ENERGY CONSTRAINED
WIRELESS SENSOR NETWORKS
Communication Principles and Sensing Aspects

Erik Björnemo

January 2009

**UPPSALA
UNIVERSITET**

Abstract
Björnemo, E. 2009. Energy Constrained Wireless Sensor Networks. Communication Principles and Sensing Aspects. 279 pp. Uppsala. ISBN 978-91-506-2043-6.

Wireless sensor networks are attractive largely because they need no wired infrastructure. But precisely this feature makes them energy constrained, and the consequences of this hard energy constraint are the overall topic of this thesis. We are in particular concerned with principles for energy efficient wireless communication and the energy-wise trade-off between sensing and radio communication.

Radio transmission between sensors incurs both a fixed energy cost from radio circuit processing, and a variable energy cost related to the level of radiated energy. We here find that transmission techniques that are otherwise considered efficient consumes too much processing energy. Currently available sensor node radios typically have a maximum output power that is too limited to benefit from transmission-efficient, but processing-intensive, techniques. Our results provide new design guidelines for the radio output power. With increasing transmission energy –with increasing distance– the considered techniques should be applied in the following order: output power control, polarisation receiver diversity, error correcting codes, multi-hop communication, and cooperative multiple-input multiple-output transmissions.

To assess the measurement capability of the network as a whole, and to facilitate a study of the sensing-communication trade-off, we devise a new metric: the network measurement capacity. It is based on the number of different measurement sequences that a network can provide, and is hence a measure of the network's readiness to meet a large number of possible events. Optimised multi-hop routing under this metric reveals that the energy consumed for sensing has decisive impact on the best multi-hop routes. We also find support for the use of hierarchical heterogeneous network structures.

Model parameter uncertainties have large impact on our results and we use probability theory as logic to include them consistently. Our analysis shows that common assumptions can give misleading results, and our analysis of radio channel measurements confirms the inadequacy of the Rayleigh fading channel model.

Keywords: wireless sensor network, communication under processing cost, wireless channels, fading, probability theory as logic, uncertainty, measurement capacity

Erik Björnemo, Department of Engineering Sciences, Signals and Systems,
Box 534, Uppsala University, SE-75121 Uppsala, Sweden

Sammandrag
(Summary in Swedish)

Trådlösa sensornätverk – Energieffektiv kommunikation och mätaspekter

Små självständiga sensorer utrustade med mikroprocessorer och radio för trådlös kommunikation kan tillsammans bilda trådlösa sensornätverk. Tack vare den trådlösa kommunikationen kan sensornätverken användas för en rad uppgifter som annars skulle ha varit praktiskt ogenomförbara. Ett av de mest kända exemplen är ett forskningsprojekt på Great Duck Island i Kanada. Där studerades fåglar och deras häckningsbeteende med hjälp av sensorer som registrerade fåglarnas närvaro vid boet samt omgivande omständigheter såsom temperatur, luftfuktighet och lufttryck. Med trådbundna sensorer hade projektet blivit ytterst komplicerat och dyrt. Ett annat exempel är övervakning och diagnos av byggnadsverks och konstruktioners tillstånd för att kunna undvika olyckor som broras eller vingbrott hos flygplan. Vidare kan sensornätverk användas för hälsoövervakning, trådlös automatisk styrning av industriprocesser, diagnos och styrning av funktioner i "smarta" hus, detektion av utsläpp av farliga kemikalier, övervakning i säkerhetssyfte, målföljning i militära sammanhang, understöd i katastrofområden samt forskningsprojekt på svårtillgängliga platser som glaciärer.

Begränsad energi. Många användningsområden är beroende av den trådlösa kommunikationen, men tyvärr orsakar just trådlösheten ett av de största bekymren för sensornätverken och deras funktion, nämligen en starkt begränsad energitillgång. Denna begränsning får långtgående konsekvenser för utformningen av allt från elektronik till algoritmer för mätdataanalys. I avhandlingen undersöker jag konsekvenserna för den trådlösa kommunikationen och vilka avvägningar som måste göras på sensor- respektive nätverksnivå. Jag studerar specifikt avvägningen mellan den fasta energiförbrukningen i radioelektroniken och den rörliga nyttoförbrukningen av sändningsenergi. En slutsats är att många av dagens tillgängliga sensornoder inte kan utnyttja de

iii

optimala avvägningarna på grund av att den fasta energiförbrukningen är för stor i förhållande till den maximala nyttoenergin radion kan uppbåda. En del allmänt accepterade sändningstekniker bör därför ifrågasättas eftersom de medför just stor fast energiförbrukning. Jag undersöker också hur storleksförhållandet mellan kommunikationsenergi och mätenergi påverkar det totala nätverkets funktion och mätkapacitet, och finner att förhållandet kan ha avgörande betydelse för vilken kommunikationsteknik som bör användas.

Ofullständig information och osäkerheter. En viktig aspekt i alla beräkningar är vår osäkerhet rörande verkligheten och hur vi ska kunna dra befogade, kvantitativa, slutsatser även när viktiga faktorer bara är ofullständigt kända. Min kvantifiering av osäkerheter och användandet av denna i mina analyser är ett genomgående bidrag i avhandlingen, och visar tydligt att osökerheten inte bör ignoreras.

Jag använder i den här avhandlingen sannolikhetslära som en utvidgning av den deduktiva logiken och ser sannolikheter som representationer av ofullständig information. Denna logiska tolkning av sannolikhet – framförd av personer som Laplace, Jeffreys, Cox och Jaynes – bygger på ett fåtal generella och grundläggande principer för kvantitativ slutledning. Pierre Simon de Laplace och Harold Jeffreys använde framgångsrikt sannolikhetslära för kvantitativ slutledning, men det var Richard Cox som formellt ställde teorin på stabil grund. Han visade att de välkända produkt- och summareglerna i sannolikhetsläran är de enda räkneregler som uppfyller ytterst basala krav på en konsekvent slutledningsmetod som aldrig står i uppenbar motsättning till sunt förnuft. Den generella utgångspunkten ger sannolikhetsläran närmast obegränsat tillämpningsområde.

Det logiska synsättet skiljer sig i några avseenden markant från det etablerade synsättet. Det etablerade synsättet är att sannolikheter i någon mening är fysiska och i princip kan bestämmas som den relativa frekvensen för ett utfall – till exempel klave i en slantsingling – i ett oändligt antal upprepade försök. "Mätbar" sannolikhet har följaktligen giltighet endast för "slumpmässiga" fenomen, och är inte tillämpbar för andra storheter – de som är "deterministiska men okända". Trots att tolkningsfrågan kan verka rent filosofisk så får synsättet praktiska konsekvenser.

> Ställda inför insikten att frekvensdefinitionen av sannolikhetslära inte kunde tillämpas på de flesta verkliga vetenskapliga problemen så uppfanns ett nytt ämne – *statistik*.
>
> Sivia (1996)

Den frekventistiska tolkningen som genomsyrar konventionell statistik – förknippad med namn som Fisher, Feller, Neyman och Pearson – medför onödiga

begränsningar i antalet användningsområden för sannolikhetslära och berövar oss också ett par användbara, och ibland nödvändiga, verktyg. Jag behöver i mina problem kunna hantera otygsparametrar[1], ansätta prior-fördelningar baserade på "datalös" information, samt få ett relevant osäkerhetsmått i slutet av mina analyser.

Mätningar av trådlösa kanaler

Den trådlösa kanalens egenskaper har mycket stor inverkan på sändningsenergin. En viktig faktor är graden av variation i den mottagna signalen, den så kallade fädningen, och ett mycket vanligt antagande är att kanalen varierar enligt Rayleigh-modellen. Dock finns indikationer på att andra modeller bättre beskriver fädningen genom att ta med olika *grader* av fädning. För oss är fädningsgraden viktig eftersom den påverkar sändningsenergin kraftigt, och ett Rayleigh-antagande bör därför vara mycket välmotiverat för att användas. Jag har genomfört mätningar inomhus och utomhus under förhållanden som är typiska för sensornätverk: antennerna nära marken, väggen eller golvet. Min analys visar ett fädningsgraden relativt ofta avviker märkbart från Rayleigh-modellen och resultatet stöder istället användandet av Nakagami-m-modellen. Denna modell inkluderar fädningsgraden genom parametern m som därmed är en viktig parameter – en otygsparameter – i mina analyser.

Kraftiga kanalvariationer kan motverkas genom att sprida den sända informationen över flera kanaler som varierar olika, till exempel med hjälp av flera antenner, och därmed minska risken att information går förlorad. För de små sensornoderna är problemet att rumsspridning typiskt kräver ett större antennavstånd än sensornodens storlek. Däremot skulle antenner för polarisationsspridning kunna göras kompakta och därmed vara ett bra alternativ. Förutsättningen är dock att de olika polarisationskanalerna varierar oberoende av varandra. Mina mätningar visar lovande resultat med närmast obefintliga korrelationer mellan de vertikala och horisontella polarisationerna. Dessutom minskar korrelationen när fädningsgraden ökar – spridningsmöjligheten är bäst när den behövs mest.

