$M_f^{d_s}(x) = \frac{\max\{f(t)\}_{t \in [x-s, x+s]} + \min\{f(t)\}_{t \in [x-s, x+s]} - 2f(x)}{s} \tag{4.50}$$

To accurately and robustly detect the relevant signal singularities, the scale of the transform (s) should be as large as possible but less than $T_w F_s$, where T_w is the duration of the shortest ECG characteristic wave we want to detect and F_s is the sampling frequency of the signal. The shortest wave of interest is the QRS complex, whose duration is usually between 0.06 s and 0.12 s. Since the sampling frequency used in this work is 250 Hz, the

scale of the transform should be between 15 and 30. In this study we choose the lowest value, 15, in order to detect correctly all the waves, even if their duration is very short. If we choose a higher value of s, the transformed signal will be smoother and the bounds of the waves will be more difficult to detect, specially for waves with a very short duration as the QRS complex. Figure 4.12 depicts the original signal (a), the signal after the morphological filtering (b) and the result of applying the morphological derivative to the filtered signal (c).

ECG wave delineation
The delineation of the fiducial points of the ECG signal is performed only taking into account the transformed signal. This delineation detects the local minima and maxima of the transformed signal, since, as aforementioned, the MMD transform converts the singular points of the original ECG signal into local maxima and minima.

Limitations
This approach causes misdetections due to the fact that, in some cases, a fiducial point of the original ECG signal is not transformed into a maximum or minimum, as depicted in Figure 4.13. This figure represents an excerpt of the original ECG signal and its transform, and illustrates the wrong detection of the end of the QRS complex by this offline algorithm and the real position of this point. The procedure followed by the algorithm to detect the end of the QRS complex is to search for a local minimum starting from the right of the R peak (vertical solid line). Therefore, it is incorrectly detected in the point indicated by the rightmost vertical dotted line. The actual end of the QRS complex is indicated by the leftmost vertical dotted line.

4.2.3.2. Real-time embedded MMD-based delineation

The limitations of the original offline MMD algorithm are due to the fact that, very often, at a fiducial point of the wave there is an abrupt change of the derivative, but not necessarily a local maximum or minimum. Then, a more efficient and accurate way of detecting these fiducial points is to look for abrupt changes in the slope of the transformed signal instead of looking for the local maxima and minima.

Furthermore, as was stated in the introduction, the aim of this work is to run this algorithm for ECG wave delineation on an embedded wearable platform. Due to the memory restrictions in this kind of platforms, only a small excerpt of the ECG record can be stored for its subsequent processing. On the other hand, we need as much data as possible to accurately calculate the thresholds that will be used in the delineation process. Taking into account both a high delineation accuracy and a low memory usage to make

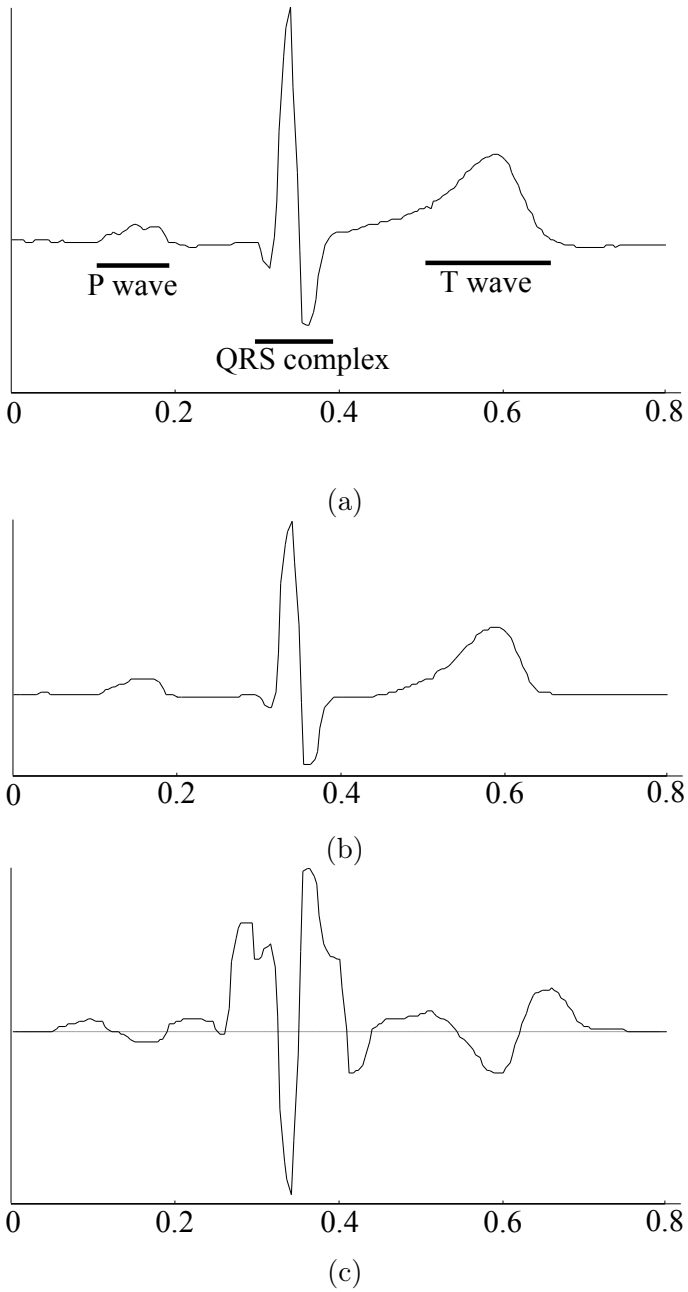

Figure 4.12: (a) Original ECG input signal (b) Signal after morphological filtering (c) Result of applying the multiscale morphological derivative transform at scale 15 to the filtered signal

the algorithm suitable for the node, the length of the sliding window needed to keep the last samples of the signal has been experimentally chosen to be 500. Hence, our implementation has been adapted to use a circular buffer of

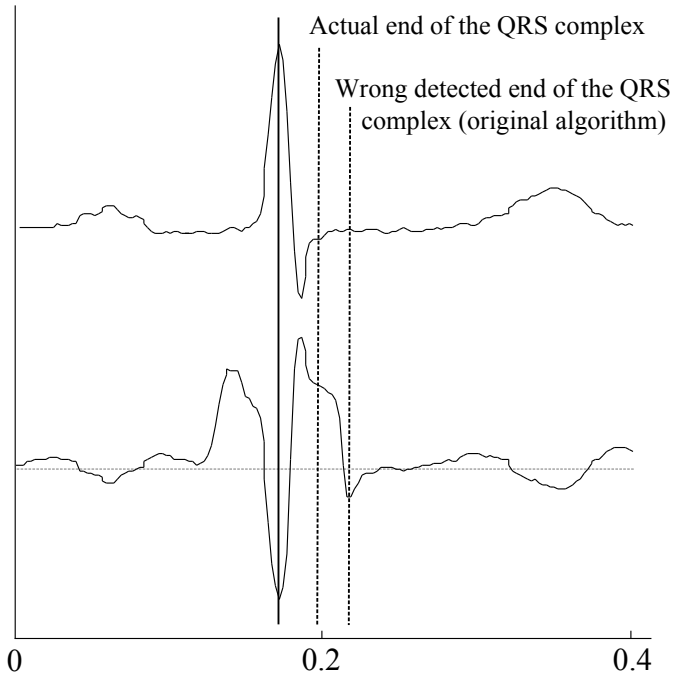

Figure 4.13: Misdetection of the end of the QRS complex due to the absence of a clear minimum

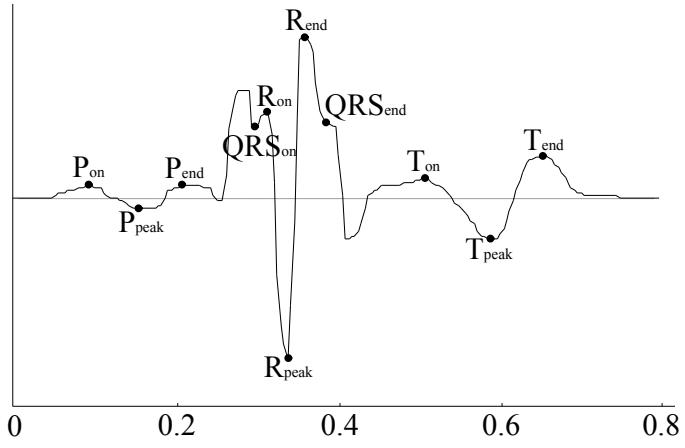

Figure 4.14: Detection of all the fiducial points of a complete ECG wave

only 1000 bytes, which contains the last 500 samples of the read ECG signal. Since the sampling frequency used in this work is 250 Hz, 2 s of signal are stored in the node.

Thresholds definition

In order to find all the characteristic points of the original ECG signal (onsets, peaks and ends of the P wave, QRS complex and T wave), we have to find the points in the transformed signal where there is a sudden change in the slope. Figure 4.14 represents an example of our delineation process of a complete ECG wave. First, the algorithm calculates a set of thresholds which will be used to find the points that correspond to the R peak (Th_R), the bounds of the Q and S waves (Th_{QS}), the bounds and peak of the P wave (Th_P) and the bounds and peak of the T wave (Th_T). To this end, the algorithm makes two histograms of the transformed signal. The first one, which takes into account the complete excerpt of signal that is stored in the buffer, and finds the maximum absolute amplitude and the most dense parts of the signal. Th_R and Th_{QS} are values between them. The second histogram considers only the part of the signal with more density, where the P and T waves appear. The higher resolution in the histogram achieved by taking into account only this part of the signal, allows to find more accurately Th_P and Th_T.

ECG detection process

The first point to be found is the R peak, which appears in the transformed signal as a local minimum with absolute amplitude greater than Th_R. As illustrated in Figure 4.14, at the left of the R peak, the algorithm looks for two points featuring sudden changes in the slope and with absolute amplitude greater than Th_{QS} (onset of the R and Q waves, respectively). At the left of the QRS complex, the algorithm looks for two zero-crossings. The absolute minimum between these two zero-crossings is the P peak if its absolute amplitude is greater than Th_P. The first local maximum, from the left of the zero-crossings, with absolute amplitude greater than Th_P is the onset of the P wave and the first local maximum, from the right of the zero-crossings, with absolute amplitude greater than Th_P is the end of the P wave.

The same procedure is followed at the right of the R peak to find the end of the R wave, the end of the S wave and the onset, peak and end of the T wave.

Next, I explain in more detail the general method for the detection of the characteristic points of the ECG wave on the embedded sensor node. Once the buffer, which stores the last 500 samples of the signal, is full for the first time, the algorithm calculates the thresholds (Th_R, Th_{QS}, Th_P and Th_T) based on the histogram of the signal, as we previously introduced. The calculation of the thresholds is repeated every time a R peak is detected. Once the thresholds are set, the algorithm proceeds as follows.

- Starting from the last detected point (or from the beginning of the buffer if it is the first time that the buffer is completely filled), the algorithm searches for a local minimum with absolute amplitude larger

than Th_R. This point is detected as an R peak. Depending on the placement of the electrodes in the patient, we can obtain an ECG signal in which the R peak is inverted, this behavior is considered by the algorithm. In this case, the detected minimum will be either the onset or the end of the R wave and the algorithm will find the R peak, that will be a local maximum with amplitude larger than Th_R.

- Once the R peak has been detected, the algorithm searches back trying to find the onset of the R wave, the onset of the Q wave and the P wave. Then, starting again from the R peak, it will search forward for the end of the R wave, the S wave and the T wave.

- From the R peak to the left, the algorithm looks for a sudden change in the slope of the transformed signal. This point will be detected as the onset of the R wave if its absolute amplitude is larger than Th_{QS}. From the onset of the R wave to the left, again it looks for another sudden change in the slope. This is considered as the onset of the Q wave if its absolute amplitude is larger than Th_{QS}.

- From the onset of the Q wave to the left, the algorithm looks for a big reduction in the slope of the signal and afterwards for a zero-crossing from a positive to a negative point. The algorithm then looks for a second zero-crossing, now from a negative point to a positive one. The minimum point between these two consecutive zero-crossings is considered to be the peak of the P wave if its absolute amplitude is greater than Th_P. The first local maximum from the left of the left zero-crossing with absolute amplitude greater than Th_P is detected as the onset of the P wave. Then, the first local maximum from the right of the right zero-crossing with absolute amplitude greater than Th_P is detected as the end of the P wave.

- If a valid P wave is not found, the step is repeated starting from the second zero-crossing that was detected to the left.

- From the R peak to the right, the algorithm looks for a sudden change in the slope of the transformed signal. This point will be detected as the end of the R wave if its absolute amplitude is larger than Th_{QS}.

- From the end of the R wave to the right, again it looks for another sudden change in the slope. This is considered as the end of the S wave if its absolute amplitude is larger than Th_{QS}.

- From the end of the S wave to the right, the algorithm looks for a big reduction in the slope of the signal and afterwards for a zero-crossing from a positive to a negative point. The algorithm then looks for a second zero-crossing, now from a negative point to a positive one.

The minimum point between these two consecutive zero-crossings is considered to be the peak of the T wave if its absolute amplitude is greater than Th_T. The first local maximum from the left of the left zero-crossing with absolute amplitude greater than Th_T is detected as the onset of the T wave. Then, the first local maximum from the right of the right zero-crossing with absolute amplitude greater than Th_T is detected as the end of the T wave.

- If a valid T wave is not found, the step is repeated starting from the second zero-crossing that was detected to the right.

In order to implement this algorithm for its execution in the Shimmer™ platform and make the final system to operate in real time, two types of optimizations need to be performed in this algorithm: suppression of the floating-point operations and changes in the underlying parameters of the algorithm.

Suppression of floating-point operations

Since there is no hardware support for floating-point operations, all the floating-point operations contained in an application compiled for one of these platforms will be replaced by software emulation code, which leads to a very long execution time of the application. For this reason, the first optimization that has to be carried out is replacing all the floating-point operations by integer operations, which can be quickly executed in the target microcontroller.

Changes in the underlying parameters

The execution time of the overall algorithm can be further reduced by choosing intelligently the parameters used in the filtering, morphological transformation and other intermediate operations, so that the algorithm makes as little as possible use of arithmetic operations for which there is not specific hardware support. Two operations of this type that are intensively used are the division and the modulo.

The division is used mainly to calculate the morphological derivative and the histogram of the signal needed to set the thresholds. Then, most of the divisions, which were manually optimized to use a power-of-two divisor, are translated by the compiler into shifts that can be performed extremely quickly.

As an example, the scale of the morphological derivative, that was initially selected to be 15, was changed to 16, which is also a very low value in the range from 15 to 30. The number of ranges used in the histogram to calculate the thresholds was also changed, being reduced from 20 to 16.

As described before, the algorithm uses a circular buffer to store the last samples of the ECG signal, as well as two more circular buffers for the filtered

signal and the signal after the MMD transform, respectively. The algorithm uses several pointers to access the data stored in the buffers, everytime one of these pointers is moved, a modulo operation is used to make sure that the pointer is always within the bounds of the buffers. The length of the buffers was changed from 500 to 512. Thus, the modulo operation can be transformed into a simple mask operation and executed much faster.

The impact of these changes in terms of delineation accuracy is negligible, as confirmed by the validation results for a representative embedded sensor node, while they help to significantly speed up the execution time of the algorithm.

4.2.3.3. Validation and experimental results

For the validation of this work, I run on a WBSN platform the proposed online version of the multiscale morphological derivative transform-based detector using as input the 105 records of the QTDB [LMGM97]. In this evaluation, we compare the proposed improved algorithm with the original MMD detector, as shown in Table 4.6.

Validation on a PC

The first experiment I run is the validation using the online algorithm on a PC, with floating-point operations in order to have the highest achievable delineation accuracy. The first column of Table 4.6 shows the results of this validation. The values in the last column of the table are the limits in the standard deviation that are set by the CSE committee [fQEWP85]. All the results of our algorithm are below these limits, unlike the original MMD algorithm (third column of Table 4.6), which violates the threshold given by the CSE committee for the end of the P wave (row P_{end}). Thanks to the improvements that have been included in this new version of the algorithm, the mean error has been significantly reduced (by 4 ms on average, if we calculate it using the absolute value of the mean error in every detection). This means that, on average, the points detected by the proposed algorithm are 4 ms (i.e., one sample) closer to the real ones than the ones detected by the original offline algorithm.