Trådlös kommunikation med fasta kostnader

Jag studerar den totala energiförbrukningen för radiokommunikationen, både den fasta förbrukningen i radioelektroniken och den rörliga sändningsenergin,

---

[1]Engelska: *nuisance parameters*.

och valet av sändningsteknik bestäms till stora delar av kvoten

$$\rho' = \frac{\text{rörlig sändningsenergi}}{\text{total fast kretsenergi}},$$

som jag använder genomgående i avhandlingen. Det optimala valet tenderar att balansera fast och rörliga kostnader så att, typiskt, $0,5 < \rho' < 5$. En radio bör alltså, med lite marginal, ha en maximal kvot på åtminstone $\rho'_{\max} = 5$ för att kunna välja den optimala tekniken. Dagens radior har sällan $\rho'_{\max} > 0,5$.

Jag studerar följande sätt att väga av fast mot rörlig energiförbrukning.

Effektreglering. Sändaren justerar sin utsända effekt efter kanalens variationer till priset av att återkopplingsinformation om kanalens kvalitet medför mer sändningar. Effektreglering är ett bra alternativ för långsamma kanalvariationer, då lite återkoppling behövs, och kan ge besparingar för en del av de befintliga radioeneheterna.

Felrättande kodning. Genom att sprida informationen över tiden med hjälp av koder kan sändningsenergin minskas, men till priset av utökad sändningstid. Många av dagens radioenheter bör undvika kodning, men med lite större sändareffekt, $\rho'_{\max} > 1$ skulle besparingarna bli märkbara.

Adaptiv modulation. Då fasta kostnader dominerar, $\rho' < 1$, kan ökad sändningstakt spara energi genom att minska sändningstiden. Adaptiv kvadratur amplitudmodulering (QAM) möjliggör detta, men jag drar slutsatsen att den energikrävande elektronik som krävs för QAM inte uppvägs av besparingarna. Enklare metoder är att föredra.

Polarisationsspridning hos mottagaren. Mina mätningar visade att förutsättningarna för polarisationsspridning finns, och jag finner att polarisationsspridning är ett energieffektivt sätt att klara fädande miljöer. Jag undersöker två metoder för mottagaren, växling mellan två polarisationer och koherent kombination av polarisationerna. Växlingen medför mycket mindre fast energiförbrukning än kombinationen, men har å andra sidan sämre mottagningsprestanda. Slutsatsen är att växlingen är att föredra för dagens radioenheter, medan större sändningseffekt, med $\rho' > 1$, skulle göra kombinationsmetoden attraktiv.

Multi-hopp. När radions räckvidd inte räcker ända fram till slutmålet måste informationen skickas via andra sensorer – en kedja av hopp leder fram. Men är detta ett energieffektivt sätt? Studeras enbart

sändningsenergin är svaret ja, och en vanlig slutsats är därför att multi-hopp sparar energi. Jag kan dock genom att studera totalförbrukningen se att multi-hopp medför så stora fasta kostnader att det bör undvikas av alla befintliga radioenheter – hoppa bara när det krävs.

Kooperativ MIMO. Genom att samarbeta kan sensorerna, även om de bara har en antenn var, sprida informationen över flera rumskanaler in en MIMO-sändning[2]. I fädande miljöer kan detta spara mycket sändningsenergi, men precis som för multi-hopp medför samarbetet stora fasta energikostnader som överväger i de flesta fall. Dessutom är besparingen i sändningsenergi starkt beroende av fädningsgraden och min analys visar att osäkerheten är stor för eventuella besparingar.

Sensornätverkets mätkapacitet

Sett ur ett nätverksperspektiv är det inte säkert att energieffektiva metoder är de bästa. Om vissa sensorer belastas hårt, till exempel för att de måste vidarebefordra mycket data, kan det leda till att deras energi snabbt tar slut och att nätverket därefter fungerar sämre. Jag inför ett nytt mått, kallat mätkapacitet, som automatiskt inbegriper både energieffektivitet och balans. Jag utgår från *antalet olika sekvenser av mätningar* som sensornätverket kan göra på den givna energibudgeten. Om nätverket kan göra många mätningar – det är energieffektivt – betyder det stor mätkapacitet *endast om mätningarna är någorlunda jämnt fördelade* mellan sensorerna – om endast en nod har energi att mäta finns det bara en möjlig mätsekvens. Mätkapaciteten är användbar för att avgöra vilka kommunikationsmetoder som är bra, men också för att planera mätningar så att det minskar den kvarvarande mätkapaciteten så lite som möjligt.

Mätkapaciteten inbegriper också en avvägning mellan mätenergi och kommunikationsenergi. Om den fasta *kommunikationsenergin* per mätning är större än själva *mätenergin* så straffas metoder som multi-hopp eftersom varje hopp då kostar många mätningar. Förhållandet mellan mätenergin och den fasta kommunikationsenergin kan vara avgörande för valet av sändningsteknik. I de fall multi-hopp ändå lönar sig så kan optimerade sändningsvägar öka mätkapaciteten, men inte drastiskt. När jag däremot studerar hierarkiska nätverk finner jag en kraftigt ökad mätkapacitet som tillsammans med mina övriga resultat talar starkt för hierarkiska strukturer.

---

[2]MIMO, Multiple-Input Multiple-Output, flera sändar- och mottagarantenner.

Till min älskade familj
Anna, Hugo, Ebba och Albert

Contents

3   Fading and Polarisation Diversity, Channel Measurements and Analysis   73

4   Power Control and Rate Adaption   93

Ingen kan ha allting,
men alltid har man nå't
Alltid kan man vifta
på det lilla man har fått.

Pelle Svanslös

Acknowledgements

I will remember my years as a PhD student with warmth thanks to the open and friendly attitude of my colleagues and friends at the Signals and Systems group. Thank you all! I would especially like to thank my supervisor Anders Ahlén for providing a place in his research group and giving me freedom to explore the research topics I have found interesting. Thank you for your patience and the trust you have shown. I would also like to thank my assistant supervisor Mikael Sternad for encouragement and a fundamentally positive attitude towards new ideas. Thank you for your optimistic feedback. Catharina Carlemalm-Logothetis, it was a pity that we did not get the time to work together, but I certainly think you got your priorities right.

Daniel Aronsson and Mathias Johansson, it might be that I am the oldest of us, but in the end I find that it is I who look up to you. It has been a pleasure to walk the Bayesian path with you, Mathias leading the way and Daniel making sure that no stone is left unturned. I think we have all learned a lot from our journey and hope that, some day, our professional paths will join again.

A special thanks to Nora Masszi for being genuinely considerate.

Barbro and Dag Björnemo, my parents-in-law, thank you for your love and support, and for being such wonderful grandparents to my children.

Eva-May and Lars Ohlander, my parents, I owe a lot to you. You have always been there for me and I am very fortunate to have parents like you.

Finally, my beloved family, you mean the world to me. Hugo, Ebba and Albert, you bring the sunshine into my everyday life with your happy faces and your endless energy. Anna, words can not express how grateful I am for all your love and support during these years, especially during the last couple of months. I love you.

Erik Björnemo
Uppsala, December 2008.

Notation and Symbols

Here follows a list of commonly used symbols.

b	Number of bits per symbol
B	Bit error rate
c	Branch envelope correlation coefficient
C	Channel capacity
$\mathcal{C}$	Measurement capacity
d	Spatial distance
$\mathcal{E}$	Energy per unit (per bit, for instance)
$\mathcal{E}_{\mathrm{RP}}$	Total radio processing energy
$\mathcal{E}_{\mathrm{Pt}}$	Transmitter processing energy
$\mathcal{E}_{\mathrm{Pr}}$	Receiver processing energy
$\mathcal{E}_{\mathrm{rad}}$	Radiated energy
$\mathcal{E}_{\mathrm{T}}$	Transmission energy (including losses)
$\mathcal{E}_{\mathrm{S}}$	Measurement (sensing) energy
$E(x|y)$	Expected value for x given the value of y
$f_k(x)$	The kth constraint function in entropy maximisation
F_k	The kth constraint value $F_k = E(f_k(x)|I)$
g	Power amplifier efficiency degradation exponent
G	Power or energy gain
m	Nakagami-m fading figure
M_{l}	Link margin
n_{t}	Number of transmit nodes in cooperation
n_{r}	Number of receive nodes in cooperation
N_{f}	Receiver noise figure
N_0	Noise power spectral density
$P(A|x, I)$	Probability for A given x and I
$p(x|H, I)$	Probability density function for x given H and I
$\mathcal{P}$	Power consumption
$q(x)$	Ignorance (invariance) measure for x
R_{c}	Code rate for error correcting codes
w	Fraction of saved energy
W	The multinomial coefficient
W_{L}	The Lambert-W, or Product-Log, function
x	Channel power gain

α	Receiver-transmitter processing ratio $\mathcal{E}_{\mathrm{Pr}}/\mathcal{E}_{\mathrm{Pt}}$
β	Processing-processing ratio $\mathcal{E}_{\mathrm{RP},2}/\mathcal{E}_{\mathrm{RP},2}$
γ	Signal-to-noise ratio per bit
$\gamma_{\mathcal{P}}$	Signal-to-noise (power) ratio
$\Gamma(x)$	The gamma function
$\Gamma(x,y)$	The complementary incomplete gamma function
Δ	Diversity order gain
η	Power amplifier efficiency
κ	Power-law propagation loss exponent
λ	Sensor node density
ξ	Sensing-to-processing ratio $\mathcal{E}_{\mathrm{S}}/\mathcal{E}_{\mathrm{Pt}}$
ρ	Transmission-to-processing ratio $\mathcal{E}_{\mathrm{T}}/\mathcal{E}_{\mathrm{Pt}}$
ρ'	Transmission-to-total-processing ratio $\mathcal{E}_{\mathrm{T}}/\mathcal{E}_{\mathrm{RP}}$
$\rho_{\max}$	Transmitter's maximum ρ
σ	Standard deviation
ς	Channel envelope gain

We use the following notation when appropriate.