The standard deviation is also improved with respect to the original MMD detector for the case of the P wave and the QRS complex. In the case of the T wave (rows T_{onset} and T_{end}), the standard deviation obtained by our detector is higher than in the original implementation, but this is not a problem since the results obtained by our algorithm are significantly lower than the limits given by the CSE committee (last column of Table 4.6), because the T wave is not decisive in the diagnosis of cardiac pathologies.

Table 4.6: Results of our algorithm (both in the PC and the Shimmer™ platform) and other state-of-the-art algorithms for ECG signal delineation

Method	Improved MMD floating-point PC	Improved MMD 16-bit int Shimmer™	Original MMD	Tolerances ($2s_{CSE}$)
Parameter	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	σ (ms)
P_{onset}	99.75 96.66 3.6 ± 9.6	99.15 96.70 3.5 ± 10.7	97.2 - 9.0 ± 9.4	10.2
P_{peak}	99.91 96.67 2.6 ± 7.7	99.28 96.65 1.2 ± 7.8	- - -	-
P_{end}	98.56 96.98 2.2 ± 9.4	98.68 96.95 1.1 ± 10.0	94.8 - 12.8 ± 13.2	12.7
QRS_{onset}	99.37 100.0 1.0 ± 6.4	99.03 100.0 0.7 ± 6.7	100.0 - 3.5 ± 6.1	6.5
QRS_{end}	99.89 99.97 −0.4 ± 6.8	99.78 99.97 −0.2 ± 7.0	100.0 - 2.4 ± 10.3	11.6
T_{onset}	95.11 83.42 5.7 ± 26.7	96.10 83.96 8.7 ± 25.8	99.8 - 7.9 ± 15.8	-
T_{peak}	99.75 100.0 −0.3 ± 15.0	99.60 100.0 0.7 ± 13.1	- - -	-
T_{end}	95.23 97.92 −7.3 ± 19.0	95.37 97.80 −6.0 ± 17.0	99.6 - 8.3 ± 12.4	30.6

Validation on the Shimmer™ Platform

After performing all the previously described optimizations for embedded sensor platforms, the algorithm was validated in the Shimmer™ platform, using as an input the whole QTDB, which is sent from a PC to the Shimmer™ through the serial port in small blocks of 512 kB as in Section 4.2.2.2. The second column of Table 4.6 shows the results of the algorithm running on the Shimmer™ platform. As it is shown, the degradation in the results due to the loss of accuracy of the optimizations performed to make the algorithm suitable for the embedded platform is negligible. Regarding the mean error, the same values, on average, are obtained in the version of the algorithm for the Shimmer™ (using integer operations and the optimized underlying

parameters previously described) and in the version of the PC (using double-precision floating-point operations and the initial underlying parameters). The sensitivity is slightly better, on average, for the version running in the Shimmer™ than for the one running in the PC.

4.2.4. Multi-lead WT-based ECG delineation

In clinical practice, however, several leads are simultaneously acquired, be it the standard 12-lead ECG in traditional clinical settings or the 3-lead configuration in emerging ambulatory ECG monitoring. This calls for the deployment of delineation approaches able to exploit the multiple available leads to improve the accuracy, stability and resilience to artifacts of the characteristic waves measurements, compared to single-lead delineation.

There are two approaches to multi-lead delineation. The first approach consists in first delineating each and every lead individually, then applying a median postprocessing selection rule over the characteristic wave peaks, to select the "best" delineated ones. Then, the wave boundaries with longest wave duration are chosen, with some outlier protection, in order to retain the "best" global single-lead delineation [LJC94, MAO+04]. When only two leads are available, this approach is however not applicable, and must be replaced by a "genie" selection that chooses for each point the lead with less error. However, this "genie" selection, which represents the optimal delineation, can only be used for validation purposes since the algorithm cannot know which is the lead with less error. The second approach instead proposes to combine the individual leads into a multi-lead signal – which provides an overall view of the cardiac phenomena and is independent of the lead system used – on which a single-lead delineation algorithm is further applied. Different multi-lead signals can be found in the literature: (1) the combination of the ECG leads in the root mean squared (RMS) sense [IvOH06, SML06, RBB+09]; (2) a derived spatial lead obtained via a multi-step iterative search on the spatial WT loops of the 2 or 3 original orthogonal leads [AMRL09]; (3) a combination of the derivatives of the ECG leads [IM10]. Given the limited processing and storage resources of our target embedded sensor node, we herein retain the well-established RMS-based multi-lead delineation approach due to its lower computational complexity and better scalability. We will additionally benchmark its performance with respect to the selection-based multi-lead delineation.

Accordingly, the available L ECG leads $x_l[n]$, with $l = 1, \ldots, L$, are first combined in a single multi-lead signal $x_{RMS}[n]$:

$$x_{RMS}[n] = \sqrt{\frac{1}{L}\sum_{l=1}^{L} x_l^2[n]}, \tag{4.51}$$

where n denotes the discrete-time index, on which single-lead delineation is

performed. For a meaningful combination, it is crucial to remove baseline wander (mainly caused by respiration, electrode impedance changes due to perspiration and body movements) on each of the leads before computing the RMS [IvOH06, SML06]. Therefore, since the quality of the subsequent delineation depends on the baseline wander correction, we assess in this work the effectiveness of the two state-of-the-art approaches previously studied in Section 4.1: cubic spline baseline estimation and morphological filtering.

The delineation of the signal x_{RMS} is performed using the single-lead WT-based ECG delineator proposed in Section 4.2.2, since it clearly outperforms the single-lead MMD-based ECG delineator in terms of execution time, therefore achieving a lower energy consumption, while having a high accuracy in the detection of the ECG characteristic points.

4.2.4.1. Validation and experimental results

The accuracy of our online multi-lead delineation using the two considered baseline wander removal techniques, running on the Shimmer™ platform for the QTDB, are compiled in Table 4.7. For comparison, this table also includes the delineation results of the online single-lead delineation algorithm applied on each of the two leads of the database.

Comparing the accuracy of single-lead delineation on individual leads (third and fourth columns of Table 4.7) and that of the multi-lead RMS approaches (first and second columns of Table 4.7), it is confirmed that exploiting the multiple leads has the potential to provide more accurate boundary locations than any lead by itself. The two considered baseline removal techniques exhibit equivalent accuracy figures in terms of mean (m) and standard deviation (σ), which are off the tolerances by only a fraction of the sample duration for all wave locations, except the onset of the P wave. For this particular wave boundary, none of the methods discussed in this section fulfills the loose criterion on the tolerance for the dispersion of automatic delineation errors. Interestingly, multi-lead delineation with cubic spline baseline estimation offers the best positive predictivity performance, P^{+}_{min}, among all the methods of Table 4.7.

Furthermore, Table 4.7 also shows that multi-lead delineation performs equivalently to the best single-lead delineation (i.e., obtained on lead 1), and outperforms the worst single-lead delineation (i.e., obtained on lead 2) by one sample duration. This further confirms the previously introduced evidence [SML06] that RMS-based lead combining alleviates the dependency of delineation results on specific leads or lead systems, and that it provides a more robust and stable boundary locations than single-lead delineation. The validation herein produced is limited to 2-lead ECG excerpts of the QTDB database, due to the memory limitations of the target embedded sensor platform. This validation should, however, be extended to higher dimensional excerpts to fully evaluate the performance and advantage of multi-lead deli-

Table 4.7: Performance comparison of online multi-lead and single-lead delineation

Method	2-lead 16-bit int Cubic spline	2-lead 16-bit int Morpho.	Single-lead 16-bit int Lead 1	Single-lead 16-bit int Lead 2	Toleranc. ($2s_{CSE}$)
Param.	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	Se (%) P^{+}_{min} (%) $m \pm \sigma$ (ms)	σ (ms)
P_{on}	94.02 95.15 4.1 ± 17.1	96.24 91.19 2.5 ± 16.4	98.84 92.66 13.4 ± 14.8	97.31 91.76 10.4 ± 19.4	10.2
P_{peak}	94.02 95.42 12.8 ± 12.1	96.24 91.38 14.7 ± 13.4	98.84 92.93 15.6 ± 12.1	97.09 92.21 6.9 ± 16.9	-
P_{end}	94.05 95.49 −2.1 ± 14.2	96.27 91.55 −1.9 ± 15.5	98.87 93.24 1.6 ± 13.2	96.81 91.43 −8.5 ± 17.9	12.7
QRS_{on}	99.67 99.23 3.2 ± 8.7	99.75 97.07 6.9 ± 8.0	99.61 99.56 5.4 ± 8.4	99.67 98.61 8.6 ± 12.6	6.5
QRS_{end}	99.67 99.20 7.3 ± 11.9	99.75 97.07 9.0 ± 9.1	99.61 99.56 1.5 ± 10.1	99.72 98.72 8.7 ± 13.1	11.6
T_{peak}	98.00 99.23 1.9 ± 17.4	98.33 98.58 5.3 ± 19.3	99.35 99.49 5.3 ± 18.4	99.35 98.90 3.5 ± 23.5	-
T_{end}	97.94 98.72 −4.1 ± 25.3	98.11 98.08 −11.4 ± 24.9	99.32 99.24 −5.3 ± 22.7	99.18 98.38 −4.6 ± 27.2	30.6

neation over its single-lead counterpart.

Additionally, Table 4.8 reports the memory usage and duty cycle on the Shimmer™ of the delineation algorithms validated in Table 4.7. In particular, it shows that RMS-based multi-lead delineation requires up to 2 times the memory usage of its single-lead counterpart. This is because it must filter the two leads separately for baseline removal, combine them, and finally delineate the resulting signal. Interestingly, both RMS-based implementations exceed the 10 kB RAM memory of the platform. The corresponding results in Table 4.8 have in fact been obtained using an alternative version of the MSP430 microcontroller (MSP430F5438), which is equipped with 16 kB of RAM memory. Moreover, the duty cycle of the RMS-based methods is considerably higher than that of single-lead delineation, with most of the computational burden dedicated to baseline wander removal (i.e., 26.44 % for cubic spline baseline estimation vs. 12.39 % for morphological filtering).

Table 4.8: Computational burden and memory footprint

Algorithm	RAM usage (kB)	Duty cycle (%)
Single lead	7.2	6.78
2-*lead RMS* *cubic spline filtering*	14.9	33.22
2-*lead RMS* *morph. filtering*	10.9	19.17

4.3. Arrhythmia diagnosis

The results of the previously studied ECG delineation algorithms can be used to perform autodiagnosis onboard the sensor node to provide the patient with information about his heart's state as well as significantly reduce the amount of data to be wirelessly transmitted, and therefore extend the lifetime of the node. To illustrate this, this section proposes a diagnosis algorithm that evaluates a set of rules that a normal ECG should meet. If any of the rules is not satisfied, the algorithm generates an alarm that will be sent to the WBSN coordinator. Moreover, this section proposes a more accurate approach, where the diagnosis algorithm is tailored to the detection of a specific cardiac arrhythmia. In this case, an an automated real-time atrial fibrillation (AF) detection algorithm is proposed, which also uses the output of an ECG delineator to perform a detailed analysis of characteristic irregularities of AF episodes.

4.3.1. General diagnosis algorithm

After processing the original ECG signal and obtaining the peaks and boundaries of the characteristic ECG waves, this diagnosis algorithm is applied to check if the signal presents any anomalous behavior, which may hint that the patient is suffering a cardiovascular pathology. This algorithm evaluates iteratively five different conditions based in the points detected in the ECG signal, according to the valid ranges reported in [Sch71, PG85], namely:

1. The time from QRS_{onset} to QRS_{end} must be less or equal to 0.10s.
2. The time interval from P_{onset} to QRS_{onset} must be in the range from 0.12s to 0.20s.
3. The amplitude of T_{peak} must always be positive.
4. The time from QRS_{onset} to R_{peak} must not be longer than 0.03s.

Table 4.9: Normal values of the QTc. HBR = heart beat rate (per minute), RR = R-R interval (in seconds).

HBR	RR	QTc *and normal limits*
40	1.5	0.46 (0.41 - 0.51)
50	1.2	0.42 (0.38 - 0.46)
60	1.0	0.39 (0.35 - 0.43)
70	0.86	0.37 (0.33 - 0.41)
80	0.75	0.35 (0.32 - 0.39)
90	0.67	0.33 (0.30 - 0.36)
100	0.60	0.31 (0.28 - 0.34)
120	0.50	0.29 (0.26 - 0.32)
150	0.40	0.25 (0.23 - 0.28)
180	0.33	0.23 (0.21 - 0.25)
200	0.30	0.22 (0.20 - 0.24)

5. The QT interval rule, which establishes a relation between the interval from QRS_{onset} to T_{end}. This rule indicates the valid interval between the heart beat rate, and the last R-R interval (i.e., the interval from the last R peak to the current one). To find the valid QT interval, Bazet's formula is used to calculate the QT coefficient for an input signal (QTc) as shown in Equation (4.52). Then, the valid values of the QTc are reported in Table 4.9.

$$QTc = \frac{\text{time interval from } QRS_{onset} \text{ to } T_{end}}{\sqrt{\text{previous R-R interval}}} \tag{4.52}$$

As a result, according to the results of checking these previous five conditions, the sensor node reports to the WBSN coordinator that the heart of the monitored patient can be suffering one of the pathologies of Table 4.10.

4.3.2. AF detection

Atrial fibrillation (AF) is the most common cardiac arrhythmia. It is associated with high risk of stroke, dementia, heart failure and death [Bea09], and is predicted to be responsible for over 3 million hospitalizations by 2025 [WC03]. AF occurs when disorganized electrical signals cause the heart's two upper chambers (called the atria) to contract very fast and irregularly. This results in a desynchronization between the atria and the lower chambers (called the ventricles), which leads to an inefficient pump of blood. The early detection of AF is important to ensure dangerous diseases to be identified and treated accordingly. However, it can be problematic since the AF can appear

Table 4.10: Pathologies detected by the diagnosis algorithm

Problem	*Possible pathology*
Time from QRS_{onset} to QRS_{endt} longer than 0.10s (it could even reach 0.12s)	Block of the His bundle. Supraventricular rhythm with aberrant conduction. Abnormal conduction over accesory pathways. Ventricular rhythm or pacemaker rhythm.
The time interval from P_{onset} to QRS_{onset} is longer than 0.20s (long PR interval)	Disorder in the conduction between atriums and ventricles at the atrioventricular node level, His bundle (or its branches) or Purkinje system.
The time interval from P_{onset} to QRS_{onset} is shorter than 0.12s (short PR interval)	Presence of an anomalous accesory pathway that produces a faster conduction or the presence of a rhythm with origin in the atrioventricular union, in the left atrium or in the lower part of the right atrium. Generally, this anomaly is due to a ventricular preexcitation.
The amplitude of T_{peak} is negative (negative T wave)	Primary alterations of the repolarization phase (due to ischemia or myocardial infarction, subacute pericarditis or myocarditis). Secondary alterations of the repolarization phase (due to alterations of the ventricular repolarization).
The time from QRS_{onset} to R_{peak} is longer than 0.03s	Delay in the ventricular activation time.
QTc is greater than the values specified in Table 4.9	The ventricular repolarization has slowed down, which can be due to acquired or congenital causes. It is related to the appearance of arrhythmias.
QTc is lower than the values specified in Table 4.9	This problem is usually related to the use of some medicines, hypercalcemia or hyperpotassemia.

only during certain periods of time and may be asymptomatic. To overcome this problem, a plethora of projects on telemedicine and health monitoring started in recent years: MyHearth [LSD+06], Health@Home [SSS+10], MobiHealth [vHBW+04], CodeBlue [KKG07], etc. The interest in this field outlines the importance of telemedicine and remote health monitoring for future

development of medical assistance and drives ICT services and technologies development to new frontiers [Bea08]. Thus, in this work we advocate the use of WBSN technologies, to enable continuous and advanced biomedical monitoring of the patients in real-time and therefore automated early diagnosis for better healthcare.