$\equiv$	Definition
$\overline{x}$	Average (expected value) of x
x_{med}	The median value of $p(x|I)$
x_{φ}	The φth percentile for $p(x|I)$
$\breve{x}$	Normalised x
$\mathrm{Nak}(\varsigma|m,\overline{\varsigma})$	The Nakagami-m probability distribution
$\mathrm{Gam}(x|m,\overline{x})$	The gamma probability distribution
$\mathrm{LogN}(x|\overline{x},\sigma_{\mathrm{dB}})$	The log-normal probability distribution

Chapter 1

Energy Constrained Wireless Sensor Networks

IMAGINE a set of small, self-contained, electronic devices equipped with sensors and the ability to communicate with each other without wires. These so-called sensor nodes can then together form a *wireless sensor network*. Such a network can monitor a region or phenomenon of interest and provide useful information about it by combining measurements taken by individual sensor nodes and then communicated over the wireless interface. We are in this thesis concerned with energy efficient wireless communication and energy based compromises between sensing and communication.

Wireless sensors with computing capabilities facilitate a range of applications that have previously been infeasible, or at lest too expensive. For instance, in a research project on Great Duck Island (Ontario, Canada), the breeding behaviour of a bird, Leach's Storm Petrel, was monitored by the use of a wireless sensor network (Mainwaring et al., 2002). Sensor nodes equipped with infrared sensors detected the presence of birds inside their nesting burrows, while other sensors registered environmental parameters such as temperature, pressure and humidity. Without the wireless sensors, the April-to-October monitoring would have been practically infeasible. Another application area gaining strong interest is that of structural health monitoring, in which wireless sensor networks are used to monitor structures such as bridges and nuclear power plants in order to detect damages or other changes in the structures. Yet other areas, to mention but a few, are health care, surveillance and security, wireless automation and military target tracking. Further examples can be found in Romer and Mattern (2004).

In many applications the key feature of a wireless sensor network is that it is just that, wireless. The use of wired node-to-node connections would in most applications constitute a severe complication, both practically and economically, in particular when hundreds of sensors are envisioned. But, while radio communication is a major enabling technology, the absence of wires is also the cause of one of the most prominent concerns, *limited energy*. Limited energy considerations is a nearly inescapable topic in wireless sensor network design as it imposes strict constraints on the network operation, and for this reason limited energy is the underlying theme of the present work.

1.1   Energy is a limited resource

Recharging batteries in a wireless sensor network is sometimes impossible due to the placement of the sensor nodes, but more commonly it is merely practically and/or economically infeasible. At any rate, it is widely recognised that, generally, energy is a strictly limited resource in wireless sensor networks and that the consequences of this limitation must be considered (Estrin et al., 2001, Shih et al., 2001, Sohrabi et al., 2000).[1]

Ultimately, if we want to have the sensor network performing satisfactorily for as long as possible, the energy constrained operation of the sensor nodes forces us to compromise between different activities in the network. Compromises are needed on the node level as well as on the network level. Saving energy is tantamount to finding the best compromise, the best trade-off, between different energy consuming activities and their design.

EXAMPLE 1.1 Communication or Processing (Pottie and Kaiser, 2000)

Assume that 1024 bits of data are transmitted 100 m at a carrier frequency of 1 GHz by the use of binary phase shift keying. The communication channel suffers from fourth order power distance loss (that is an attenuation loss proportional to d^4 where d denotes distance) and Rayleigh fading (severe fluctuations in channel quality). If our target bit error rate B equals 10^{-6}, then Pottie and Kaiser (2000) find that the transmission consumes approximately 3 J. Contrast this with a computer processor that performs 10^8

---

[1]A possible counter measure to limited energy is the use of so-called energy harvesting, that is techniques for extracting energy from environmental sources such as the sunlight. It appears to us that energy harvesting will be very important, but we note that most present suggestions deliver relatively little power and in addition require quite specific conditions. One can for this reason not rely on energy harvesting as a panacea.

instructions/J (a 100 MIPS/W processor). It will then execute 300 million instructions on the same energy budget as the transmission. Large energy savings could thus be achieved if the amount of data to transmit is reduced by local processing; we can trade a small increase in data processing energy for a larger decrease in communication energy. Assuming that the data consists of 128 eight-bit sensor readings from which we can compute a single eight-bit arithmetic average using less than one million instructions, we could save more than 99 percent of the original energy cost by sending only the average value.[2]

Based on the recognition in the literature that the radio is a prominent energy consumer (Raghunathan et al., 2002, Shih et al., 2001), we devote the present thesis largely to energy efficient radio communication and the related compromises.

1.1.1   Sharing resources: sensing and communication

Communication between sensor nodes is an idle exercise unless the nodes perform measurements and have relevant information to send. In the end, wireless communication is necessary but undesired. It is the information provided by the sensing that we are after, and the communication techniques should be chosen to allow the network as a whole to, simply put, deliver many measurements. The issue of energy efficient communication is of course implicitly included in this requirement, because the sensors and the radio share the same battery, but the matter is not quite as simple as choosing the most efficient transmission scheme from the textbook.

Because individual sensor nodes experience different communication and sensing energy costs – each possibly depending on the node's location, the network topology, the distribution of the monitored events, the communication environment, etc. – the choice of a certain communication scheme may be energy-wise good for one sensor node but bad for another. At the end of the day we do not care about individual nodes, but the overall network performance. However, the uneven energy consumption between nodes *will* influence the network's ability to provide good performance over time: in the extreme case, nodes in a certain area can be quickly drained of all en-

---

[2]We have modified the example slightly by correcting an error in the calculation and also exemplified an approximate energy saving for a special case. Further note that all assumptions are not given in the paper, but these specific details are in any case not needed for our present illustrative purpose.

ergy, stop functioning completely and thereby significantly decrease overall
network performance. So, when communication energy efficiency comes at
the price of a severely imbalanced energy consumption across the network,
the efficiency might not be worth its price. How do we know if it is? We
introduce in this thesis a novel way of quantifying a network wide sensing
resource, which we shall call the *measurement capacity* of the network, based
on the fundamental question of how many different measurement tasks the
network can respond to at a given energy budget. This metric automatically
trades off energy efficiency and energy balance, always having the sensing in
focus, and improves on the existing network lifetime metrics (Dietrich and
Dressler, 2006).

Having observed the importance of seeing the network as a whole, we
recognise that before we address the overall network performance we need to
understand the communication specific compromises that underpin it. With
a better understanding of energy efficient wireless communication in simpler
node-to-node scenarios, and the energy trade-offs involved, we will reduce
the risk of unwarranted assumptions on the network level. Moreover, in the
important class of (heterogeneous) hierarchic networks (Mhatre et al., 2005,
Yarvis et al., 2005) it turns out that low-level energy efficiency is closely
connected to the measurement capacity, much more so than in flat, non-
hierarchical, networks.

1.1.2   Wireless communication under processing costs

The ultimate limit on transmit energy efficient communication over a static
Gaussian channel was given by Shannon (1948a,b), and performance limits
for many other channel conditions are also well-explored in the communica-
tion theory literature, see (Goldsmith, 2005). But when we shift from the
traditional focus on *transmission* energy to *total* energy many of the common
"truths" in communication theory are invalidated. And this shift is neces-
sary in our context of energy constrained wireless sensor networks because
the amount of energy consumed by the sensor nodes' electronic circuits is
well in parity with the amount of transmission energy for most operating
conditions. Example 1.2 illustrates the profound impact that the inclusion
of processing energy can have.

EXAMPLE 1.2 Processing Energy and Intermittent Transmission (Youssef-Massaad, 2005)

The channel capacity defined by Shannon (1948a) for a bandlimited static channel with additive white Gaussian noise was derived under a pure *transmission* power constraint. Achieving error free reception at the Shannon transmission rate requires infinitely long channel codes and the energy-optimal transmission is continuous, spreading the energy of each symbol over a long time. Now, if the constraint is instead posed for the *total* energy, including a processing energy cost, Youssef-Massaad (2005) shows that the capacity achieving transmission may have to be intermittent. The presence of a non-negligible processing cost makes continuous transmissions sub-optimal energy wise, and changes the structure of an optimal approach entirely.

The most obvious analogy to the discussion above is manufacturing costs; the production of a "widget" will incur fixed costs (machines, factory buildings, etc.) and variable costs (raw material, labour, etc.). We would probably be dismissed if we ignored the fixed costs when pricing the widgets, unless the production would be on a scale large enough to make fixed costs negligible relative to the variable costs. In this sense, traditional communication theory is in the realm of "mass production" where only the variable transmission costs count. We are in this thesis considering the "small production series" of wireless sensor networks for which fixed processing costs do count.

When considering the energy trade-off between transmission and processing costs it is intuitively evident that it must be the relative, not the absolute, costs that matter: the same transmission choice would be energy-optimal whether we are considering kJ or nJ per bit as long as the relative cost is the same.[3] Curiously enough, this simple observations is almost never considered explicitly in the literature on energy constrained wireless sensor networks. Results are commonly given for specific absolute energies and system parameters, and are often presented in terms of a threshold distance: the inter-node distance at which one transmission technique becomes superior to another. There is nothing erroneous per se in this approach, but it introduces an unnecessarily large sensitivity to parameter choices that does not affect the relative cost. A consequence, which we exemplify later in

---

[3]Of course, to the real sensor node and its battery the opposite holds: it is the absolute values that matter.

Example 2.1, is that the threshold distance might be misleading if interpreted without care. For this reason we use, to the extent it is possible, the *transmission-to-total-processing ratio*

$$\rho' = \frac{\text{transmission energy}}{\text{total processing energy}},$$

which puts the focus on the energy compromise itself: just how dominating must the transmission energy be to motivate a processing intensive technique? In this way we avoid sensitivity to many model parameters, but certainly not all.

1.2   Design choices under uncertainty

Studying a node-to-node wireless link it is fairly straightforward to calculate the required transmission energy by the use of the well-established standard models of communication theory, *for a given set of model parameter values.* In reality, the model parameters for a planned wireless sensor network are not accurately known beforehand, and the impact of erroneous assumptions can be quite dramatic.

EXAMPLE 1.3 Sensitivity to Assumptions

Previously, in Example 1.1 borrowed from Pottie and Kaiser (2000), we saw that 3 J of transmission energy could be used to perform 300 million instructions on a processor. Now, let us consider two channel assumptions that went into their calculation, namely the fourth order propagation loss and the Rayleigh fading assumption. These assumptions would probably be considered within reason by most researchers and developers in the communications field, and they are quite commonly used. On the other hand, the use of a second order, free-space, propagation loss and a static Gaussian channel is also common in the communications literature. By the use of these assumptions we find that the transmission energy in Example 1.1 is reduced by a factor $2.2 \cdot 10^8$! Consequently, instead of the 300 million instructions Pottie and Kaiser concluded we had available, we now get *one single* instruction on the transmission energy budget. The conclusion is suddenly that we should avoid the computer processing.

Upon comparing Example 1.1 and Example 1.3 the unavoidable conclusion is that fixed parameter-value assumptions must be used with care and

only when we are quite sure they hold in reality, or when the impact of deviations from them is negligible. Still, it is the two extreme fading assumptions on Rayleigh fading and completely non-fading channels that are in almost exclusive use in the research on wireless sensor networks.

The design of a wireless sensor network is inherently a design under great uncertainty, and disregarding its presence can obviously lead to unwarranted conclusions. We will therefore make use of probability theory as extended logic to reach conclusions that are defensible under our present uncertainty regarding influential model parameters. Our interpretation and use of probabilities as carriers of incomplete information – in the spirit of scientists like Laplace, Jeffreys and Jaynes – differs considerably from the conventionally taught interpretation of probabilities as the long-run relative frequencies in repeated trials – following statisticians like Fisher, Neyman, Pearson and Feller. The frequentist interpretation maintains that probabilities are, at least in principle, physically measurable as relative frequencies and that probabilities pertain to "random" phenomena only. All other phenomena or entities about which we are uncertain, the ones that are "deterministic but unknown", can according to the frequentist theory not be associated with probabilities. Frequentist probability theory is unable to deal with them consistently.

> Faced with the realisation that the frequency definition of probability theory did not permit most real-life scientific problems to be addressed, a new subject was invented – *statistics*!

> Sivia (1996)

In contrast to the conventionally taught (frequentist) statistics, the recognition of probabilities as representing a specific state of knowledge facilitates a unified approach to inference problems. The theorems of Cox (1946) show that the ordinary rules of probability theory are (the only) consistent rules for quantitative reasoning under uncertainty. There is no restriction to relative frequencies – they are merely a special case – and the need to classify entities as "random" or "deterministic" is disposed of. Probability theory as logic helps us to do what we really want, perform the best possible inference with the information at hand, with the additional benefit of providing a defensible measure of our uncertainty.

We highlight these different attitudes toward probabilities, not to continue a century-long debate on the matter, but because we need tools that are not available to a frequentist, and we want to give the reader a chance to understand the rationale behind our approach and make sense of our analyses.

The extended logic approach to inference, mastered by Edwin Thompson Jaynes, deals effortlessly with two features which cause considerable trouble in conventional statistics (Jaynes, 2003). First, the parameters that we are uncertain about are not of primary interest but still require attention due to their impact on the results – they are *nuisance parameters*. Conventional statistics has no coherent principles to apply, while nuisance parameters represent no principal problem to someone using probability theory as logic. Second, we must base many of our conclusions on "non-data" prior information, and as the use of prior probabilities is denounced in the statistical theory but recognised as an integral part of inference in the extended logic framework, our choice is clear.[4]

1.3   Our topic and related work

The research topic of wireless sensor networks is a large umbrella covering a very wide range of interests. Almost whatever your interest is, you can find your niche under this umbrella. Consequently there is an enormous spread in the contributions to this field, representing different viewpoints, prerequisites and goals. Considering the wealth of interesting studies we feel that there are, probably due to the application dependent nature of the subject, surprisingly few fundamental results specific to energy constrained wireless sensor networks. In the list below we therefore give an overview of topics and techniques that illustrate some important aspects and general results, and references that mainly serve as starting points for further reading. The list is by no means comprehensive, and the reader is referred to surveys like Akyildiz et al. (2002), or handbooks like Swami et al. (2007) and Ilyas and Mahgoub (2005), for further reading and references.

Duty cycling. A key technique frequently used is to put nodes to sleep when they are not actively performing a task; there can be long idle periods. The electronics will in sleep mode consume orders of magnitude less energy than in active mode (Shih et al., 2001). We assume that duty cycling is used.

Multi-hop communications. Energy losses in wireless transmissions increases super-linearly, d^κ, $\kappa > 1$, with distance d and at some point several short hops will be more energy-efficient than a single hop. The assumption of multi-hop communication is so common that it almost

---

[4]We do not denounce the use of conventional statistics, but personally the author sees no benefit to do so once the efficiency and consistency of probability theory is appreciated.

appears synonymous to sensor network communication (Akyildiz et al., 2002), but the trade-off between the radios' processing and transmission energies is sometimes overlooked.

Cooperative transmissions. Multi-hopping is one cooperative transmission technique, but there are other forms of cooperative transmissions. One proposed technique to lower the communication energy consumption is cooperative MIMO[5] involving several nodes transmitting and/or receiving messages jointly, (Cui et al., 2004).

Distributed processing and local decisions. Sensor nodes need not be pure sensing devices but can be given some autonomy which allows them to make local decisions and only spend energy on long-range communication with the central unit when required by the situation. The processing could be performed by neighbouring nodes in cooperation.

Hierarchies and heterogeneous networks. Locally, it might be energy efficient to assign a leader node when, for instance, a few neighbouring nodes perform distributed processing. The use of a leader node can simplify communication and sleep scheduling (Bandyopadhyay and Coyle, 2003, Heinzelman et al., 2000). Another hierarchic alternative is to use different node types designed to be good at different tasks. Such heterogeneous networks can, for example, alleviate a majority of the sensor nodes from energy consuming computations on their ill-suited micro-processors and let a few nodes with more well-suited processors do the computations (Tsiatsis et al., 2005).

Sensor selection. By adaptively scheduling the activities of the sensor nodes the networks can utilise gathered information on the present state of affairs to inactivate sensors not needed and let the nodes best positioned perform the task (Yang and Heinzelman, 2008).

Scalable hardware. Closely related to duty cycling is scalable hardware solutions, in this case meaning a hardware platform which adaptively can scale its energy consumption by shutting down parts of the electronics or use it at different levels of performance (Wentzloff et al., 2004). For instance, by lowering the clock frequency of the processor one can sometimes decrease the power consumption more than the increase in required processing time: the energy consumption is reduced.

---

[5]Multiple Input Multiple Output

Wake-up radio. A problem with putting nodes to sleep is that they can
not be contacted other than during short scheduled time intervals.
One proposed solution is an always-on ultra-low power radio intended
only for external activation; an extremely small listening power allows
a large on-time (Gu and Stankovic, 2004). Such a radio would be
extraordinarily useful in event driven applications if it could be made
energy-efficient enough. The ideal would be a passive, zero power,
wake-up radio with sufficient sensitivity.

Quantisation, data fusion and packet aggregation. When sensor read-
ings are strongly correlated the amount of data to be sent over long
ranges can be reduced by local extraction of the important informa-
tion. Similarly, if multi-hopping is used the data from several sensors
can be fused, and packets aggregated, to facilitate more efficient com-
munication (Rajagopalan and Varshney, 2006).

We are in this thesis primarily concerned with energy constrained oper-
ation in fixed wireless sensor networks designed to monitor a specific spatial
region. In our setting there will always exist a central unit which is "re-
sponsible" for the network operations and to which the information finally
should reach. We are thus broadly considering the class of many-to-one data
gathering networks. In Figure 1.1 we illustrate this type of sensor network
together with some typical activities.

1.4 Outline and contributions

Chapter 2 introduces the models and the related assumptions that we will
make use of in subsequent analyses. These include energy trade-off models
and wireless channel models. We also give, for those unfamiliar with the
theory, an overview of the extension of deductive logic to probability theory
and discuss our use of the theory, which is to provide probability distributions
corresponding to the present uncertainties. The probability distributions we
will use are assigned at the end of the chapter.