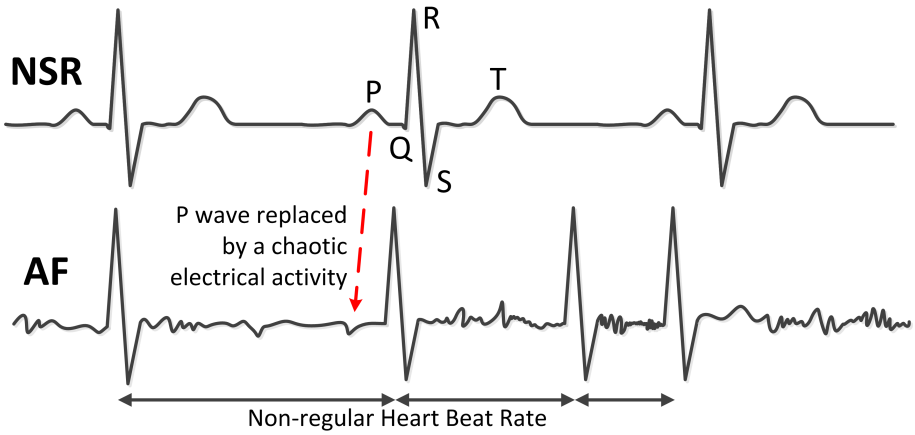

Figure 4.15: ECG signal during normal sinus rhythm (above) and an AF episode (below). During AF episodes, HBR is not regular and P wave is replaced by a chaotic electrical activity between consecutive QRS complexes

The main characteristics of AF are an irregular heart beat rate (HBR) and the absence of the P wave, which is replaced by a sawtooth pattern [Fea06]. This is observed in Figure 4.15, which shows an ECG signal during normal sinus rhythm (NSR, above), where the different ECG waves can be clearly distinguished and the HBR is regular, and an AF episode (below), where the HBR is completely irregular and the P wave is replaced by a chaotic electrical activity between consecutive QRS complexes. Many studies on the development of algorithms able to recognize these two features and therefore automatically detect the occurrence of an AF episode can be found in the literature [Cea97, TG00, LH05, Cea02, Cea08, SSS+10]. However, the implementation of such algorithms in a WBSN platform that can provide real-time monitoring and diagnosis is unheard of.

During the last two decades, several algorithms and techniques for the automatic detection and classification of AF episodes were presented. Such algorithms analyze the ECG signal and search for anomalies and peculiar characteristics of AF episodes.

When AF occurs, the cardiac frequency becomes highly irregular, leading to a dramatic increase of the R-R interval variance. R-R interval dynamics are analyzed by Cerutti et al. [Cea97] to detect AF episodes. This work proposes the use of linear and non-linear indexes for characterization of R-R series. Experimental results suggested a possible application of them for

automatic detection of AF episodes. Tateno and Glass [TG00] describe a method for automatic detection of AF episodes based on R-R intervals using the Kolmogorov-Smirnov test to estimate the similarities between standard and test density histograms. This technique is able to reach 93.2 % sensitivity and 96.7 % specificity. Logan et al. [LH05] propose an approach based on the variance of the R-R intervals over a sliding window, achieving 96 % sensitivity and 89 % specificity. However, these approaches require high computation and storage requirements, which do not make them applicable for WBSN platforms, since they work offline using the whole ECG signal or very long windows of it.

Other main characteristic of the AF episodes is the absence of the P wave. A different approach based on this feature is proposed by Clavier et al. in [Cea02], which presents a single-lead automatic P wave analysis of patients prone to atrial fibrillation. The P wave segmentation was done using a hidden Markov model by taking into account some statistical and electrophysiological properties of the signal. Even if the paper does not present very good results in terms of sensitivity and specificity, it suggests that combining various symptoms can be a key point for the implementation of accurate and efficient automatic detection algorithms. Concerning the combination of various AF symptoms, an interesting work was presented by Couceiro et al. in [Cea08]. In this paper, the authors combine three different algorithms to detect AF episodes: P wave detection, heart rate analysis and atrial activity analysis. The output of these algorithms is subsequently combined using a neural network classifier which is able to detect the presence of AF episodes. According to their measurements, the proposed algorithm outperforms existing techniques with 93.8 % sensitivity and 96.09 % specificity. However, the atrial activity analysis involves quite complex frequency-domain calculations that exceed the capabilities of state-of-the-art WBSN nodes [BGM+10]. Also, the neural network classifier is very expensive in terms of computing power and memory resources to be implemented in such limited WBSN platforms.

In addition, advances in electronics and development of portable and energy efficient devices for medical care allow the development of innovative applications for medical assistance and monitoring. Regarding AF detection, Sánchez-Tato et al. [SSS+10] address the necessity of an affordable monitoring and tele-healthcare system for patients. In particular, they present AF detection as one of the most relevant case studies for a diffusion of mobile healthcare systems to end users. Therefore, real-time detection of AF episodes in people with risk of cardiac failure is one of the main concern in current development of telemedicine.

4.3.2.1. Atrial Fibrillation Detection Approach

As aforementioned, the AF detection approach proposed in this work relies on two algorithms that perform, for each heart beat, an analysis of the

HBR and the detection of the absence of the P wave, respectively. Then, the outputs of both algorithms are combined using fuzzy logic to classify the heart beat under analysis as AF or normal. The scheme of the proposed technique, depicted in Figure 4.16, is similar to the one described in [Cea08], but avoids the complex atrial activity analysis and replaces the neural network classifier by a much simpler, but yet highly accurate, fuzzy classifier. These changes allow the algorithm to meet the limited resources of the Shimmer™, especially in terms of memory and processing power. This section provides a detailed description of these algorithms as well as our new fuzzy classifier.

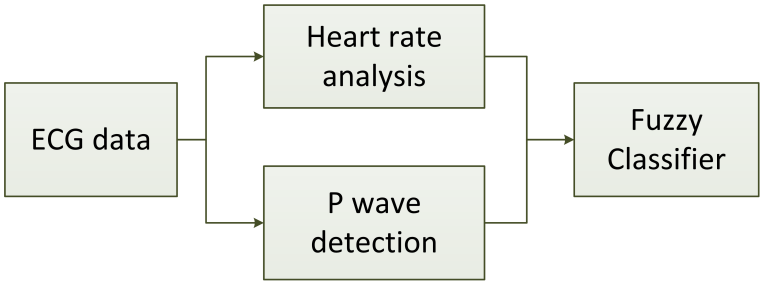

Figure 4.16: Overview of the proposed AF detector algorithm

Heart Rate Analysis

Heart rate analysis is based on the *R-R interval variance*, which is higher during AF episodes due to the irregular activity of the heart. First we need to detect the most significant peaks of the ECG signal (i.e., the R peaks), that are related to the contraction of the myocardium. For this purpose, we use as starting point the delineation algorithm introduced in Section 4.2.2.2. This algorithm performs wavelet-based single-lead detection of the major ECG characteristic waves (i.e., the QRS complex, P and T waves), i.e., ECG delineation.

Then, when the R peaks are detected, we calculate the time difference between two consecutive peaks to obtain the R-R intervals. Next, we normalize the R-R intervals according to the following equation:

$$RR_{norm} = \frac{RR}{\bar{RR}} * 100; \tag{4.53}$$

where RR is the current R-R interval, $\bar{RR} = \frac{2}{3}\bar{RR} + \frac{1}{3}\bar{RR'}$ and $\bar{RR'}$ is the average of the R-R intervals in a 10-second window centered in the current beat. This formula is inspired by the feature normalization proposed in [MM83], and updates $\bar{RR}$ with local information about the current R-R interval, compensating for different patient resting HBR. The variance of RR_{norm} is then computed over a 10-second sliding window.

At this point, [LH05] applies hard thresholding to the variance in order to compute an initial AF detection. This initial classification is then smoothed

to eliminate spurious errors using a majority voting scheme over a 600-beat window. Instead, in order to reduce the computational complexity and storage requirements to be able to run in real-time in WBSN nodes, we provide the variance to our newly developed fuzzy classifier, that will interpret this data accordingly. Moreover, this new approach we propose enables a much faster diagnosis, with an average delay to classify a beat of only 5 seconds, since it uses a 10-second sliding window centered in the current beat, while the average delay of the method proposed in [LH05] is 300 heart beats, which means 225 seconds for a typical HBR of 80 beats per minute.

P wave detection

We perform the P wave detection by comparing the candidate P wave automatically obtained by the previously mentioned wavelet-based delineation algorithm with a model of the P wave, inspired on the approach proposed by [Cea08]. However, in our case this model, depicted in Figure 4.17, was built by averaging all the annotated P waves found in the QTDB [LMGM97]. Then, for each beat, we calculate the correlation coefficient between the P wave detected by the delineation algorithm and our P wave model. This correlation value is then given as an input to our fuzzy classifier.

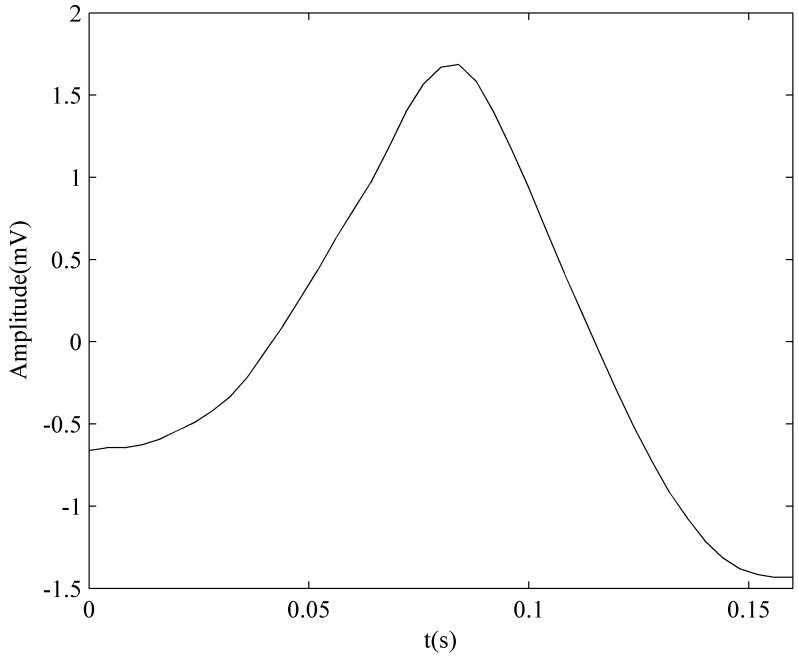

Figure 4.17: P wave model

4.3.3. Fuzzy Classifier

To combine the outcome of both sub-algorithms and to deal with limited resources, we have developed a specific fuzzy classifier for AF detection. Thanks to its low complexity, it can be implemented with very few resources and, therefore, it fits in the limited resources of state-of-the-art WBSN nodes, such as the Shimmer™.

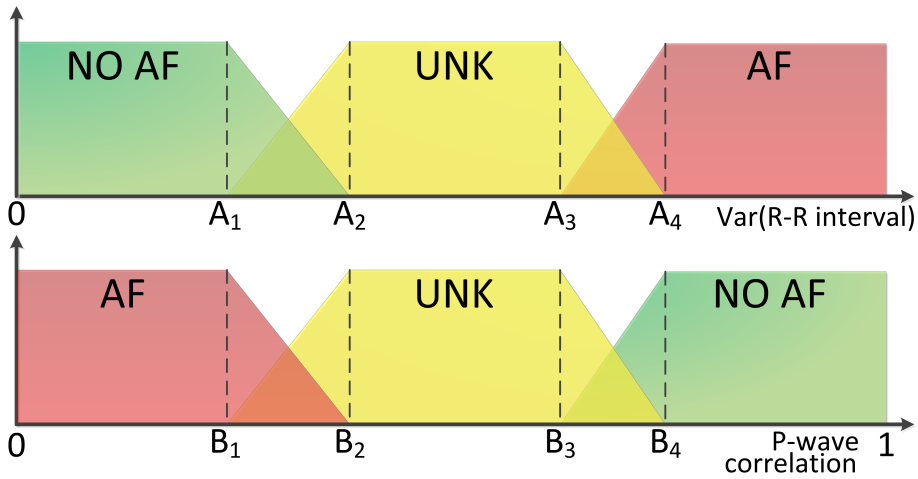

Figure 4.18: Membership functions of our fuzzy classifier

The classifier takes, as input, the output of the sub-algorithms and provides, as output, the outcome of classification process (NSR or AF). Trapezoidal membership functions are used to calculate fuzzy values. They translate the raw output of the two sub-algorithms into a fuzzy value that is subsequently used by fuzzy rules to detect and classify AF episodes. Each membership function is characterized by four parameters, which are represented by **A** (for HBR analysis) and **B** (for P wave correlation), that define the borders of fuzzy values in the process of fuzzification and can be changed to tune the algorithm for specific needs. In this specific problem, each fuzzy variable can be represented with three fuzzy values. They indicate the output of the corresponding sub-algorithm. These values are: **AF**, **NO AF** and **UNK**. *AF* indicates that an AF episode is detected, *NO AF* indicates a normal sinus rhythm, and when the output is equal to *UNK*, it is not clear if an AF episode is happening or not; in the latter case, additional information is required (i.e. the output of the other membership function). To reliably detect AF episodes, membership function parameters should be set correctly. After a careful experimental exploration, we have set these parameters to the values indicated in Table 4.11, which provide the best results. The developed membership functions are depicted in Figure 4.18.

To detect AF episodes, our proposed classifier combines fuzzy values with a set of rules. Rules are weighted according to the accuracy and the

Table 4.11: Configuration of the parameters of our fuzzy classifier

A_1	150	B_2	0.5
A_2	200	B_2	0.6
A_3	250	B_3	0.7
A_4	300	B_4	0.8

Table 4.12: Fuzzy rules used for AF classification

Var(R-R interval)	P wave correlation	output value	output weight
AF	AF	AF	1
AF	-	AF	0.5
-	AF	AF	0.2
AF	UNK	AF	0.8
UNK	AF	AF	0.4
NO AF	UNK	NO AF	1
UNK	NO AF	NO AF	0.5
NO AF	-	NO AF	0.9
-	NO AF	NO AF	0.5
NO AF	NO AF	NO AF	1

correlation between symptoms and correct AF classification. The set of rules I used in this work is summarized in Table 4.12.

The configuration of the parameters of the fuzzy classifier shown in Table 4.11 was set to obtain the best results on all the ECG records of the database used to validate the proposed approach, each of which belongs to a different patient. However, since the ECG signals differ from patient to patient and depend on physiological aspects like heart size or chest structure, the parameters of the classifier can be tuned further by training to improve classification accuracy on a specific patient. In fact, our results indicate that a personalized configuration for each patient always enhances the accuracy of the AF detection, achieving up to 100 % sensitivity and specificity for several signals of the used database. This training can be done using small excerpts of available ECG traces of each patient that have been previously annotated by the cardiologist. Moreover, training not only improves classification itself, but can be used to improve single-algorithm accuracy.

4.3.3.1. Real-Time Embedded AF Detection

The actual implementation of the proposed algorithm has been optimized for the Shimmer™ WBSN node. The algorithm is optimized to reduce the memory usage and execution time, in order to achieve a long lifetime of the platform while performing AF detection in real-time.

One of the most important parts in the design of the proposed AF detection approach is the module that performs the detection of the R peaks, which is needed by the heart rate analysis algorithm to compute the R-R interval variance. In particular, a precise computation of the R-R intervals relies on an accurate detection of the R peaks. As previously mentioned, the algorithm that performs this duty is the wavelet-based delineator presented in Section 4.2.2.2. We mainly selected this algorithm for its low memory usage (it only needs 7.2 kB of RAM) and its low computation requirements (in only requires a 6.78 % duty cycle). This algorithm also performs the delineation of the P and T waves, and therefore we can reuse the delineation of the P wave as an input for our own module of the AF detection algorithm that performs the analysis of the absence of the P wave. We have optimized this algorithm for the target Shimmer™ WBSN platform, and the implementation details are described next.

To enable the real-time implementation of the proposed AF detection approach, I needed to perform several code optimizations. Since the MSP430 microcontroller of the Shimmer™ platform is not equipped with a floating-point unit, the first adaptation needed is the conversion of all the floating-point operations to use 16-bit integers. This caused a dramatic fall in the performance of the algorithm, due to the significant loss of accuracy in the calculation of the variance and the correlation coefficient needed by the heart rate analysis and the P wave detection modules, respectively. Therefore, the operands involved if these operations were converted to 32-bit integers. However, this conversion did not solve the accuracy loss, and I had to increase the width of 28 % of the operands to 64 bits in order to overcome the problem. In addition, I optimized the computation of the correlation coefficient, which needs a square root operation to calculate the deviation of the P wave under analysis. Hence, the C library function `sqrt`, that involves several floating-point operations, was replaced by our own implementation of the Longhand square root algorithm [Rol87], that only uses integer operands and performs well on platforms without hardware division support, such as typical WBSN nodes.

After these optimizations, the total amount of memory used by our complete AF detection approach algorithm is 8.4 kB of RAM, which means that all the required processing for ECG classification and AF detection after the delineation algorithm only needs 1.2 kB of RAM. This low memory usage is mainly due to the use of our new fuzzy classifier, that requires very few resources while achieving a highly accurate classification. Overall, the duty cycle of the complete real-time AF detection approach is only 7.73 %, enabling the MSP430 to go to sleep mode during the remaining 92.27 % of the time and therefore resulting in an extended lifetime of the Shimmer™.

4.3.3.2. Validation and experimental results

To assess the AF detection accuracy of the proposed algorithm, I have used the MIT-BIH AF database (AFDB) [MM83], which is composed of 23 10-hour ECG recordings sampled at 250 Hz, with manual annotations of the AF episodes that appear in the signals. To this end, a beat-to-beat comparison between the results obtained by the algorithm and the manual annotations of the database is performed. Then, we calculate the sensitivity (Se) and specificity (Sp), which are defined as:

$$Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP} \tag{4.54}$$

where TP is the number of true positive detections, FN is the number of false negative detections, TN is the number of true negative detections and FP is the number of false positive detections.

Table 4.13: Performance comparison between the proposed AF detection approach and state-of-the-art offline algorithms

Algorithm	*Se* (%)	*Sp* (%)
Proposed AF detection algorithm	96	93
Couceiro et al. [Cea08]	93.8	96.09
Moody et al. [MM83]	93.58	85.92[1]
Cerutti et al. [Cea97]	93.3	94[1]
Tateno et al. [TG00]	93.2	96.7
Shkurovich et al. [Sea98]	78	92.65
Logan et al. [LH05]	96	89

The results obtained after running the algorithm over the AFDB show that the proposed system achieves 96 % sensitivity and 93 % specificity. These results are comparable to state-of-the-art offline AF detection algorithms, even outperforming several of them as Table 4.13 shows, while the proposed AF algorithm is able to operate online in WBSN nodes and provides real-time AF monitoring and detection.

4.4. Energy versus performance trade-off study

Based on the power characterization of the Shimmer™ node, which was introduced in Section 3.4.1, this section proposes a comparison of the energy consumption of a Shimmer™ node running a simple ECG streaming application and the algorithms considered in this chapter, in order to evaluate

[1]These values correspond to positive predictivity ($P+ = \frac{TP}{TP+FP}$)

how onboard processing affects the total energy consumption of the sensor node

4.4.1. ECG delineation

Table 4.14 reports the energy consumption of a Shimmer™ node an ECG streaming application and the various ECG delineation algorithms proposed in Section 4.2. The first row reports the duty cycle in percentage of ECG signal acquisition time, while the second row indicates the time interval between the generation of two consecutive data packets of 114 bytes. Given that a sample is represented using 1.5 bytes and that the considered sampling frequency is 250 Hz, the streaming application for instance generates one packet every 304 ms. For the delineation algorithms, on the other hand, 30 bytes of data (i.e., amplitude and timing information of the 9 fiducial points of the ECG beat) are generated every time a heart beat is detected. Thus, a packet can store up to 3 heart beats. Assuming a typical heart beat rate of 80 beats per minute, a full packet is queued for transmission every 2250 ms. The third row shows the beacon interval, which is adjusted according to the time interval between consecutive packets. The fourth row reports the total energy consumption per second for every algorithm. Finally, the last row is directly calculated from the fourth one, assuming the energy supply of the Shimmer™ platform is a 280 mAh Li-ion battery at 3.7 V.

Table 4.14: Node lifetime for the delineation algorithms

	ECG Stream.	Single lead WT	Single lead MMD	2-lead Morph. filt.	2-lead spline filt.
Duty cycle (%)	0	6.78	18.25	19.17	33.22
Packet ready every... (ms)	304	2250	2250	2250	2250
Beacon interval (ms)	983	7864	7864	7864	7864
Energy consumption (mJ)	7.70	7.44	8.62	8.71	10.15
Lifetime (h)	134.6	139.2	120.2	118.9	102.1

Table 4.14 shows that the total energy consumption increases with the computational burden of the delineation algorithm. Surprisingly, it also evidences that even the computationally-light single-lead WT-based delineation algorithm, which only consumes 6.78 % of the CPU time *and* significantly reduces the amount of data to transmit (from 3.29 to 0.44 packets per second), offers a marginal lifetime extension with respect to the plain ECG streaming. These results highlight that, for WBSN applications, the radio is not always responsible for most of the energy consumption of the node, as widely assumed. To understand the limited lifetime extension provided

by embedded signal pre-processing, we subsequently characterize and analyze the energy consumption breakdown of the considered platform. The two main sources of energy consumption are the radio and the microcontroller.

The radio

The energy consumption of the radio depends on the MAC protocol [Std03], which adds communication overhead to control the access to the shared medium. The fraction of energy consumption due to the used MAC protocol only represents 1.3 % for the cubic spline delineation, and 14.3 % for streaming. Therefore, the considered GTS-based protocol is a good choice for the continuous ECG monitoring with periodic reporting scenario. It is worthwhile mentioning that this energy characterization was done for line-of-sight communications between the sensing node and the WBSN coordinator. A detailed study of the impact of various body-area propagation conditions should be performed. It is however outside of the scope of this thesis.

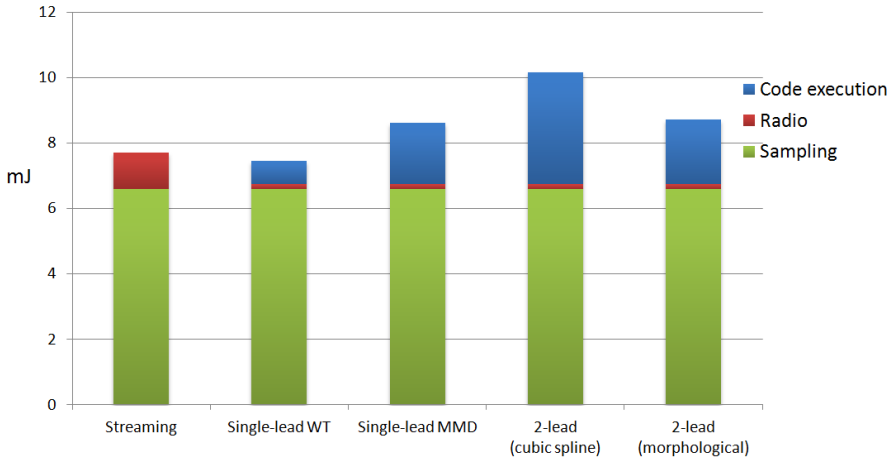

Figure 4.19: Breakdown of energy consumption of Shimmer™

The microcontroller

The consumption due to the CPU has a significant impact on the node lifetime for delineation algorithms that require more computational power than wireless ECG streaming. Except for single-lead WT-based delineation, the increased CPU activity leads to a shorter node lifetime, although the radio energy consumption is decreased by a factor of 8.6× (from streaming to delineation). Thus, the code execution can be very expensive if the power consumption of the microcontroller is high or the algorithm to be executed has not been carefully optimized for the target platform. Therefore, a careful study must be done when choosing or designing the microcontroller of each final WBSN platform. Indeed, it is key to take into account the po-

wer consumption in active mode as well as the possibility of having a low voltage operation or using low-power modes that disable some parts of the microcontroller when they are not used.

Nevertheless, for the Shimmer™ platform, the dominant process of the microcontroller in terms of energy consumption is not the code execution. As illustrated in Figure 4.19, most of the energy consumption is due to sampling the ECG signal (i.e., 65.2 % for multi-lead delineation with cubic spline baseline estimation and 85.7 % for the streaming scenario). Therefore, the sampling process can be very expensive in terms of energy consumption depending on the hardware implementation and the sampling frequency. In this particular case, for the considered sampling frequency of 250 Hz, it was not possible to turn the internal reference voltage of the A/D converters off and on between two consecutive samples, because the time that the internal voltage regulator takes to reach the appropriate level for the conversion is longer than the sample duration. Therefore, the voltage reference is always active, increasing significantly the total energy consumption. Hence, a careful choice of an ultra-low-power A/D converter can potentially decrease to a large extent the energy consumption of the whole system, as the sampling process has proven to be the greatest energy-draining task in the design of our wearable ECG WBSN system.

4.4.2. Arrhythmia diagnosis

Table 4.15: Node lifetime for streaming and the proposed arrhythmia diagnosis applications

	ECG streaming	General diagnosis	AF detection
Duty cycle (%)	0	6.85	7.73
Packet ready every... (ms)	304	21000	21000
Beacon interval (ms)	983	62915	62915
Energy consumption (mJ)	7.70	7.32	7.41
Lifetime (h)	134.6	141.5	139.9

This section provides an energy consumption analysis of the arrhythmia diagnosis algorithms proposed in Section 4.3 as well as a comparison with ECG data streaming. Table 4.15 shows the energy consumption of the Shimmer™ running an ECG streaming application, as well as the proposed algorithms for general arrhythmia diagnosis and AF detection. Both algorithms send only the obtained results (i.e., the decision of the classifier for each ECG beat, indicating whether or not an arrhythmia was detected). The first row shows the code execution in percentage of ECG signal acquisition

time, while the second row shows the time interval between the generation of two consecutive packets. In the case of the arrhythmia diagnosis algorithms, only the results of the classification are sent to the WBSN coordinator, which means that, for each beat, the node sends the time where the beat was detected (28 bits) and a code that contains the result of the classification (4 bits). This code has 9 possible values, to indicate if the heart beat is normal, or if an AF or any other of the 7 pathologies described in Table 4.10 was detected. Therefore a packet can store 28 heart beats. Assuming a typical HBR of 80 beats per minute, a packet is generated every 21 s. The third row shows the beacon interval, that is adjusted taking into account the maximum throughput required by the algorithm. The fourth row shows the energy consumption of the considered sensor platform per second. Finally, the last row indicates the total lifetime of the node by considering the energy supply of the Shimmer™, which has a 280 mAh Li-ion battery at 3.7 V.

From Table 4.15 we can conclude that, although advanced processing is performed in the Shimmer™ WBSN node and therefore the energy consumption of the microcontroller increases, the total lifetime of the WBSN node including the real-time AF detection algorithm is higher than using a baseline ECG raw transmission. The reason for this situation is the significant reduction in the amount of transmitted information to the WBSN coordinator. In fact, thanks to this reduction, the algorithm only generates a packet every 21 s, which decreases the energy consumption of the radio to only 0.22 % of the total energy consumed by the whole platform in the case of AF detection. However, if the packet generation frequency is reduced even more, which is possible if the node only sends an alarm packet when an arrhythmia is detected, the additional energy savings would be negligible. For example, if no alarm is detected, the radio's energy consumption would be 0.04 % of the total, and the lifetime would be only increased from 139.9 to 140.1 hours. Hence, this observation concludes that the only way to further reduce the energy consumption of the platform lies in the optimization of the ADC or the microcontroller.

Chapter 5

Conclusions

Pioneering spirit should continue,
not to conquer the planet or space...
but rather to improve the quality of life.

Bertrand Piccard

This chapter summarizes the conclusions that can be extracted from the work presented in this Ph. D. thesis. Moreover, several lines of promising research work will be proposed. Finally, the applications and achievements of this work are described and a list of the publications produced during this thesis is provided.

5.1. Final remarks

As previously mentioned in the introduction of this thesis, cardiovascular diseases are the major cause of mortality worldwide, responsible for 30 % of deaths worldwide in 2005 and economic fallout in billions of Dollars [PW10]. Their burden is only expected to rise due to the fast aging of the world population and the increasing prevalence of unhealthy lifestyles. WBSNs for health monitoring, diagnosis and emergency detection are gaining popularity and will deeply change healthcare delivery in the next years. This technology will help to decrease the high healthcare costs and medical management needs of ECG analysis and diagnosis that, otherwise, are expected to be unsustainable for traditional healthcare delivery systems in the next years.

This thesis proposes the development of a long-lived wireless ECG monitoring system that not only continuously acquires and displays the ECG signal of the patient, but it also features on-board real-time analysis and diagnosis, instantly providing information about the operation of the heart. To this end, first, a study of the available WBSN architectures was conducted in chapter 2 and the Shimmer™ platform was chosen to be used, since

it is one that best fits the requirements of this work. This platform features a great flexibility to implement the different techniques that have been proposed, and it offers enough computational and memory resources to deal with the herein considered applications. Furthermore, it is a good representative of a WBSN platform, therefore all the results provided in this thesis are similarly applicable to other WBSN nodes.

Moreover, this work seeks to improve WBSN energy efficiency, while meeting the data fidelity requirements of the application. To achieve this goal, chapter 3 presents a MAC protocol especially tailored to fulfill the requirements of WBSNs. This MAC protocol is a reduced version of the beacon-enabled mode of the IEEE 802.15.4 protocol using guaranteed time slots (GTS). It uses a TDMA-like approach, which has been proven to be advantageous in star-topology WBSN, since its schedule-based nature eliminates the possibility of having collisions and overhearing, while keeping the control packet overhead very low and dramatically reducing idle-listening. The use of this MAC protocol contributes to improve WBSN energy efficiency, since it significantly reduces the energy consumption of the wireless communication.

Chapter 3 also proposes the use of FreeRTOS, a very light real-time OS that enables an efficient and energy-aware management of the hardware resources of the WBSN node. In addition, FreeRTOS provides an effective scheduling of the different tasks to be executed by the node, such as signal processing algorithms, MAC protocol management or sensing, as well as interprocess communication. Finally, the use of this OS allows to decouple the hardware and software resources of the node, providing more flexibility.

Then, chapter 4 presents several algorithms for real-time ECG filtering, delineation and arrhythmia diagnosis that provide the nodes with the ability to evaluate the heart's state and detect potentially dangerous cardiac disturbances. These algorithms, which were optimized and ported to the Shimmer™ platform, can also contribute to the reduction of the WBSN node's energy consumption, since they significantly reduce the amount of information that needs to be transmitted by the node. To achieve this energy consumption reduction, a trade-off between CPU usage and radio communication needs to be explored.

In particular, chapter 4 proposes two real-time single-lead ECG delineation algorithms, based on the wavelet transform and on the multiscale morphological transform, respectively. The excellent accuracy of these online ECG delineators was established using a commonly used standard database. Then, this chapter proposes two RMS-based multi-lead ECG delineation algorithms using two different techniques for baseline wander removal: cubic spline baseline estimation and morphological filtering. The multi-lead delineation performs equivalently to the best single-lead delineation for the 2-lead QTDB, within a fraction of a sample duration of the CSE tolerances. This confirmed that this multi-lead approach alleviates the dependency of

delineation results on specific leads or lead systems, and on expert intervention. Moreover, the two evaluated baseline wander removal approaches were shown to provide different trade-offs between positive predictivity, on the one hand, and computational burden and memory footprint, on the other hand.

Then, chapter 4 proposes two approaches for real-time arrhythmia diagnosis. The first one uses a set of rules that take into account the timing and morphological information extracted after the delineation of the ECG signal to evaluate, for each beat, whether it is normal or presents any irregularity that may indicate an abnormal behavior of the heart. The second approach, which is more accurate since it is especially tailored to detect a particular arrhythmia, consists of an AF detection algorithm that relies on the observation of two characteristic irregularities of AF episodes in the electrocardiogram (ECG) signal, namely the R-R interval irregularity and the absence of the P wave. This further shows the importance of a sensible design and optimization of the AF detection approach for the limited processing and storage capabilities of the target WBSN platform. As a result, the performance results of the complete real-time AF detection system are similar to state-of-the-art off-line AF detectors, achieving 96 % sensitivity and 93 % specificity for the AFDB.

Chapter 4 also presents a comprehensive evaluation of the Shimmer™'s energy consumption for the various delineation and arrhythmia detection implementations investigated, with respect to a simple raw ECG streaming application. The resulting energy consumption breakdown showed:

1. Although advanced signal processing is performed in the sensor node and thus the energy consumption of the microcontroller is higher, the lifetime of the overall system can be increased with respect to streaming, since the amount of information that needs to be sent to the WBSN coordinator is dramatically reduced.