A contribution which stretches across the thesis is the quantitative inclu-
sion of uncertainties regarding channel parameters and sensor node positions.
Our aim is to present defensible conclusions based on our present state of
knowledge, but also the uncertainty that they come with. In many cases
the uncertainty can be significant, for instance regarding the possible energy
savings from multi-hop communications.
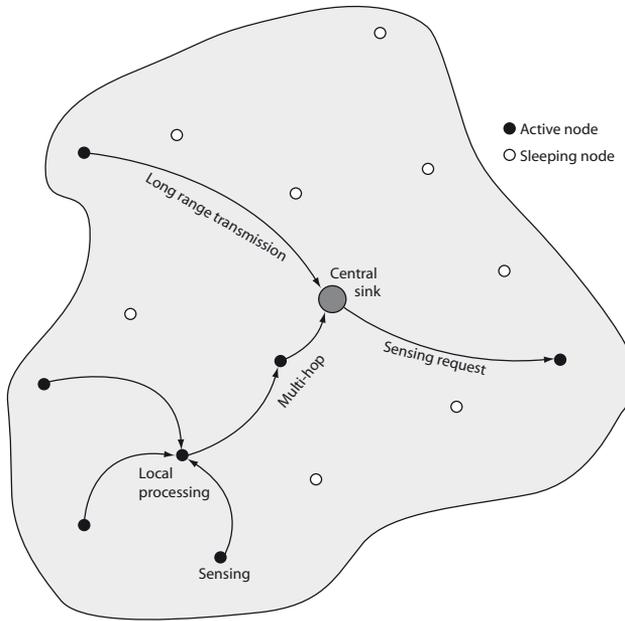
Figure 1.1: Data gathering many-to-one network. Wireless sensor nodes perform sensing tasks, perform local processing and transmit the requested information to the central sink (at some stage in the process). All operations are run on limited resources as no nodes, with the central sink as the possible exception, have wire-line power connection.

1.4.1 Channel measurements and analysis

Chapter 3. There are indications in the literature that the commonly used Rayleigh fading model has insufficient flexibility to describe many measured channels accurately. However, no reported measurement campaigns have been carried out to investigate this under typical sensor network conditions. We here analyse measurements performed in forest and office environments for both line-of-sight and non-line-of-sight channels. Transmit and receive antennas were positioned and/or moved close to the ground, the floor or the walls to capture conditions relevant for sensor networks. We find from our analysis that the Rayleigh fading model is indeed inadequate as a general fading model. In many cases the fading is less severe than asserted by the Rayleigh model, and our results support the use of the more flexible Nakagami-m fading model. This model lacks a precise physical interpretation, but by studying it from a maximum entropy viewpoint we can at least pin down the macroscopic constraint which, if imposed together with

an average power constraint, leads to the Nakagami-m model.

Additionally, our measurements reveal that different polarisations show very good diversity properties. We find very small correlations between the vertical and horizontal branches we analyse, in particular in the cases of severe fading. This is encouraging because these are the cases when channel diversity is most needed. Thanks to a compact configuration, polarised antennas should thus be of value to small wireless sensor nodes.

1.4.2   Communication under processing costs and uncertainty

By including processing energy – both at the transmitter and the receiver sides – and optimising for a small total communication energy, we find the important general trend that the optimal choice approximately equalises transmission and processing energy costs. To the extent balancing is achieved depends of course on the initial conditions, but there are still some general implications. First, we see that present radios often have insufficient output power (relative to their processing power consumption) to make use of the optimal transmission choices. For example, many sensor radios forces the node to resort to multi-hopping at a stage where a single-hop would be more energy efficient. On the other hand, the use of very powerful amplifiers is discouraged by the amplifier efficiency degradation with back off, and we include this effect in our calculations. Radio design should thus include, in addition to the concerns regarding absolute energy consumption, considerations of the maximum achievable transmission-to-total-processing ratio $\rho'_{\max}$. Second, the balancing effect accentuates the need to include processing energies: the processing energy cost will be non-negligible no matter how we turn.

Chapter 4.   Here we study the energy efficiency of transmit power control, error correcting codes and adaptive modulation. First, by the use of the Shannon limit we study some general, but idealised, properties of rate optimisation. We note the interesting result that the optimised spectral efficiency will be inversely proportional to the transmission-to-total-processing ratio ρ'. This signifies the impact of processing costs: only when the transmission costs are entirely dominant can the usual wide-band limit for energy consumption be used adequately.

The feedback required to perform (truncated) channel inversion penalises the use of power control, but not significantly as long as the fading is slow enough to avoid excessive feedback rates. Most, but not all, existing sensor radios will benefit from the use of power control, but there is significant

uncertainty regarding the amount of saved energy.

Error correcting codes reduce the transmit energy at the price of increased transmission time. The processing cost this introduces outweighs some of the coding benefits, and we study this trade-off for a class of adaptive block codes. Many existing radios will not benefit from error correcting codes.

Adaptive quadrature amplitude modulation (QAM) can be used to increase the transmit rate at short distances and thereby reduce the processing energy. However, we find that the processing intensive hardware required outweighs the possible benefits, a conclusion reached in a similar context by Shih et al. (2001).

Chapter 5 provides an assessment of receiver polarisation diversity from an energy perspective. Simple and processing-cheap schemes such as switched (switch-and-stay, threshold) diversity provide sizeable energy savings in severe fading, but is more sensitive to the degree of fading than more processing intensive schemes like maximum ratio combining. Almost all existing nodes can afford to pay a little extra processing for the robustness provided by switched diversity, while maximum ratio combining becomes attractive at larger transmission-to-total-processing ratios ρ'. In any case, we conclude that receiver polarisation diversity is certainly an attractive diversity technique as it achieves energy efficient transmissions in a simple manner.

Chapter 6 presents results for multi-hop communication. With the exception of applications where significant packet aggregation and/or data fusion is possible along a multi-hop route, we find that multi-hop within the transmission range of present sensor radios is wasteful of energy. One main reason is the drastically increased processing cost associated with the involvement of one or more relay nodes. Consequently, to avoid multi-hopping when it is wasteful, sensor radios should be designed to have more output power, relative to the processing power consumption, than they presently have. For the time being, it is advisable to hop as far as possible.

Due to the increased number of hops, the scheme also suffers from "inverse diversity", that is an increased risk of errors in the end-to-end communication. Multi-hop is therefore sensitive to fading, and especially in Rayleigh fading the performance is decreased markedly. Hence, the use of receiver polarisation diversity improves the situation considerably by improving the fading resilience. Furthermore, the multi-hop approach is sensitive to the position of the relay node, and the result is that the possible gains become

significantly more uncertain in sparse networks than in dense networks.

By joint optimisation of the number of hops and the error correcting code rate, we find that they balance their contributions as the number of hops increase. The result is that it suffices with reasonably high code rates in most cases, typically above 2/3. Also, the transmission-to-total-processing ratio ρ' stabilises at levels that, while out of reach for existing nodes, would be attainable with modest increases in transmit power.

Chapter 7 considers the use of cooperative MIMO, that is single-antenna nodes forming arrays and performing joint transmissions (here with the use of space-time block codes). Like multi-hopping, cooperative MIMO introduces too much additional processing energy to be useful for existing sensor radios. The gains over a single-input single-output (SISO) transmission rely to a great extent on the poor performance of SISO in severe fading. For this reason the gains are sensitive to the degree of fading, and small deviations from the Rayleigh case have large impact, and the energy gains become uncertain even at large transmission-to-processing ratios ρ'.

If we complement the SISO transmissions with receiver polarisation diversity to reduce the channel variations, and on top of that channel inversion through transmit power control, the relative gains from the use of cooperative MIMO are further reduced.

In the comparison between multi-hop and cooperative MIMO for long range transmissions, the results show great variability with the degree of fading and the propagation loss behaviour.

Summary. An overall picture now emerges, with somewhat fuzzy edges due to the present uncertainties, of when to apply different communication techniques. We show this overview in Figure 1.2. At the horizontal midpoint, $\rho' = 1$, transmission and processing energies are equal: to the right the former dominates, and to the left the latter dominates. Note that energy efficient multi-hop and cooperative MIMO is outside the transmission range of all the shown nodes, while techniques which can be introduced at lower processing costs are feasible alternatives.

1.4.3   Measurement capacity – a network resource metric

Chapter 8 outlines the need for another network wide metric than the commonly used concept of lifetime, which is typically defined as the time until the first node, or a certain percentage of the nodes, runs out of energy. We propose a novel metric, the network measurement capacity, which
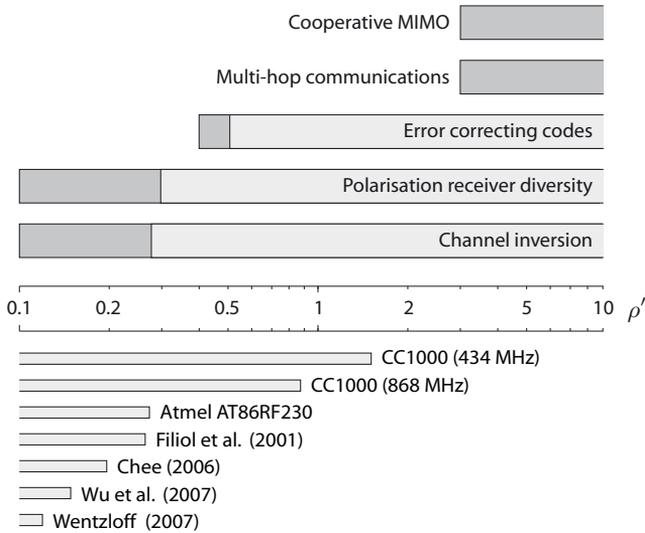
Figure 1.2: Approximate transmission-to-total-processing regions of energy efficiency for the considered techniques (upper bars), and the attainable transmission-to-total-processing ratios $\rho'_{\max}$ for existing radios (lower bars). The dark grey areas show indicate uncertain regions where no firm conclusion can be drawn, while light gray areas indicate more reliable energy savings.

is based on the *number of different sequences of measurements/detections that the network can perform successfully*, for a given distribution of energy. This leads to automatic incorporation of both energy efficiency and load balancing, two concepts that are widely recognised as desirable but have not been combined other than in *ad hoc* ways previously.