2. Provided an optimized schedule-based MAC protocol and corresponding optimized radio shutdown, a single-lead embedded delineation only offers a marginal node lifetime extension.

3. This marginal advantage vanishes for multi-lead approaches, which entail up to 5 times increase of CPU usage.

4. The ECG sampling is the greatest energy-draining task for the considered sampling frequency.

Accordingly, our system-level exploration suggests that a careful choice of the WBSN platform's components, taking into account the final application during design time, can potentially decrease to a large extent the energy consumption of the whole system, therefore node lifetime extension can be achieved provided:

1. An ultra-low-power A/D converter and associated architecture is used to rationalize the ECG sampling cost.

2. A truly ultra-low-power microcontroller is chosen to carry out signal pre-processing.

3. Last, but not least, a sensible algorithmic and software optimization of the embedded signal processing is performed to take into account the limited processing and storage resources of wireless sensor nodes.

5.2. Applications and achievements of this Ph. D. Thesis

All the algorithms presented in this thesis have been first implemented and tested on a PC. Then, they have been ported for their real-time execution on the Shimmer™ platform and validated using standard databases, in order to evaluate their performance and compare them to other similar state-of-the-art algorithms. Finally, these algorithms have been carefully optimized to reduce as much as possible their execution time and memory footprint, and used in real conditions. To this end, I invested a great amount of time in implementation and experimental validation, until everything perfectly and smoothly worked. Everyone that has ever dealt with hardware will understand me when I say that this task has been many times extremely unpleasant and frustrating. But at the end, the satisfaction of seeing the system working and used by people, and obtaining useful results, compensates for all those bad moments. There is an extra motivation in this thesis that makes it exciting and rewarding, beyond reducing healthcare cost and energy consumption, which is making people's lives better, since wearable round-the-clock monitoring devices will bring about new types of analysis, leading to new treatments and ultimately save lives all around the world.

In order to make the proposed ECG monitoring system easily usable by patients and doctors, graphical user interfaces for PC, iPhone, iPod Touch, iPad, Nintendo DS and Android tablets have been developed, with the collaboration of Nicolas Boichat, Rubén Braojos, Karim Kanoun, Rafael de la Hoz, Pablo Fernández and Miguel Márquez. Figure 5.1 shows a Shimmer™ node with four electrodes, which is wirelessly connected to an iPhone. The developed application displays on the iPhone the ECG signal of the patient, the heart beat rate and the results of the delineation received from the Shimmer™ in real-time, and records all the data in a file. Moreover, if an arrhythmia detection algorithm is being used, the application shows the real-time diagnosis of the ECG signal and sends an email to the doctor when an abnormal event is detected. The email includes a diagnosis of the arrhythmia and a few seconds of the ECG signal around the detected event.

Figure 5.1: Electrode sensors on the skin link to the Shimmer™ platform clipped to a patient's belt sending continuous data to an iPhone, which displays the ECG signal, heart beat rate and delineation results

Many demonstration of this system have been shown in conferences, exhibition booths, university research days, etc. During one of these demonstration, the EPFL media services recorded a video that was given widespread coverage in blogs, newspapers, radio and TV, including its appearance in CNN [Lab].

Moreover, this system has been used under real settings during the first Solar Impulse virtual flight [Imp]. During this event, which took place in the installations of Solar Impulse in Dübendorf (Switzerland) in February, 2012, André Borschberg (one of the pilots and founders of the project) was at the controls for a flight simulation of 72 hours in a full-size mock-up of the cockpit of the HB-SIB solar airplane. For the whole duration of the flight the pilot was using the ECG delineation system and his ECG signal was recorded, as well as the delineation results. With the help of Rubén Braojos and Professor Jean-Marc Vesin, from EPFL, the recorded results were used to evaluate the fatigue and vigilance of the pilot. To this end, the heart rate variability was studied, and the relation between the sympathetic and parasympathetic nervous systems was extracted, to be able to detect when the pilot is feeling fatigue and is not fully concentrated in controlling the plane. This information is extremely important for the sleep management, enabling to develop the best flight and rest strategy for the pilot during the real flight. The use of the system was monitored by a team of doctors from the Hirslanden Private Hospital Group, and the results for fatigue detection were approved by them. Figure 5.2 shows Rubén Braojos and I placing the

Figure 5.2: The ECG delineation system is installed on the pilot some minutes before the "virtual" takeoff.

ECG monitoring system on André Borschberg, just a few minutes before the beginning of the virtual flight. The system recorded and analyzed the ECG signal of the pilot during the whole mission.

The use of the proposed ECG monitoring and diagnosis system is not limited to humans, but it can also be applied to animals. In this regard, several experiments were conducted in a farm in La Forclaz (Switzerland), where the ECG signal of sheep was monitored. It has been proved that sheep can feel a dangerous situation and detect the presence of wolves. The aim of these experiments is to extract the stress level of sheep from their ECG to anticipate the potential attack of a wolf. Figure 5.3 shows the system installed on a sheep, as well as a screenshot of the iPhone that receives and displays the ECG signal and delineation results.

Finally, this work gave rise to the US patent entitled "Automatic online delineation of a multi-lead electrocardiogram bio signal", which covers all the herein presented ECG signal processing algorithms.

5.3. Ongoing work derived from this thesis

During the development of this thesis, two lines of work have been initiated, motivated by some of the conclusions that have been extracted. The first one involves the creation of a general analytical model targeted to no-

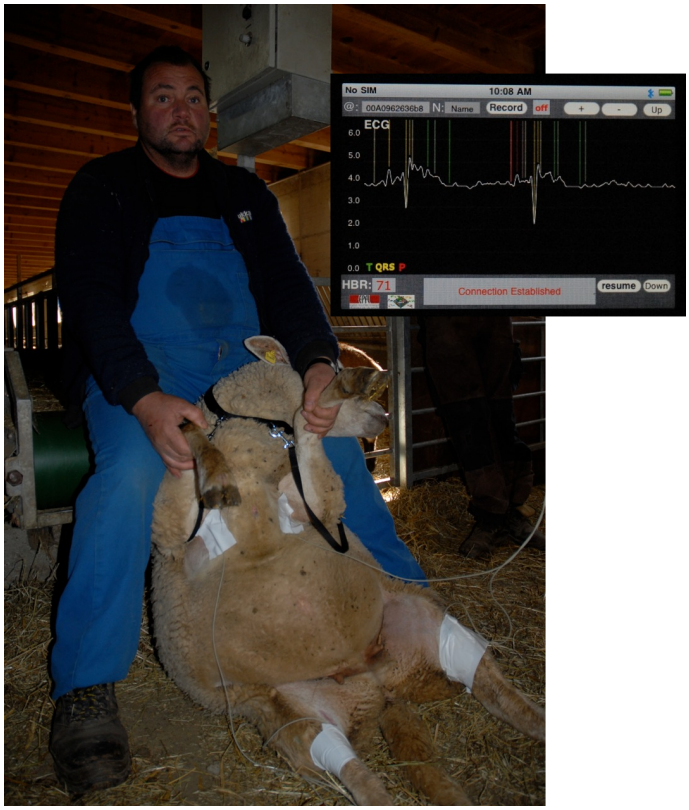

Figure 5.3: ECG delineation system installed on a sheep. The received ECG signal and delineation results are displayed on an iPhone (cf. right image)

des whose architecture is similar to the Shimmer™'s. The model is part of an optimization framework that performs an evaluation of the node configurations and the corresponding performance tradeoffs to find the optimal solution in the design space, contributing to a further reduction of the WBSN node's energy consumption. The second line is related to the design of energy-efficient microcontrollers for online biomedical processing that exploit low-voltage computing.

5.3.1. Multi-objective model-based design for WBSN platforms

As discussed in the previous chapters of this thesis, the design of WBSN nodes is mainly focused on maximizing the lifetime of the node by reducing the energy consumption, while keeping into account performance requirements such the delay and quality of the delivered data. Since a design may depend on tens of parameters, an efficient multi-objective optimization fra-

mework is required to explore the design space and to identify the Pareto-optimal solutions. The most critical part of this process is to provide a fast yet reliable estimation of all the optimization metrics. There are three possible techniques to evaluate a solution: a) an exhaustive set of experiments, which however cannot be automated; b) a network simulation, which is slow and hence impractical when a large number of potential solutions needs to be explored; c) an analytical model of the node, which favors a quick optimization and a better analysis of the node behavior.

Model-based evaluation has a long history, as many models have been proposed to describe the basic components of a node (e.g., memory, radio, etc. [KOKE06][GC97][SRS03]). However, combining those components to form a model of the entire node is not an easy task, as the model should include meaningful information of the specific node, while being reusable and not requiring a massive amount of experimental data to be constructed. In order to cope with the difficulty of building reliable node characterizations, a promising trend is to generate statistical models from a properly-selected set of experimental data [BDCD11]. The experimental data is used to estimate the parameters of a set of simple equations, which however do not provide an application-aware evaluation of the node.

This research line tackles the model-based optimization problem from a different perspective, i.e., by narrowing the scope of the model to WBSNs and focusing on their most typical features (defined in [UHB+10]). In particular, this allows us to discard those aspects that are not generally required, such as complex networking or task assignment. In this way, we are able to analytically capture aspects like lifetime, transmission quality and application performance with a high precision, while still keeping the model general and reusable. Then, if the size of the design space is not too big, we can simply use an exhaustive search to perform the design space exploration. Otherwise, we need to use a more sophisticated approach, like the adapted version of the multi-objective simulated annealing algorithm proposed in [BRK+12b], to perform the design space exploration and find the Pareto-optimal configurations. The proposed model has been tested on a real application for ECG monitoring that uses compressed sensing [MKAV11], and is implemented on the Shimmer^{TM}platform.

5.3.1.1. Related work

Model-based optimization of wireless sensor nodes is a topic that has been already explored in the literature. However, most of the related works characterize the energy consumption of one of the node components shown in Figure 5.4, whereas just few of them aim at optimizing multiple components and performance metrics, which is instead the purpose of our work.

Power models of single components such as the microcontroller [GC97] and the memory [KOKE06] are available outside the scope of WSNs. Ho-

wever, most of the effort in the context of sensor networks with low-power requirements has been devoted to the wireless communication part, with the design of new MAC protocols and the characterization of the radio. In [SRDV08], the authors propose an alternative MAC protocol for the *ZigBee* standard that introduces new power-saving policies. In [LZA09], a model that relates the routing performed at the MAC level to the node lifetime is proposed. However, both works assume a multi-hop routing, and thus they cannot be applied to the star topologies used in WBSNs [UHB+10]. A characterization of the radio has been proposed in [SRS03], where the energy consumption is related to parameters like the bit error rate and the modulation. Similarly, [CCGCG10] provides a model for an IEEE 802.15.4 transmitter, which is supported by a set of physical measurements. However, the models of the single components cannot capture the interdependencies that exist between the different parts of the node, and in particular they often discard the effects of the application.

A few existing works have tried to propose an optimization that considers several components of the node. In [BCSV06] the authors propose a platform-based design methodology for industrial control. Although the work considers all the aspects of the node design, it is not based on an analytical model, and it mainly focuses on the generation of a complex network, which is not a critical aspect in WBSNs. In [KMW08], different energy/delay tradeoffs are explored by exploiting voltage and modulation scaling. Similarly, [YFG07] proposes a model-based optimization framework for star-topology networks, and a genetic algorithm to reduce the energy consumption by acting on the voltage and the modulation level. However, the number of parameters involved in the optimization is a small subset of the ones that can be tuned on real nodes. In [NWP10], the authors propose an optimized transmission schedule to minimize the packet delay. The work shows good potential if transmission delay is the main objective, but it is not proved to scale in more sophisticated multi-objective optimization scenarios. A very relevant example of application-driven design is proposed in [HEC+08], where a multi-objective optimization involving all the system components is provided in the field of wildlife monitoring. The work shows how a deep knowledge of the final application can lead to an optimized node lifetime, while guaranteeing the quality of service (high data rate and low distortion). However, it heavily relies on the experimental data and hence is very specific for the target domain.

5.3.1.2. The proposed analytical model

The proposed model is targeted to WBSN nodes whose architecture is similar to the one depicted in Figure 5.4, which was introduced in Section 2.1. This model, which estimates the relevant quality metrics associated to the system, can be part of an optimization framework that also includes an

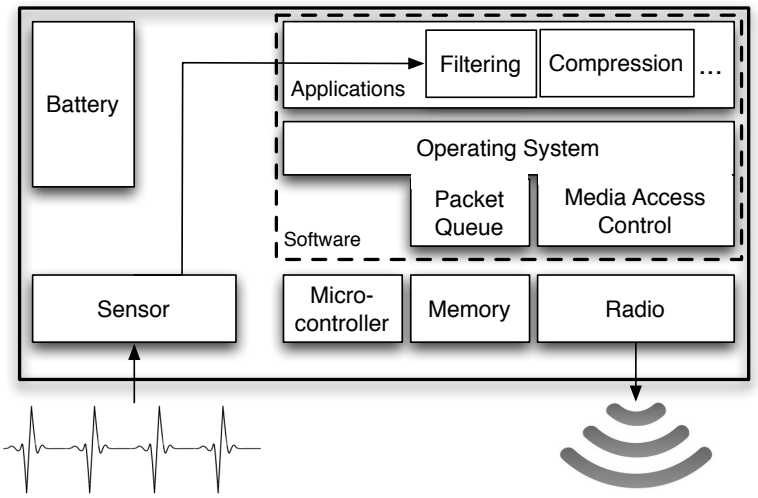

Figure 5.4: Block diagram of the reference WBSN node architecture

algorithm to find the optimal configurations. Due to the limited size of the design space in this work the algorithm can be reduced to a simple exhaustive search that will always find the best solution.

The node model includes fundamental parts that describe the common structures of every WBSN node, and advanced parts that can be further detailed according to the specific scenario (e.g., the application, the communication channel), and parameters that need to be determined through experimental data. The contribution to the energy consumption of each component depicted in Figure 5.4 is characterized, although the level of detail differs from component to component depending on the number of relevant design parameters.

The proposed model is comprehensive as it captures the interdependencies between different components. The main metric, i.e., the overall energy consumption, is expressed as a function of all the hardware components of the system (sensor, microcontroller, memory and radio), whose behavior is influenced by the applications and the OS. In particular, the energy consumption per second is expressed as:

$$E_{node} = E_{sensor} + E_{\mu C} + E_{memory} + E_{radio} \tag{5.1}$$

However, since the straightforward reduction of the energy consumption may lead to the loss of performance in one or more components of the node, we defined a set of performance metrics that only involve one or two components, in order to keep them monitored during the design.

Sensor

The sensor component consists of a transducer to detect the signal, and a

hardware circuit to sample the data at a frequency $f_{sampling}$ and quantize it with a resolution of ρ_{ADC} bytes. The energy consumption of the former can be considered as a constant related to the specific sensor, while the latter is linearly related to the sampling frequency [BB95]:

$$E_{sensor} = E_{transducer} + [\alpha_1 \cdot f_{sampling} + \alpha_0] \ , \tag{5.2}$$

where α_1 is a constant depending on the capacitance and the square of the supply voltage [BB95] of the A/D converter, while α_0 describes leakage effects and is determined experimentally.

Although specific metrics can be defined to estimate the performance of the sensor, e.g. the signal-to-quantization-noise ratio, a more meaningful evaluation can be obtained by combining it with the subsequent application component.

Applications

The applications are software programs that do not directly dissipate energy, but influence the performance of the microcontroller, the memory and the radio. In particular, the applications define the duty cycle (ψ_{app}) of the microcontroller, the memory requirement (σ_{app}) and the average number of memory accesses per second (γ_{app}), which can be determined using software profiling. In this case, we assume that data can be compressed at the application level with ratio CR, which generates the following amount of packets to be transmitted per second (R_p):

$$R_p = f_{sampling} \cdot \frac{CR \cdot \rho_{ADC}}{H_{payload}} \ , \tag{5.3}$$

where $H_{payload}$ is a parameter that defines the number of data bytes included in each communication packet, and hence $H_{payload}/\rho_{ADC}$ denotes the number of samples per packet.

In order to evaluate the performance of the applications, domain-specific metrics can be defined. If compression is the only processing performed at the application level, such a metric is defined as the quality of the reconstructed signal.