We make use of the measurement capacity metric to study optimal routing in a many-to-one network. It turns out that the sensing energy has a fundamental impact on the choice of routing pattern, and multi-hopping is very unfavourable if sensing consumes less energy than the communication processing.

By including shadowing of communication paths within the network, we find that the possibility to circumvent shadowing objects provides a much better motivation for using multi-hop than the gain from shorter hops that is most commonly put forward.

A network's measurement capacity can be increased by optimising the multi-hop routing patterns, but the gains are relatively small. They are somewhat larger in shadowed environments than in non-shadowed environments, but they are still modest and the single-hop approach performs sur-

prisingly well. A significant increase in measurement capacity is however achieved by introducing local sinks which form a second layer in the network. Such heterogeneous-hierarchic structures are strongly supported by this result, and the results from Chapters 4 to 7 give guidelines for how large the sub-networks in such tiered structures should be.

1.5   Outlook toward future research

The more we know about the channel characteristics, the more accurately we can predict which communication technique that will be most energy efficient. Considering the relative lack of sensor network specific measurements, we plan to carry out extensive channel measurement campaigns. With enough data from different environments, proper model selection can show us which models to use. Incorporation of relevant channel models in the development of higher layer protocols for routing and scheduling is crucial if these are to be efficient and robust under the large channel variations we can expect sensor networks to encounter.

The concept of measurement capacity will be studied and developed further. Application of the metric to specific problems, first and foremost to scheduling for target search and detection, will enable us to test it, generalise it and hopefully also to include the full aspect of the uncertainties present. Generalisations towards the related concept of *sensing capacity* (Aeron et al., 2007, Rachlin et al., 2005) offers an interesting avenue of research. The measurement capacity can also be useful in deployment planning, and the deployment of tiered networks under monetary constraints (Mhatre et al., 2005) and uncertainty deserve, in our view, more attention. The issues here involve robustness to uncertainties such as the ones we have explored in this thesis, but also regarding possible higher-level node failures.

Chapter 2

Models, Methods and Assumptions

THIS chapter comprises definitions of, and assumptions regarding, our basic models for the sensor nodes and their energy consumption, and also the important channel models used for calculating the wireless transmission costs. Importantly, we define the transmission-to-processing ratio which will accompany us throughout the thesis as our primary trade-off variable. We provide a fairly detailed account of probability theory as logic and how we make use of its recognition of probabilities as carriers of incomplete information to include our prior knowledge quantitatively in the subsequent chapters. Our assigned probability distributions for channel and deployment parameters are closing this chapter.

2.1   Energy models for networks and nodes

We conceptually divide each sensor node into three functional constituents: the sensor itself, the (data) processor and the radio; see Figure 2.1. They are all supported by the onboard battery and must share its energy. Now, let

$$
\begin{aligned}
\mathcal{P}_{\mathrm{S}} &\equiv \text{Sensor power consumption,} \\
\mathcal{P}_{\mathrm{DP}} &\equiv \text{Data Processor power consumption,} \\
\mathcal{P}_{\mathrm{R}} &\equiv \text{Radio power consumption.}
\end{aligned}
\tag{2.1}
$$

Since our focus is on *energy* the power consumptions $\mathcal{P}$ are not interesting per se, it is rather how they translate into an energy consumption through

17

Usage time T
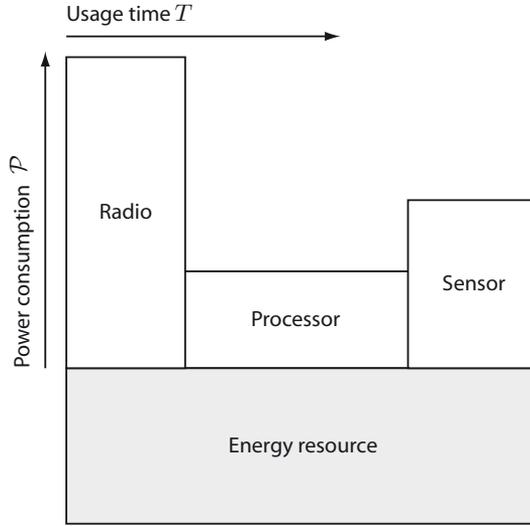
Figure 2.1: Illustration of the energy consumption in a wireless sensor node. The sensor(s), the data processor and the radio are the three major conceptual building blocks whose energy consumption should be minimised to maximise the node's lifetime.

the active time required to perform a certain task.[1] We will therefore, for a given "activity", study the energy consumed per "unit". Here, "activity" can refer to a wireless transfer of data, the reading of a value from the sensor with subsequent processing, or a combination thereof. The "unit" could be a data information bit, a measurement value, an in-network inference, or whatever is appropriate for the application. Typically we will consider energy per transmitted bit and/or energy per delivered measurement value when assessing different communication and sensing schemes. Let the per-unit energy consumptions, for sensing, data processing and radio communication respectively, be denoted

$$
\begin{aligned}
\mathcal{E}_{\text{S}} &\equiv \mathcal{P}_{\text{S}} T_{\text{S}}, \\
\mathcal{E}_{\text{DP}} &\equiv \mathcal{P}_{\text{DP}} T_{\text{DP}}, \\
\mathcal{E}_{\text{R}} &\equiv \mathcal{P}_{\text{R}} T_{\text{R}},
\end{aligned}
\tag{2.2}
$$

where $T_{(.)}$ denotes the active time for each part respectively.

---

[1]Of course, power limitations exist and one can not ignore the issue completely. For instance, the battery has a peak power limitation which sets limits for the sensor node activities. However, we will assume that such power limitations are well beyond the operating points we consider.

Assumption 2.1

Unless otherwise stated, all sensor nodes are identical, and the central sink has unlimited energy resources. ∎

Remark 2.1 *Our sensor energy model with three blocks is indeed somewhat simplistic. Some nodes can scale their processing power (Wentzloff et al., 2004), the radio usually have several modes of operation (sleep, wake-up, transmit, receive, etc.) and the three blocks can in reality not operate completely independent of each other. In spite of the model's simplicity it will serve its purpose well in our general trade-off analysis and provide energy-optimisation insights. Our main objective is to capture the first-order trade-off effects.*

2.2 Communication under processing costs

A wireless transmission of data consumes energy at both the transmitter and the receiver. Only a part of this total energy is actually radiated from the antenna, while there is an overhead part consumed by the nodes' active circuitry which process data and signals.

2.2.1 Trade-off model for the sensor node radio

Our goal is not a complete and detailed sensor radio energy model with every component explicitly included because such a model would most likely be too implementation specific. What we need is a reasonably generic energy model which captures the prominent trade-off between processing and transmission energies in the choice of transmission scheme. Toward this end, we begin with a simple basic model of the *total radio energy consumption* $\mathcal{E}_{\text{tot}}$ expended during transmission *and* reception of a bit (or other relevant entity). Let

$$\mathcal{E}_{\text{tot}} = \mathcal{E}_{\text{RP}} + \mathcal{E}_{\text{T}}, \tag{2.3}$$

where

$$\begin{aligned} \mathcal{E}_{\text{RP}} &\equiv \text{total Radio Processing energy consumption,} \\ \mathcal{E}_{\text{T}} &\equiv \text{Transmission energy consumption.} \end{aligned} \tag{2.4}$$

Hence, $\mathcal{E}_{\text{RP}}$ represents the energy consumption for all radio processing that does not change with the radiated output power while $\mathcal{E}_{\text{T}}$ represents the transmission costs which do change with output power (but not necessarily

in a linear manner). Processing energy is expended at both the transmitter
and the receiver and we write

$$\mathcal{E}_{\mathrm{RP}} \equiv \mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}} \tag{2.5}$$

where Pt and Pr denote transmit processing and receive processing respec-
tively.

If a change in transmission technique can reduce the transmission power
significantly, for example by employment of error correction coding, at a
small expansion of the transmission time one might save in total energy
consumption. The key issue is the relation between the radio processing
energy $\mathcal{E}_{\mathrm{RP}}$ and the transmission energy $\mathcal{E}_{\mathrm{T}}$. Because the transmission energy
$\mathcal{E}_{\mathrm{T}}$ generally increases with the inter node distance d it is common to present
a threshold distance $\tilde{d}$ at which two techniques $\mathcal{T}_1$ and $\mathcal{T}_2$ under consideration
perform equally well,

$$\mathcal{E}_{\mathrm{tot},\mathcal{T}_1}(\tilde{d}) = \mathcal{E}_{\mathrm{tot},\mathcal{T}_2}(\tilde{d}). \tag{2.6}$$

The threshold distance is appealing in that one can immediately relate it to
network deployments and required sensor node densities. However, as Ex-
ample 2.1 below shows, the threshold distance is a slightly deceptive metric
as it may conceal the influence of several parameters.