Operating system

The OS is composed of software routines that implement services such as the packet queue and the MAC layer. The software routines of the OS can be modeled as any other software application, thus requiring a duty cycle ψ_{OS} from the microcontroller, and a maximum memory σ_{OS} that is accessed γ_{OS} times per second on average. For the sake of analysis, we separate the memory required by the transmission queue from the remaining memory occupied by the OS, because a detailed model of the transmission queue is crucial to characterize the throughput of the system.

The MAC layer implemented in the OS manages the access to the wireless channel shared among a known number of nodes (N_{nodes}) connected to the WBSN coordinator. The access policy defined by the MAC algorithm can be modeled using two quantities: a transmission window of length Δ_{tx} when the node can transmit without conflicts, and the number of times this window is repeated per second, N_{tx}. Those quantities can be directly computed for contention-free access mechanisms, but they can also be determined statistically for contention-based policies. In order to enforce the access policy, the algorithm may require a number of control messages to be exchanged from the node to the coordinator. We denote this number as $\Phi_{Node \to C}$, and the length of those messages as $H_{Node \to C}$), and the opposite as $\Phi_{C \to Node}$, of length $H_{C \to Node}$. Finally, the MAC protocol defines the control information that must be included in each packet (typically a header and a checksum), thus determining the final length H_{packet} of each packet. The transmission queue, which can contain up to λ packets, has then a size of λH_{packet} bytes.

A set of performance metrics can be defined for the OS component, and in this work we focus on the throughput and the packet delivery delay. In particular, the system should guarantee a throughput of R_p packets per second as required by the application, but certain packets might need to be retransmitted with a probability equal to PER. We define $R_p^{(r)}$ as the packet rate including the retransmissions, which is equal to $R_p \cdot [1 + \varpi(PER)]$, where $\varpi(PER)$ is the estimated number of retransmissions, including possible multiple retransmissions of the same packet. Then, a sufficient condition to guarantee that the desired throughput is met is the following:

$$N_{tx} \left\lfloor \frac{\Delta_{tx}}{T_{packet}} \right\rfloor \geq R_p^{(r)} \ \wedge \ \lambda \cdot N_{tx} \geq R_p \ , \tag{5.4}$$

where T_{packet} is the packet transmission time, and $N_{tx} \cdot \lfloor \Delta_{tx} / T_{packet} \rfloor$ is the maximum channel capacity guaranteed by the MAC layer. The second condition prevents the capacity of the queue from being a bottleneck and to avoid dropped packets. If the conditions are satisfied, we can provide the following worst-case estimation of the packet delivery delay:

$$Delay = \frac{1}{N_{tx}} \cdot \frac{\sum_{i=N_{tx}-\nu}^{N_{tx}-1} i}{\nu} \ , \ \nu = \left\lceil \frac{R_p}{R_p^{(r)}} \cdot N_{tx} \right\rceil \ , \tag{5.5}$$

where ν indicates the fraction of the N_{tx} transmission windows that are required to send R_p packets. $Delay$ is then computed by considering that all the retransmissions occur before the R_p packets are transmitted.

Microcontroller

Similarly to the sensor, the consumption of the microcontroller is expressed as a function of its frequency $f_{\mu C}$. The processor needs to be active for a duty

cycle defined by the application and the OS, before switching to a low-power mode where only leakage effects occur:

$$E_{\mu C} = (\psi_{app} + \psi_{OS}) \cdot \beta_1 \cdot f_{\mu C} + \beta_0 \ , \tag{5.6}$$

where β_1 depends on the capacitance and on the square of the supply voltage, and β_0 describes the leakage effects and should be determined experimentally. Note that, if the specific scenario does not allow the microcontroller to switch to a low power state, $\psi_{app} + \psi_{OS}$ should be set equal to one.

Memory

The system memory is used for the execution of the applications and the OS, and to store the packets queue. The memory size M, which is also the main quality metric for the design of the memory component, can be written as:

$$M = \sigma_{app} + \sigma_{OS} + \lambda H_{packet} \ . \tag{5.7}$$

The energy consumption of a memory component is due to two factors [KOKE06]: a dynamic consumption due to the memory accesses, and a leakage that is known to be proportional to the memory size and appears when the memory is not being accessed. The software applications and the OS access the memory γ_{app} and γ_{OS} times per second, respectively. The transmission queue is filled with a number R_p of packets per second and, since we defined the throughput to guarantee that no packet is dropped, it is eventually read at the same rate. Assuming that the memory access in read and write modes has the same cost, we can express the energy consumption as:

$$\begin{aligned} E_{memory} = (2 \cdot R_p + \gamma_{app} + \gamma_{OS}) \cdot T_{mem} \cdot \zeta_{access} + \\ [1 - (2 \cdot R_p + \gamma_{app} + \gamma_{OS}) \cdot T_{mem}] \cdot M \cdot \zeta_{idle} \ , \end{aligned} \tag{5.8}$$

where T_{mem} is the access time in read or write mode, while ζ_{idle} and ζ_{access} are hardware parameters that define the consumption in idle and accessing modes.

Radio

The energy consumed by the radio depends on the number of packets that are sent and received. In particular, when a transmission of one bit takes place, the energy consumption can be expressed as:

$$E_{tx} = [P_{carrier} + P_r] \cdot T_{bit} \ , \tag{5.9}$$

where $P_{carrier}$ is the power required to generate the signal carrier, P_r is the remaining consumption related to the radio circuit. T_{bit} indicates the average time to transmit one bit, which also includes all the control information added by the physical layer [SRS03]. The value of $P_{carrier}$ can be determined

according to the desired PER. In particular, given the level of noise at the receiver, it is possible to compute the signal-to-noise ratio and consequently the bit error rate (BER), as a function of $P_{carrier}$ and the modulation scheme [SRS03]. Once the BER is known, the packet error rate can be expressed as the probability of one bit being erroneous in a packet of length H bytes, i.e., $1-(1-BER)^{8H}$.

The energy required to receive a bit (E_{rx}) is computed as in Equation 5.9, where $P_{carrier}$ is equal to zero, and P_r has a different value during the receiving phase. As a consequence, the energy consumption of the radio can be expressed as:

$$\begin{aligned} E_{radio} =& E_{tx} \cdot \left[R_p^{(r)} \cdot 8H_{packet} + \Phi_{Node \to C} \cdot 8H_{Node \to C} \right] \\ &+ E_{rx} \cdot \left[\Phi_{C \to Node} \cdot 8H_{C \to Node} \right] \ . \end{aligned} \tag{5.10}$$

5.3.1.3. A case study

This section shows how the previously proposed model can be easily adapted to cope with real WBSN applications and widely-adopted standards, and how the design space can be explored to find the optimal solution considering different performance tradeoffs.

Case Study Overview

We consider a real system that samples the ECG and uses the compressed sensing [MKAV11] to reduce the amount of data to be transmitted. The compressed signal is then sent to a smartphone that acts as a central coordinator, reconstructs the ECG and performs analysis and detection tasks. The choice of compressed sensing is motivated by the improved node lifetime, indeed, experimental results [MKAV11] showed that compressing and sending data can increase the node lifetime by 9.7 % when compared to transmission of the raw ECG.

The system is implemented on the Shimmer^{TM}platform [BGM^{+}10]. On the software side, FreeRTOS [biba] was ported on the node, and it controls the sensing, the queue services and the beacon-enabled mode of the IEEE 802.15.4 MAC layer [Std03]. In this MAC protocol, previously introduced in Section 3.2, a beacon is periodically sent by the coordinator to define the time structure in terms of superframes. A superframe is a time interval divided into an inactive and an active part, the latter being further divided into a contention-free and a contention-active portions. In this case study, only the contention-free part is used, so the transmission only occurs during *guaranteed time slots* (GTSs).

Mapping the case study on the analytical model

The model introduced in Section 5.3.1.2 provides a good characterization of

many parts of the target node, but additional information can be included to further describe the application, the memory, the MAC and the radio modulation.

At the application level, we estimated the duty cycle ψ_{app} required by the compressed sensing, and in the current implementation it only marginally depends on the value of CR. The performance metric considered for this component is the percentage root-mean-square difference (PRD), which quantifies the difference between the original ECG and the one reconstructed by the coordinator from the compressed data. By analyzing the experimental data provided in [MKAV11], the PRD can be expressed as a fifth-order polynomial function of CR:

$$PRD = \omega_5 CR^5 - \omega_4 CR^4 + \omega_3 CR^3 - \omega_2 CR^2 + \omega_1 CR - \omega_0 \ , \qquad (5.11)$$

where the coefficients ω_n are positive constant values.

The total available memory on the node is 10 kB. According to the experimental results, 6.5 kB are required by the compressed sensing application (σ_{app}), while 3.5 kB are reserved for the FreeRTOS routines (σ_{OS}) and for the transmission queue, whose size is then upperbounded. In particular, for a packet length H_{packet} of 127 B (i.e., the maximum value for the selected MAC), λ must be lower than 10.

The beacon-enabled IEEE 802.15.4 MAC layer can also be easily included in the node. Two protocol-specific parameters need to be defined: the superframe order (SFO), and the beacon order (BCO). The former determines the active period or superframe duration (SD), while the latter defines the interval between two beacons (BI) as follows:

$$SD = 15.36ms \cdot 2^{SFO} \ , \ \ BI = 15.36ms \cdot 2^{BCO} \qquad (5.12)$$

The superframe structure can be mapped on the transmission window Δ_{tx} of our model, as the average transmission time per second is equal to SD divided by the number of nodes in the network. Similarly, BI defines how many times a superframe is repeated, hence it can be related to N_{tx}:

$$\Delta_{tx} = \frac{SD}{N_{nodes}} \ , \ \ N_{tx} = \frac{1}{BI} \qquad (5.13)$$

In terms of control messages, the standard does not require any control message from node (thus $\Phi_{Node \to C} = 0$), whereas the coordinator sends a number of beacons that depends on BI, and an acknowledgment for each transmitted packet, hence:

$$\Phi_{C \to Node} = R_p^{(r)} + \frac{1}{BI} \ . \qquad (5.14)$$

Finally, the estimation of the PER for this case study can be obtained by computing the BER for the 4-PSK modulation [SRS03] with the selected value of $P_{carrier}$.

Node optimization

The design parameters available on the target platform are $f_{\mu C}$, CR, $H_{payload}$, λ, BCO, SFO, and P_{tx}, while the cost function includes E_{node}, PRD, $Delay$, and PER. For the sake of illustration, we adopted a coarse discretization of the parameters and reduced the design space to 10^8 solutions, which can be explored by an exhaustive search algorithm. In case the size of the design space is higher, we can use an algorithm like the MOSA approach proposed in [BRK+12b] in order to optimize the search of the Pareto-optimal node configurations.

The estimation provided by the proposed model proves to be effective as the error with respect to the experimental data is very low (i.e., it does not exceed the 1.9 %). Moreover, results show that the MOSA algorithm proposed in [BRK+12b] effectively explores the Pareto set, as the optimal solutions found by the MOSA perfectly match the ones found by the exhaustive algorithm, and cardinality of the Pareto set scales well with the number of executions. The solutions show a wide range of tradeoffs, e.g., the difference between the extreme values of E_{tx} exceeds 44 %, the values of the PRD span from 0 to 93 (the maximum range is up to 100), and it is possible to achieve real-time transmission as well as packet latencies of tens of seconds. In Figures 5.5 and 5.6, we can see a comparison between the solutions found by the MOSA algorithm and an exhaustive search to analyze the delay vs. energy consumption (Figure 5.5) and the PRD vs. energy consumption (Figure 5.6) [BRK+12a].

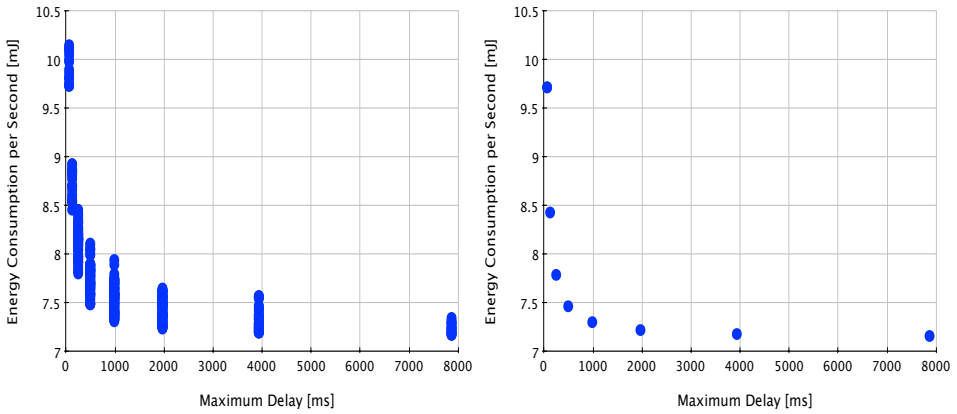

Figure 5.5: Delay and energy consumption of the solutions found by the approach proposed in [BRK+12b] (on the left), and an exhaustive search (on the right)

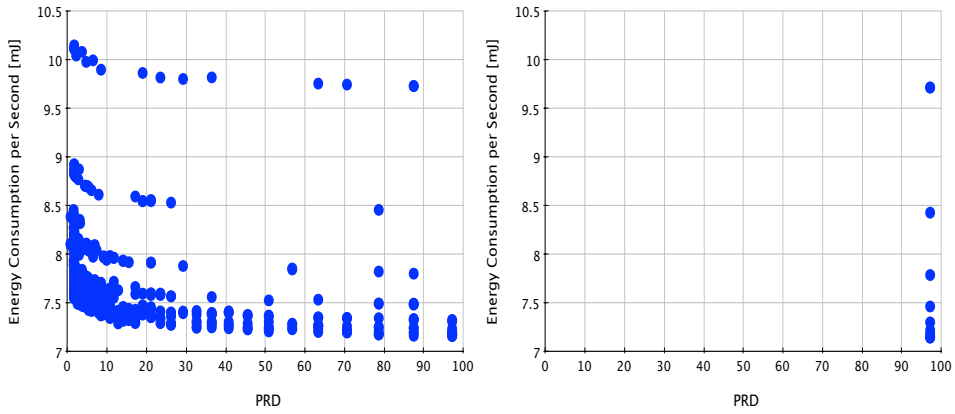

Figure 5.6: PRD and energy consumption of the solutions found by the approach proposed in [BRK+12b] (on the left), and an exhaustive search (on the right)

5.3.2. Low-power microcontroller design

In order to illustrate the work on low-power microcontroller design, we introduce two different processing cores, *Firat* and *TamaRISC* [DCA+12], to compare different architectural choices in terms of energy efficiency and performance for online biomedical processing requirements, while exploiting low-voltage computing. *Firat* is a custom implementation of a PIC24 from Microchip [Tecc], a well-established ISA, whereas *TamaRISC* is a custom ISA with a very reduced number of instructions. In the specific architecture of the processing cores, we compare the power vs. performance of different architectural choices, including sequential and parallel processing for different biomedical workloads while exploiting near-threshold computing. We limit the scaling of the operating voltage to the transistor threshold voltage level to avoid performance variability and functional failure issues occurring mainly at sub-threshold voltages.