EXAMPLE 2.1 The Deceptive Threshold Transmission Distance

Cui et al. (2004) study the energy efficiency of multiple input multiple
output (MIMO) techniques and present a threshold transmission distance
$\tilde{d}$ above which MIMO requires less total energy per bit than a single-input
single-output (SISO) transmission. They find that a 2×1 MIMO system
will consume less energy than a SISO system for distances above $\tilde{d} \approx 60$ m.

At first, when Mr. A sees their result he thinks that 60 m is a long
distance and that MIMO is not suited for his sensor network. But then
he realises that the result is based on a free-space propagation assumption
where the propagation loss is proportional to d^2, and he recalculates the
threshold to suit the environment he is considering. By the use of a fourth
order model in which the propagation loss is proportional to d^4, he arrives
at $\tilde{d}_{\mathrm{new}} = (\tilde{d})^{2/4} = \sqrt{60} \approx 7.7$ m. This puts everything in a new light and
Mr. A is happy to choose MIMO for his sensor network application.

Mrs. B on the other hand, she plans a very sparse sensor network
and for her 60 m is a relatively short distance considering that the net-
work will cover about a square kilometre of open fields. But then, upon
closer inspection, she realises that the threshold $\tilde{d}$ is based on a link safety

margin $M_l = 10000$ (40 dB), that is a margin intended to guard against un-modelled phenomena and thereby ensure that the received power is large enough. Since her application environment is very predictable she is sure that a margin $M_{l,new} = 2$ (3 dB) is enough, and she recalculates the threshold. To her disappointment, the distance at which cooperative MIMO becomes superior has now grown to $\tilde{d}_{new} = \sqrt{M_l}\tilde{d} > 4200$ m. Therefore, SISO will be more energy efficient than cooperative MIMO in Mrs. B's application.

In fact, by the use of a single threshold distance one can "prove" almost anything by reallocating part of the transmission energy from a distance dependent propagation loss to any other parameter in the transmission model. Examples include link margins, noise levels, antenna gains and power amplifier efficiencies. We will instead focus on the trade-off between processing and transmission energies directly by using the transmission-to-processing ratio in place of distance.

The transmission-to-processing ratio

We define the transmitter's *transmission-to-processing ratio* as

$$\rho \equiv \frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{Pt}}}, \tag{2.7}$$

where $\mathcal{E}_{\mathrm{T}}$ and $\mathcal{E}_{\mathrm{Pt}}$ are the transmission and processing energies, respectively. One might object that since the whole processing cost will matter, not only transmit processing, it would be more appropriate to use the transmission-to-total-processing ratio

$$
\begin{aligned}
\rho' &\equiv \frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{RP}}} \\
&= \frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}}} \\
&= \frac{\rho}{1 + \mathcal{E}_{\mathrm{Pr}}/\mathcal{E}_{\mathrm{Pt}}}.
\end{aligned} \tag{2.8}
$$

We agree, but as we can not always express our results in terms of this total ratio ρ', for instance in Chapter 6, we sometimes choose to use ρ in (2.7) for pragmatic reasons. Moreover, by defining the *receiver-to-transmitter processing ratio*

$$\alpha \equiv \frac{\mathcal{E}_{\mathrm{Pr}}}{\mathcal{E}_{\mathrm{Pt}}} \tag{2.9}$$

we obtain from (2.8)

$$\rho' = \frac{\rho}{1 + \alpha}. \tag{2.10}$$

For any given receiver-to-transmitter processing ratio α one can thus easily find ρ' from ρ.

Assumption 2.2

To simplify interpretation of our numerical results, we will use the reasonable receiver-to-transmitter processing ratio $\alpha = 1$ whenever we must choose a value. Thereby, one can always find the total transmission-to-total-processing ρ' ratio from the transmission-to-processing ratio ρ numerically as $\rho' = \rho/2$. ∎

Observe that the transmission-to-processing ratio varies with the transmitter output power and is not a constant of the transmitter architecture. However, we assume that a sensor node radio has a maximum achievable transmission-to-processing ratio denoted

$$\rho_{\max} \equiv \max\left(\frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{Pt}}} \right). \tag{2.11}$$

Typically, $\mathcal{E}_{\mathrm{Pt}}$ is assumed constant and the maximum $\rho_{\max}$ corresponds to the use of the transmitter's maximum transmit power level.

The threshold ratio at which two schemes $\mathcal{T}_1$ and $\mathcal{T}_1$ are energy-wise equal in performance is denoted

$$\begin{aligned} \tilde{\rho}_{\mathcal{T}_{\mathrm{ref}}} \equiv{}& \text{threshold transmission-to-processing ratio} \\ & \text{for transmission scheme } \mathcal{T}_{\mathrm{ref}}. \end{aligned} \tag{2.12}$$

Here $\mathcal{T}_{\mathrm{ref}}$ is the technique used as a reference in the comparison, and generally $\tilde{\rho}_{\mathcal{T}_1} \neq \tilde{\rho}_{\mathcal{T}_2}$. Note that a threshold transmission-to-processing ratio $\tilde{\rho}$ has the appealing feature of being unaffected by transmit energy reallocations such as the one given in Example 2.1. It is only affected by reallocations between processing energy $\mathcal{E}_{\mathrm{Pt}}$ and transmission energy $\mathcal{E}_{\mathrm{T}}$ – exactly the trade-off we are interested in.[2]

Remark 2.2 *Let us stress that absolute values are important – it is in the end the actual energy consumption that counts – but in the present analysis we gain insights regarding the transmission-processing trade-off that are independent of several system parameters. These general insights can later be translated into absolute values, such as a threshold distance, for a specific sensor node type.*

---

[2]Independent changes in $\mathcal{E}_{\mathrm{Pt}}$ and $\mathcal{E}_{\mathrm{T}}$ also change the transmission-to-processing ratio ρ.

Now that we have the trade-off metric ρ', let us model its constituents $\mathcal{E}_{\mathrm{RP}}$ and $\mathcal{E}_{\mathrm{T}}$, beginning with the former.

2.2.2   Node radio hardware models

The radio processing energy $\mathcal{E}_{\mathrm{RP}}$ of (2.5) is consumed by oscillators, mixers, filters, etc. The energy consumption is highly implementation specific and ranges from nJ/bit to μJ/bit depending on the radio architecture, see Wu et al. (2007) and Shih et al. (2001). Considering all the possible architectures and design choices, it is far outside the scope of this work to devise and use a detailed and generic hardware energy model. Apart from the maximum transmission-to-processing ratio ρ_{max} in (2.11) and the receiver-to-transmitter processing ratio α in (2.9), we include only a model of the transmitter's power amplifier efficiency.

Efficiency degradation of power amplifiers

The hardware characteristic which we find most important in the present trade-off analysis is the degradation in power amplifier efficiency,

$$\eta \equiv \frac{\mathcal{E}_{\mathrm{rad}}}{\mathcal{E}_{\mathrm{T}}}, \tag{2.13}$$

with decreasing radiated power $\mathcal{E}_{\mathrm{rad}}$ (amplifier back-off). The reason for us to include it is the considerable impact it has on the achievable energy savings from transmission efficient techniques. Mikami et al. (2007) argue that a reasonable amplifier model is

$$\mathcal{E}_{\mathrm{T}} = \mathcal{E}_{\mathrm{T,max}} \left(\frac{\mathcal{E}_{\mathrm{rad}}}{\mathcal{E}_{\mathrm{rad,max}}} \right)^{1/g} \tag{2.14}$$

where

$$g \equiv \text{power amplifier degradation exponent,}$$
$$\mathcal{E}_{\mathrm{rad}} \equiv \text{radiated energy per bit.} \tag{2.15}$$

According to Mikami et al. (2007), the degradation exponent is restricted to $2 \leq g \leq 2.8$, with a typical value $g = 2.6$. Rewriting (2.14) we find that the power dependent amplifier efficiency in (2.13) can be expressed

$$\eta = \eta_{\mathrm{max}} \left(\frac{\mathcal{E}_{\mathrm{rad}}}{\mathcal{E}_{\mathrm{rad,max}}} \right)^{1-1/g}, \tag{2.16}$$

where η_{max} is the maximum efficiency attained for maximum output power: $\eta = \eta_{\mathrm{max}}$ when $\mathcal{E}_{\mathrm{rad}} = \mathcal{E}_{\mathrm{rad,max}}$ (or when $g = 1$). The degradation model (2.14) signifies that a reduction in radiated energy is not translated proportionally to the transmission related energy $\mathcal{E}_{\mathrm{T}}$ because amplifiers are typically designed to operate at maximum output power, and therefore work less efficiently when backed-off to lower levels. A result quantitatively similar to (2.14) was obtained by Haenggi (2003) who found an empirical model from a study of amplifier data sheets. In Figure 2.2 we show the power amplifier efficiency according to (2.16) and the proposal of Haenggi (2003, Definition 2). As one can see, the two models are in fairly good agreement. Henceforth, we will use (2.14) under the following assumption:

Assumption 2.3

The power amplifier degradation exponent in the Mikami model (2.14)

$$g = 2, \tag{2.17}$$

unless otherwise stated.   ■

Characteristics of present sensor node radios

Here we give examples of radio modules and designs suited for wireless sensor networks. It is hard to find precise quantifications of ρ_{max} for the radios, but we make rough estimates from data sheets and publications where enough data has been found.