The first processing core, *Firat* [DCA+12], is shown in Figure 5.7. *Firat* is a Reduced Instruction Set Computer (RISC)-like architecture with a Harvard memory model. The simple three-stage pipeline with partial data bypassing is able to provide the low-to-moderate performance required by biomedical applications. The core operates on a data word length of 16 bits, comprises 16 working registers, 2 external data and 1 instruction memory ports. The instruction word size is 24 bits and almost all the instructions use only one word. Each single-word instruction consists of an 8-bit opcode, which specifies the instruction, and one or more operands, which further specify the operation of the instruction. The instruction set is a subset of the PIC24's and includes totally 66 instructions. Among them, the ISA provides 16-bit signed/unsigned multiplication as well as signed/unsigned 32/16-bit

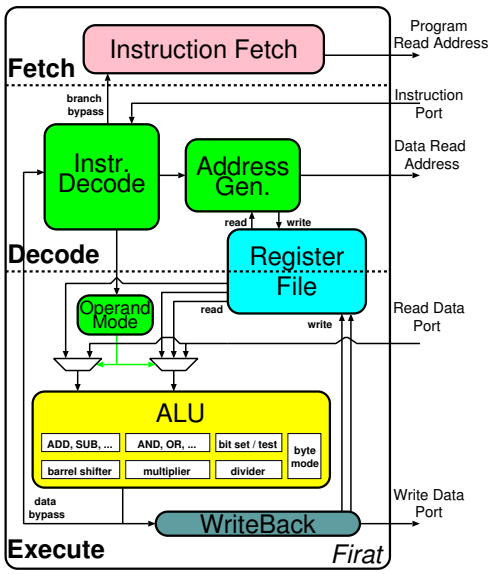

Figure 5.7: Architecture of the Firat microcontroller

integer division. Moreover, a barrel shifter is used for both arithmetic and logic single or multi-bit right/left shifts. Most of the instructions are executed in only one clock cycle, except program-flow (3 cycles), double data-move (2 cycles) and division (17 clock cycles). The instructions operate on either two or three operands. Both operands of the two-operand instructions can be either a register (with different addressing modes for the first operand), a direct memory address or a literal of various sizes (from 4 to 16 bits). The instructions with three operands have always the first operand as a working register, the second operand can be either a register with different addressing modes or a literal, and the third operand is always a register with different addressing modes. The supported addressing modes are register direct, register indirect (with pre or post-increment and decrement, and signed offset).

The second presented processing core, *TamaRISC*, is a custom-designed RISC architecture, shown in Figure 5.8. The core architecture focuses on minimizing the instruction set complexity, while still providing enough hardware support, especially regarding addressing modes, for efficient execution of biomedical signal processing applications. The processor has a three-stage pipeline (fetch, decode and execute stages). The core operates on a data word length of 16 bits, comprises 16 working registers and 3 external memory ports, for one instruction read, data read and data write, each in the same cycle. The core architecture is therefore similar to the *Firat* architecture, but with reduced complexity. The instruction word length is 24 bits, and every instruction has a single-word size. All instructions are executed in one cycle, which is guaranteed by the complete data bypassing inside the core for register and memory-write-back data. The main reduction of com-

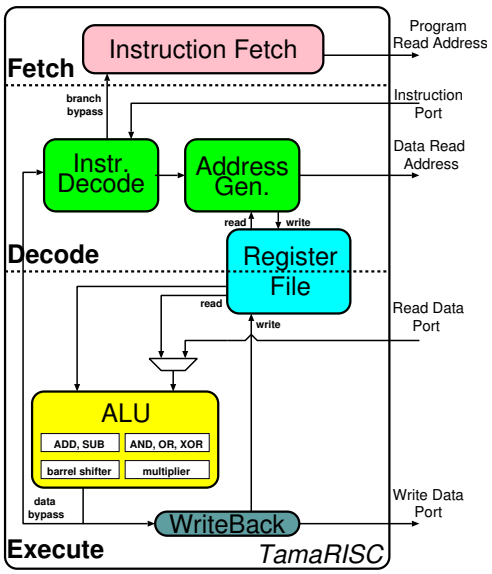

Figure 5.8: Architecture of the TamaRISC microcontroller

plexity lies in the ISA, which comprises a total of 11 unique instructions, with 8 arithmetic logic unit (ALU), 1 general data-move and 2 program flow instructions. All ALU instructions work on 3 operands, using the exact same addressing mode options for each instruction, which reduces the complexity of the architecture, since the operand fetch logic and the arithmetic operation are completely decoupled. Additionally, the instruction word encoding is designed as regular (fixed bit positions) and as simple as possible to allow for very efficient decoding of the operands and the different instruction words in general. The supported addressing modes are register direct, register indirect (with pre or post-increment and decrement) as well as register indirect with offset. The second operand also supports the use of 4-bit literals.

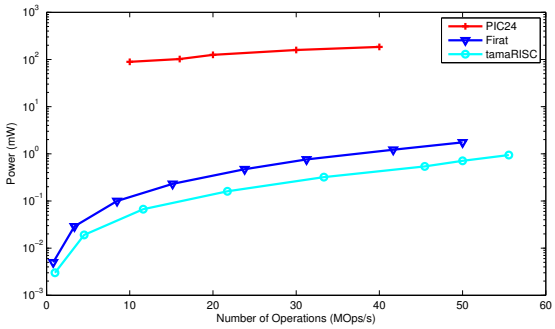

Figure 5.9: Power vs. Performance comparison of the processing cores

To compare the power vs. performance of the processing cores, *Firat* and

TamaRISC architectures have been implemented in a 90nm low leakage process technology. Figure 5.9 shows the power consumption of the processing cores and the PIC24HJ12GP201 microcontroller for various processing requirements. The power consumptions of the PIC24 microcontroller for different workloads are obtained from the datasheet, available online in [Tecd]. As opposed to the PIC24, both *Firat* and *TamaRISC* exploit voltage scaling by operating at the minimum voltage level satisfying processing requirements. As it can be seen in the figure, *Firat* achieves up to 890× power saving with respect to the PIC24. However, this saving is not achieved free of tradeoffs. As opposed to the PIC24, *Firat* architecture lacks of built-in units such as clock generator, voltage regulator, complete suite of mature and compatible peripherals, etc.

TamaRISC, is even more power efficient, achieving up to 52 % power savings with respect to *Firat*. This advantage is due to the reduced instruction set, as well as the efficient decoding of instructions and addressing modes of this architecture. *TamaRISC* also achieves higher energy efficiency compared to other state-of-the-art cores, developed for biomedical signal processing. On average, *TamaRISC* consumes only 11.8 pJ/Ops at 1.0 V. For the same supply voltage level (1.0 V), and although a 130 nm process was used, Kwong et al. [KC11] report 47 pJ/cycle energy consumption for their 16-bit core where the number of clock cycles per instruction is higher than one. In another work, Ickes et al. [ISP+11] introduce a 32-bit core implemented in 65 nm, and the energy consumption of the core [ISP+11] is estimated to be between 19.7 pJ/Ops and 27.0 pJ/Ops for 1.0 V. Compared to these state-of-the-art processing cores, *TamaRISC* features a significantly lower energy consumption per operation due to its simple architecture as well as its reduced instruction set.

5.4. Future lines of work

This thesis found an answer to each of the questions that were initially posed, but along the way, many other queries have popped up. Therefore, this section proposes the following ideas to be studied in the future:

1. Tackle the design of more efficient microcontrollers that use the benefits of advanced power reduction techniques such as sub-threshold voltage operation.

2. Include hardware blocks that work in collaboration with the microcontroller, performing tasks like filtering or compression. The implementation of these tasks in hardware instead of software will increase the energy efficiency of the WBSN platform.

3. Increase the size of the memories, since they are in most cases the main

bottleneck to port and optimize signal processing algorithms to the WBSN platform. The use of more power-efficient memory architectures or the definition a memory hierarchy will also help to reduce the energy consumption of this hardware block.

4. Consider the design of an ultra-low-power A/D converter, optimized for the requirements of biomedical applications, which require higher sampling frequency than other parameters typically used in generic WSNs.

5. Use energy scavenging mechanisms that can harvest energy from external sources (thermal gradient, kinetic energy, etc.), in order to achieve, in the best case, a zero external-energy WBSN.

6. Create multi-hop network topologies, where other intermediate nodes are used to reach the WBSN coordinator. Using this approach, the transmission range, and therefore the energy consumption, of the nodes can be significantly reduced.

7. Perform a detailed study of the impact of various body-area propagation conditions to obtain a realistic model of the wireless signal communication on the human body. This knowledge enables to accurately tune the transmission power of the radio to save as much energy as possible.

8. Low-complexity encryption and security of the network, since WBSN applications involve the transmission and storage of sensitive personal data.

9. Use more advanced sensing techniques like multiparametric or distributed sensing, which combine various signal measured by the same or different WBSN platforms.

10. Use high-level algorithms that not only consider the data of an individual, but combines and analyzes the data of a population. This can include the use of complex algorithms that can run on a PC, server or even a data center.

5.5. Publications

Here there is a list with the papers produced during my research:

Journal papers

(1) F. Rincón, A. Dogan, N. Khaled, D. Atienza, A. Burg. "Energy-Aware Design of Biomedical Wireless Body Sensor Networks". *Integration, the VLSI Journal*. 2012, to appear.

(2) F. Rincón, J. Recas, N. Khaled, D. Atienza. "Development and Evaluation of Multi-Lead Wavelet-Based ECG Delineation Algorithms for Embedded Wireless Sensor Nodes". *IEEE Transactions on Transactions on Information Technology in BioMedicine (TITB)*. 2011. vol. 15, pp. 854-865.

Conference papers

(1) I. Beretta, F. Rincón, N. Khaled, P.R. Grassi, V. Rana, D. Atienza. "Design Exploration of Energy-Performance Trade-Offs for Wireless Sensor Networks". In Proc. *IEEE/ACM Design Automation Conference (DAC)*. 2012.

(2) I. Beretta, F. Rincón, N. Khaled, P.R. Grassi, V. Rana, D. Atienza, D. Sciuto. "Model-Based Design for Wireless Body Sensor Network Nodes". In Proc. *IEEE Latin American Test Workshop (LATW)*. 2012.

(3) I. Bogdanova, F. Rincón, D. Atienza. "A Multi-lead ECG Classification Based on Random Projection Features". In Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2012.

(4) F. Rincón, N. Boichat, V. Barbero, N. Khaled, D. Atienza. "Multi-Lead Wavelet-Based ECG Delineation on a Wearable Embedded Sensor Platform". In Proc. *Computers in Cardiology (CinC)*. 2009.

(5) F. Rincón, N. Boichat, N. Khaled, D. Atienza. "ECG Delineation on a Commercial Wearable Embedded Sensor Platform". In Proc. *International Workshop on Wearable Micro and Nano Technologies for Personalized Health (pHealth)*. 2009.

(6) N. Boichat, N. Khaled, F. Rincón, and D. Atienza. "Wavelet-Based ECG Delineation on a Wearable Embedded Sensor Platform". In Proc. *IEEE International Workshop on Body Sensor Networks (BSN)*. 2009.

(7) F. Rincón, L. Gutiérrez, M. Jiménez, V. Díaz, N. Khaled, D. Atienza, M. Sánchez-Élez, J. Recas, G. de Micheli. "Implementation of an automated ECG-based diagnosis for a wireless body sensor platform". In Proc. *International Conference on Biomedical Electronics and Devices (BIODEVICES)*. 2009.

(8) F. Rincón, M. Paselli, J. Recas, Q. Zhao, M. Sánchez-Élez, D. Atienza, J. Penders, G. de Micheli. "OS-Based Sensor Node Platform and Energy Estimation Model for Health-Care Wireless Sensor Networks". In Proc. *Design, Automation and Test in Europe (DATE)*. 2008.

(9) F. Rincón, A. Susu, M. Sánchez-Élez, D. Atienza, G. de Micheli. "A Simulation Model for Wireless Sensor Networks Based on TOSSIM". In Proc. *Conference on Design of Circuits and Integrated Systems (DCIS)*. 2007.

(10) F. Rincón, F. Rivera, M. Sánchez-Élez, R. Hermida. "2D mesh NoC simulation tool using an FPGA based environment". In Proc. *Conference on Design of Circuits and Integrated Systems (DCIS)*. 2006.

(11) F. Rincón, F. Rivera, M. Sánchez-Élez, R. Hermida. "Herramienta de simulación de una NoC con topología de malla 2D usando un entorno basado en FPGA". In Proc. *Jornadas de Paralelismo*. 2006.

Bibliografía

[AAC+11] S. Abbate, M. Avvenuti, G. Cola, P. Corsini, J. Light, and A. Vecchio. Recognition of false alarms in fall detection systems. In *IEEE Consumer Communications and Networking Conference (CCNC)*, pages 23–28, jan 2011.

[AMRL09] R. Almeida, J.P. Martínez, A.P. Rocha, and P. Laguna. Multilead ECG delineation using spatially projected leads from wavelet transform loops. *IEEE Transactions on Biomedical Engineering*, 56:1996–2005, August 2009.

[BB95] T.D. Burd and R.W. Brodersen. Energy efficient CMOS microprocessor design. In *Proceedings of the Twenty-Eighth Hawaii International Conference on System Sciences*, volume 1, pages 288–297 vol.1, January 1995.

[BCSV06] A. Bonivento, L.P. Carloni, and A. Sangiovanni-Vincentelli. Platform based design for wireless sensor networks. *Mobile Networks and Applications*, 11(4):469–485, August 2006.

[BDCD11] L.S. Bai, R.P. Dick, P.H. Chou, and P.A. Dinda. Automated construction of fast and accurate system-level models for wireless sensor networks. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2011*, pages 1–6, March 2011.

[Bea08] G. Brettlecker et al. Technology in healthcare. In Michael Schumacher, Helko Schuldt, and Helkki Helin, editors, *CASCOM: Intelligent Service Coordination in the Semantic Web*, pages 125–139. Birkhäuser Basel, Basel, 2008.

[Bea09] E. J. Benjamin et al. Prevention of atrial fibrillation: report from a national heart, lung, and blood institute workshop. In *Circulation*, volume 119, pages 606–618, 2009.

[BGM+10] A. Burns, B.R. Greene, M.J. McGrath, T.J. O'Shea, B. Kuris, S.M. Ayer, F. Stroiescu, and V. Cionca. SHIMMER™: a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors*, 10(9):1527–1534, September 2010.

[biba] FreeRTOS real-time operating system.

[bibb] The GCC toolchain for the Texas Instruments MSP430 MCUs. `http://mspgcc.sourceforge.net`.

[BMT91] F. Badilini, A.J. Moss, and E.L. Titlebaum. Cubic spline baseline estimation in ambulatory ECG recordings for the measurement of ST segment displacements. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 13, pages 584–585, November 1991.

[BRK+12a] I. Beretta, F. Rincón, N. Khaled, P.R. Grassi, V. Rana, and D. Atienza. Design exploration of energy-performance trade-offs for wireless sensor networks. In *Proceedings of the IEEE/ACM Design Automation Conference (DAC)*, 2012.

[BRK+12b] I. Beretta, F. Rincón, N. Khaled, P.R. Grassi, V. Rana, D. Atienza, and D. Sciuto. Model-based design for wireless body sensor network nodes. In *Proceedings of the 13th IEEE Latin-American Test Workshop (LATW12)*, 2012.

[BS93] R.V. Boomgaard and A. Smeulders. Morphological multiscale image analysis. In *Proceedings of the First International Workshop on Mathematical Morphology and Its Applications to Signal Processing*, pages 143–150, Politecnica de Catalunya, Barcelona, Spain, 1993.

[CCGCG10] E. Casilari, J.M. Cano-García, and G. Campos-Garrido. Modeling of Current Consumption in {802.15.4/ZigBee} Sensor Motes. *Sensors*, 10(6):5443–5468, June 2010.

[CD89] C.H. Henry Chu and E.J. Delp. Impulsive noise suppression and background normalization of electromagnetism signals using morphological operators. *IEEE Transactions on Biomedical Engineering*, 36:262–272, 1989.

[Cea97] S. Cerutti et al. Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes. In *Computers in Cardiology 1997*, pages 77–80. IEEE, September 1997.

[Cea02] L. Clavier et al. Automatic p-wave analysis of patients prone to atrial fibrillation. *Medical & Biological Engineering & Computing*, 40(1):63–71, January 2002.

[Cea08] R. Couceiro et al. Detection of atrial fibrillation using model-based ECG analysis. In *Int. Conf. on Pattern Recognition*, pages 1–5, December 2008.

[CES04] D. Culler, D. Estrin, and M. Srivastava. Overview of sensor networks. *Computer*, 37(8):41–49, aug 2004.

[CK96] A. Cohen and J. Kovacevic. Wavelets: The mathematical background. *Proceedings of the IEEE*, 84:514–522, 1996.

[Cor09] Corventis. Piix, 2009. `http://www.corventis.com/AP/nuvant.asp`.

[DCA+12] A.Y. Dogan, J. Constantin, D. Atienza, A. Burg, and L. Benini. Low-power processor architecture exploration for online biomedical signal analysis. *IETCDS*, 2012.

[DGV04] A. Dunkels, B. Gronvall, and T. Voigt. Contiki - a lightweight and flexible operating system for tiny networked sensors. In *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, LCN '04, pages 455–462, Washington, DC, USA, 2004. IEEE Computer Society.

[DS88] J.D. DePree and C.W. Swartz. *Introduction to Real Analysis*. John Wiley, New York, 1988.

[EC508] ANSI/AAMI/ISO EC57:1998/(R)2008. Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms. 2008.

[FD09] G. Fang and E. Dutkiewicz. Bodymac: Energy efficient tdma-based mac protocol for wireless body area networks. In *9th International Symposium on Communications and Information Technology, 2009. ISCIT 2009*, pages 1455 –1459, sept. 2009.

[Fea06] V. Fuster et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation. *Circulation*, 114(7):257–354, 2006.

[FK11] M.O. Farooq and T. Kunz. Operating systems for wireless sensor networks: A survey. *Sensors*, 11(6):5900–5930, 2011.

[fQEWP85] The Common Standards for Quantitative Electrocardiography Working Party. Recommendations for measurement standards in quantitative electrocardiography. *European Heart Journal*, 6:815–825, 1985.

[GBPG09] B. Grundlehner, L. Brown, J. Penders, and B. Gyselinckx. The design and analysis of a real-time, continuous arousal monitor. In *6th IEEE International Workshop on Body Sensor Networks (BSN 2009)*, pages 15–21, jun 2009.

[GC97] V. Gutnik and A.P. Chandrakasan. Embedded power supply for low-power DSP. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 5(4):425–435, 1997.

[Gre69] T.N.E. Greville. *Theory and applications of spline functions*. Academic Press, New York, 1969.

[HEC+08] Z. He, J. Eggert, W. Cheng, X. Zhao, J. Millspaugh, R. Moll, J. Beringer, and J. Sartwell. Energy-aware portable video communication system design for wildlife activity monitoring. *IEEE Circuits and Systems Magazine*, 8(2):25–37, 2008.

[HJB+09] M.A. Hanson, H.C. Powell Jr., A.T. Barth, K. Ringgenberg, B.H. Calhoun, J.H. Aylor, and J. Lach. Body area sensor networks: Challenges and opportunities. *Computer*, 42:58–65, 2009.

[HMYLH+10] W. Hui-Min, L. You-Liang, M.C. Hou, L. Shih-Hsiang, B.S. Yen, H. Yu-Chieh, C. Lei-Chun, H. Shao-You, H. Sheng-Chieh, and J. Ming-Yie. A ±6ms-accuracy, 0.68mm^2 and 2.21μw QRS detection ASIC. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1372 –1375, jun 2010.

[IM10] A. Illanes-Manriquez. An automatic multi-lead electrocardiogram segmentation algorithm based on abrupt change detection. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2334–2337, 2010.

[Imp] Solar Impulse. Solar impulse: around the world in a solar airplane. `http://www.solarimpulse.com/`.

[Insa] Texas Instruments. CC2420 RF Transceiver. `http://www.ti.com/product/cc2420`.

[Insb] Texas Instruments. Code composer essentials. `http://www.ti.com/tool/msp-cce430`.

[Insc] Texas Instruments. Msp430f1611 microcontroller datasheets. `http://www.ti.com/product/msp430f1611`.

[ISP+11] N. Ickes, Y. Sinangil, F. Pappalardo, E. Guidetti, and A.P. Chandrakasan. A 10 pJ/cycle ultra-low-voltage 32-bit microprocessor system-on-chip. In *ESSCIRC (ESSCIRC), 2011 Proceedings of the*, pages 159–162. IEEE, September 2011.

[IvOH06] Z. Ihara, A. van Oosterom, and R. Hoekema. Atrial repolarization as observable during the PQ interval. *Journal of Electrocardiology*, 39:290–297, 2006.

[JBW+09] S.C. Jocke, J.F. Bolus, S.N. Wooters, A.D. Jurik, A.C. Weaver, T.N. Blalock, and B.H. Calhoun. A 2.6-μw sub-threshold mixed-signal ECG SoC. In *Symposium on VLSI Circuits*, pages 60 –61, june 2009.

[JD96] P.T. Jackway and M. Deriche. Scale space properties of the multiscale morphological dilation-erosion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:38–51, 1996.

[KC11] J. Kwong and A.P. Chandrakasan. An energy-efficient biomedical signal processing platform. *IEEE Journal of Solid-State Circuits*, 46(7):1742 –1753, july 2011.

[KKG07] G. Kambourakis, E. Klaoudatou, and S. Gritzalis. Securing medical sensor environments: The CodeBlue framework case. In *International Conference on Availability, Reliability and Security, 2007. ARES 2007*, pages 637–643. IEEE, April 2007.

[KMW08] G.S.A. Kumar, G. Manimaran, and Z. Wang. End-to-End Energy Management in Networked Real-Time Embedded Systems. *Parallel and Distributed Systems, IEEE Transactions on*, 19(11):1498–1510, 2008.

[KOKE06] H. Koc, O. Ozturk, M. Kandemir, and E. Ercanli. Minimizing Energy Consumption of Banked Memories Using Data Recomputation. In *Proceedings of the 2006 International Symposium on Low Power Electronics and Design, 2006. ISLPED'06*, pages 358–361, 2006.

[KSR03] C.H.I. Kim, H. Soeleman, and K. Roy. Ultra-low-power dlms adaptive filter for hearing aid applications. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 11(6):1058 –1067, dec. 2003.

[Lab] CNN Labs. Smartphone detects danger in a heartbeat. `http://edition.cnn.com/2011/10/24/tech/mobile/heart-monitor-smartphone-app/index.html`.

[LH05] B. Logan and J. H Healey. Robust detection of atrial fibrillation for a long term telemonitoring system. In *Computers in Cardiology*, pages 619–622, September 2005.

[LJC94] P. Laguna, R. Jané, and P. Caminal. Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database. *Computers and Biomedical Research*, 27:45–60, February 1994.

[LLK09] C. Li, H.B. Li, and R. Kohno. Reservation-based dynamic tdma protocol for medical body area networks. *IEICE Transactions*, 92-B(2):387–395, 2009.

[LMGM97] P. Laguna, R.G. Mark, A. Goldberger, and G.B. Moody. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in Cardiology*, volume 24, pages 673–676, September 1997.

[LMP+04] P. Levis, S. Madden, J. Polastre, R. Szewczyk, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler. Tinyos: An operating system for sensor networks. In *Ambient Intelligence*. Springer Verlag, 2004.

[LSD+06] J. Luprano, J. Sola, S. Dasen, J.M. Koller, and O. Chetelat. Combination of body sensor networks and on-body signal processing algorithms: the practical case of MyHeart project. In *International Workshop on Wearable and Implantable Body Sensor Networks, 2006. BSN 2006*, pages 4–79. IEEE, April 2006.

[LT10] H. Li and J. Tan. Heartbeat-driven medium-access control for body sensor networks. *IEEE Transactions on Information Technology in Biomedicine*, 14(1):44 –51, jan. 2010.

[LZA09] Y. Liu, W. Zhang, and K. Akkaya. Static worst-case energy and lifetime estimation of wireless sensor networks. In *Proceedings of the 2009 IEEE 28th International Performance Computing and Communications Conference (IPCCC)*, pages 17–24, 2009.

[LZT95] C. Li, C. Zheng, and C. Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42:21–28, January 1995.

[Mal89] S.G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37:2091–2110, 1989.

[MAO+04] J.P. Martinez, R. Almeida, S. Olmos, A.P. Rocha, and P. Laguna. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51:570–581, April 2004.

[MHM+10] W. Massagram, N. Hafner, C. Mingqi, L. Macchiarulo, V.M. Lubecke, and O. Boric-Lubecke. Digital heart-rate variability parameter monitoring and assessment asic. *IEEE Transactions on Biomedical Circuits and Systems*, 4(1):19 –26, feb. 2010.

[MK77] C.R. Meyer and H.N. Keiser. Electrocardiogram baseline estimation and removal using cubic splines and space-state computation techniques. *Computers and Biomedical Research*, 10:459–470, October 1977.

[MKAV11] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst. Compressed Sensing for Real-Time Energy-Efficient ECG Compression on Wireless Body Sensor Nodes. *Biomedical Engineering, IEEE Transactions on*, 58:2456–2466, 2011.

[MM83] G. B. Moody and R. G. Mark. A new method for detecting atrial fibrillation using r-r intervals. In *Computers in Cardiology*, volume 10, pages 227–230, 1983.

[MPL+77] P.W. Macfarlane, J. Peden, J. Lennox, M.P. Watts, and T.D. Lawrie. The Glasgow system. In *Proceedings of the IFIP Working Conference on Trends in Computer-Processed Electrocardiograms*, pages 143–150, North-Holland, New York, NY, USA, 1977.

[MZ92a] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 14:710–732, 1992.

[MZ92b] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:710–732, 1992.

[Net] IEEE 802.15 WPAN Task Group 6 (TG6) Body Area Networks. Ieee 802.15 wpan task group 6 (tg6) body area networks. `http://www.ieee802.org/15/pub/TG6.html`.

[NWP10] S. Nabar, J. Walling, and R. Poovendran. Minimizing Energy Consumption in Body Sensor Networks via Convex Optimization. In *Proceedings of the 2010 International Conference on Body Sensor Networks*, BSN '10, pages 62–67, Washington, DC, USA, 2010. IEEE Computer Society.

[PG85] F. Pérez-Gómez. *Cardiac Pacing*. Editorial Grouz, 1985.

[PGdVT07] J. Penders, B. Gyselinckx, N. de Vicq, and T. Torfs. Body area networks for multi-modal biomedical monitoring. In *International Conference on Wearable Micro and Nano Technologies for Personalized Health (pHealth)*, jun 2007.

[PL96] K.R. Park and C.N. Lee. Scale-space using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:1121–1126, 1996.

[PS05] J.A. Paradiso and T. Starner. Energy scavenging for mobile and wireless electronics. *IEEE Pervasive Computing*, 4(1):18–27, 2005.

[PW10] M. Patel and J. Wang. Applications, challenges, and prospective in emerging body area networking technologies. *IEEE Wireless Communications*, 17(1):80–88, february 2010.

[RBB+09] F. Rincón, N. Boichat, V. Barbero, N. Khaled, and D. Atienza. Multi-lead wavelet-based ECG delineation on a wearable embedded sensor platform. In *Proceedings of Computers in Cardiology*, pages 289–292, 2009.

[RGJ+09] F. Rincón, L. Gutierrez, M. Jimenez, V. Diaz, N. Khaled, D. Atienza, M. Sanchez-Elez, J. Recas, and G. de Micheli. Implementation of an automated ECG-based diagnosis for a wireless body sensor platform. In *Proceedings of the International Conference on Biomedical Electronics and Devices (BIODEVICES 2009)*, Porto, Portugal, 2009.

[Rol87] T. J. Rolfe. On a fast integer square root algorithm. *SIGNUM Newsl*, 22(4):6–11, 1987.

[RPR+08] F. Rincón, M. Paselli, J. Recas, Q. Zhao, M. Sánchez-Élez, D. Atienza, J. Penders, and G. de Micheli. OS-based sensor node platform and energy estimation model for health-care wireless sensor networks. In *Design, Automation and Test in Europe (DATE)*, mar 2008.

[SA] Shockfish SA. Tinynode 584 user's manual. `www.tinynode.com/?q=system/files/TN584_Users_Manual_v_1_3.pdf`.

[Sch71] L. Schamroth. *The disorders of cardiac rhythm.* Blackwell Scientific Publications, 1971.

[SCK02] Y. Sun, K.L. Chan, and S.M. Krishan. Ecg signal conditioning by morphological filtering. *Computers in Biology and Medicine*, 32, November 2002.

[SCK05] Y. Sun, K.L. Chan, and S.M. Krishan. Characteristic wave detection in ecg signal using morphological transform. *BMC Cardiovascular Disorders*, 5, September 2005.

[Sea98] S. Shkurovich et al. Detection of atrial activity from high-voltage leads of implantable ventricular defibrillators using a cancellation technique. *IEEE Trans. on Biomedical Engineering*, 45(2):229–234, February 1998.

[SL05] L. Sörnmo and P. Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Elsevier Academic Press, 2005.

[SML06] M.L. Soria, J.P. Martinez, and P. Laguna. A multilead wavelet-based ECG delineator based on the RMS signal. In *Proceedings of Computers in Cardiology*, pages 153–156, 2006.

[SRDV08] P. Suarez, C.G. Renmarker, A. Dunkels, and T. Voigt. Increasing ZigBee network lifetime with X-MAC. In *Proceedings of the workshop on Real-world wireless sensor networks*, REALWSN '08, pages 26–30, New York, NY, USA, 2008. ACM.

[SRS03] C. Schurgers, V. Raghunathan, and M.B. Srivastava. Power management for energy-aware communication systems. *ACM Transactions on Embedded Computing Systems*, 2(3):431–447, August 2003.

[SSS+10] I. Sánchez-Tato, J.C. Senciales, J. Salinas, L. Fanucci, G. Pardini, F. Costalli, S. Dalmiani, J.M. de la Higuera, Z. Vukovic, and Z. Cicigoj. Health @ home: A telecare system for patients with chronic heart failure. In *International Conference on Broadband and Biomedical Communications (IB2Com)*, pages 1–5. IEEE, dec 2010.

[Std03] IEEE 802.15.4 Std. Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs). In *IEEE*, 2003.

[Teca] Crossbow Technology. Micaz wireless measurement system. `http://bullseye.xbow.com:81/Products/Product_pdf_files/Wireless_pdf/MICAz_Datasheet.pdf`.

[Tecb] Crossbow Technology. Telosb mote platform. `http://www.xbow.com/pdf/Telos_PR.pdf`.

[Tecc] Microchip Technology. Pic24 16-bit microcontroller. http://www.microchip.com/en_us/family/16bit/architecture/PIC24H.html.

[Tecd] Microchip Technology. PIC24HJ12GP201 16-bit general pupose microcontroller. http://www.microchip.com/wwwproducts/Devices.aspx?dDocName=en520472.

[Tec09] Toumaz Technology. Sensium life pebble, 2009. http://www.toumaz.com/public/news.php?id=92.

[TG00] K. Tateno and L. Glass. A method for detection of atrial fibrillation using RR intervals. In *Computers in Cardiology*, pages 391–394, 2000.

[UHB+10] S. Ullah, H. Higgins, B. Braem, B. Latre, C. Blondia, I. Moerman, S. Saleem, Z. Rahman, and K. Kwak. A Comprehensive Survey of Wireless Body Area Networks. *Journal of Medical Systems*, pages 1–30, 2010.

[vHBW+04] A. van Halteren, R. Bults, K. Wac, N. Dokovsky, G. Koprinkov, I. Widya, D. Konstantas, V. Jones, and R. Herzog. Wireless body area networks for healthcare: the MobiHealth project. *Studies in Health Technology and Informatics*, 108:181–193, 2004.

[WC03] W. A. Wattingney and J. B. Croft. Atrial fibrillation hospitalizations triple since 1985, will continue to climb. 2003. http://www.americanheart.org/presenter.jhtml?identifier=3014151.

[WC04] A. Wang and A. Chandrakasan. A 180mv fft processor using subthreshold circuit techniques. In *IEEE International Solid-State Circuits Conference (ISSCC)*, pages 292 – 529 Vol.1, feb. 2004.

[Wit84] A.P Witkin. Image understanding. In *Norwood*, pages 79–95, NJ:Ablex, 1984.

[YFG07] C. Yeh, Z. Fan, and R.X. Gao. Energy-Aware Data Acquisition in Wireless Sensor Networks. In *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, pages 1–6, 2007.

[YKH+09] Y.H. Yassin, P.G. Kjeldsberg, J. Hulzink, I. Romero, and J. Huisken. Ultra low power application specific instruction-set processor design for a cardiac beat detector algo-

rithm,trondheim, norway. In *IEEE NORCHIP Conference, NORCHIP*, November 2009.

[YTP+09] R.F. Yazicioglu, T. Torfs, J. Penders, I. Romero, H. Kim, P. Merken, B. Gyselinckx, H.J. Hoo, and C.V. Hoof. Ultra-low-power wearable biopotential sensor nodes. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, sep 2009.

[ZSB+11] Y. Zhang, Y. Shakhsheer, A.T. Barth, H.C. Powell Jr., S.A. Ridenour, M.A. Hanson, J. Lach, and B.H. Calhoun. Energy efficient design for body sensor nodes. *Journal of Low Power Electronics and Applications*, 1(1):109–130, 2011.

One never notices what has been done;
one can only see what remains to be done.

Marie Curie