Chee (2006) Discusses four transmitter types designed for energy efficiency, namely direct conversion, direct modulation, injection lock, and an active antenna solution. The active antenna transceiver is the most efficient and we find that its maximum transmission-to-processing ratio $\rho_{\mathrm{max}} = 2.7$ when tested in isolation. This applies however not to a fully functional node, and when the whole solution is tested the maximum ratio is only $\rho_{\mathrm{max}} \approx 0.4$

Wu et al. (2007) Low power 17 GHz front end consuming 16 mW in transmit mode at its maximum radiated output power of 0.75 mW. Assuming a total efficiency of 10 percent, the maximum transmission-to-processing ratio becomes $\rho_{\mathrm{max}} \approx 7.5/8.5 \approx 0.9$. The receiver consumes 17.5 mW, meaning that the receiver-to-transmitter processing ratio $\alpha \approx 17.5/8.5 \approx 2.1$. Consequently, $\rho'_{\mathrm{max}} \approx 7.5/26 \approx 0.3$.
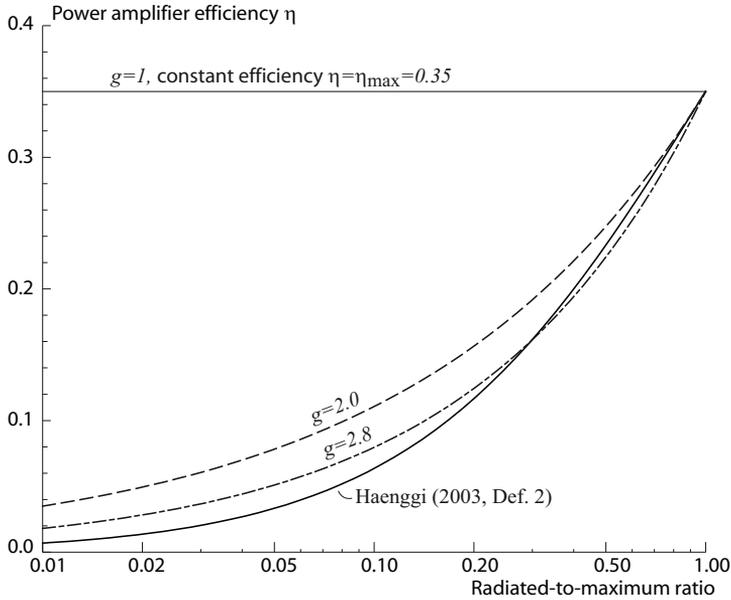
Figure 2.2: The transmitter's efficiency η is reduced when the amplifier is backed off from its maximum output. The g-curves correspond to the model due to Mikami et al. (2007). The horizontal axis displays the radiated-to-maximum ratio $\mathcal{E}_{\mathrm{rad}}/\mathcal{E}_{\mathrm{rad,max}}$. Energy gains from backing off from $\mathcal{E}_{\mathrm{rad,max}}$ to $\mathcal{E}_{\mathrm{rad}}$ are diminished by this efficiency degradation.

Atmel AT86RF230 Low power 2.4 GHz transceiver (ZigBee) with $\rho_{\max} \approx$ 0.74. It has a receiver-to-transmitter ratio $\alpha \approx 1.6$, and recalculating the equivalent $\rho_{\max}$ corresponding to $\alpha = 1$ we find that $\rho_{\max} \approx 0.56$. Note that it is not a full sensor node, only the radio.

Chipcon CC1000 Transceiver for 434 MHz and 868 MHz. Our estimate of the maximum ratios are $\rho_{\max} \approx 3.85$ and $\rho_{\max} \approx 1.95$ for the two frequencies. Receiver-to-transmitter ratios are $\alpha = 1.4$ and $\alpha = 1.12$ respectively, and recalculating for $\alpha = 1$ yields the equivalent $\rho_{\max} \approx$ 3.2 and $\rho_{\max} \approx 1.84$. Only radio, no other functionalities.

Filiol et al. (2001) Bluetooth receiver with maximum ratio $\rho_{\max} \approx 1$. The receiver-to-transmitter ratio is $\alpha = 2.73$ and the equivalent transmission-to-processing ratio corresponding to $\alpha = 1$ is thus $\rho_{\max} = 0.54$.

Wentzloff (2007) Ultra wideband transceivers with $\rho_{\max} = 0.24$.

From an energy consumption perspective one may ask whether it is desirable to have a large $\rho_{\max}$ or a small $\rho_{\max}$. Maybe it does not matter? First, we

observe that it is easy to achieve almost any $\rho_{\max}$ we please by the use of a large (or small) power amplifier, so the numbers above should not be seen as representing current *limitations* in $\rho_{\max}$. They serve as a reference for typical values and indicate that values slightly below one are most common, but that transceivers with larger $\rho_{\max}$ exist. Second, we note that ρ says nothing about the energy-efficiency since it is the ratio of to energy consumptions: a specific ρ can pertain to any energy per bit. The maximum transmission-to-processing ratio is therefore not a direct measure of transceiver quality. However, as our analysis will show, there may be strong reasons not to design radios with very small or very large $\rho_{\max}$:

- Sensor node radios with $\rho_{\max} < 0.1$ will always operate with processing energies dominating transmission energies. They can thus not benefit from transmission techniques that are transmission-efficient; the optimum transmission-processing trade-off can be infeasible for this type of node. For instance, these sensor node can be forced to use multi-hop transmissions when a longer single hop would have consumed less energy (see Chapter 6).

- Sensor node radios with $\rho_{\max} > 10$ are flexible (assuming that output power is adjustable) but pay an energy penalty caused by the power amplifier back-off degradation described by (2.14). When operating significantly below its maximum output transmit power, the sensor node radio dissipates an unnecessarily large amount of energy through its power amplifier.

We return to this discussion in Chapter 9 when we know more about the transmission-processing compromise.

Having set down our very basic radio processing model, let us turn to the more intricate modelling of the transmission energy consumption $\mathcal{E}_{\mathrm{T}}$.

2.2.3 Transmission energy and channel models

The transmission cost $\mathcal{E}_{\mathrm{T}}$ is the energy consumed from the onboard battery by the transmitter's power amplifier. It is strongly influenced by the wireless channel between the nodes, the environment in which the sensor network is deployed. Considering energies per bit, we assume, in accordance with most standard textbooks on wireless communication, the validity of the following wireless link model (Proakis, 2001, Sec. 5.5.2).

Assumption 2.4

To achieve a target signal-to-noise ratio per bit γ at the receiver's symbol detector, with a margin M_l, the transmitter antenna must radiate an average energy per bit $\mathcal{E}_{\mathrm{rad}}$ fulfilling

$$\mathcal{E}_{\mathrm{rad}} = M_l \cdot \frac{N_f}{G_{\mathrm{ant}}} \cdot x^{-1} N_0 \gamma, \tag{2.18}$$

where

$$
\begin{aligned}
M_l &\equiv \textit{link safety margin,}\\
N_f &\equiv \textit{receiver noise figure,}\\
G_{\mathrm{ant}} &\equiv \textit{antenna gain relative to isotropic radiation,}\\
x &\equiv \textit{channel gain,}\\
N_0 &\equiv \textit{noise power spectral density at the receiver,}\\
\gamma &\equiv \textit{received signal-to-noise ratio per bit, } \mathcal{E}_{\mathrm{rec}}/N_0,\\
&\qquad \textit{where } \mathcal{E}_{\mathrm{rec}} \textit{ is the received energy per bit.}
\end{aligned}
\tag{2.19}
$$

■

In (2.18), the first factor M_l is a design parameter intended to cover unpredicted variations in the other parameters; a safety link margin. The second factor N_f/G_{ant} is directly dependent on the hardware design. The third factor $x^{-1} N_0 \gamma$ is determined by the transmission environment – the channel gain between transmitter and receiver and the ambient noise level at the receiver – and the signal quality γ required for the modulation and detection scheme to attain the performance requirements.

Remark 2.3 *The* signal-to-noise ratio per bit γ *is the ratio between the received energy per bit and the noise power spectral density,*

$$\gamma = \frac{\mathcal{E}_{\mathrm{rec}}}{N_0}. \tag{2.20}$$

Both these quantities represents energy and are given in Joules (J). Hence, γ is a ratio of energies, *as opposed to the* signal-to-noise ratio $\gamma_{\mathcal{P}}$ *which is the ratio between received signal and noise* powers. *When the suffix* per bit *is used we are always referring to energy, not power.*

Observe that M_l, N_f, G_{ant} and N_0 in (2.18) are important radio design parameters, but they will have no impact in our analysis since they are common to all transmission schemes we examine. Thanks to our use of the transmission-to-processing ratio ρ in (2.7), these factors will cancel out. The channel gain x is far more influential in our analysis.

The linear multi-path radio channel model

The radiated energy per bit $\mathcal{E}_{\mathrm{rad}}$ in (2.18), required to fulfill the performance goals at the receiver, is strongly dependent on the properties of the wireless channel embodied in the channel power gain x. So, what can be said about the channel and how should it be modelled? The physical models for calculation of radio wave propagation are readily available through Maxwell's equations. In principle, we could calculate the field strengths at the receiving antenna as a function of the transmitter and the environment by specifying all boundary conditions. The unsurmountable practical problem facing us is of course that the boundary conditions are both extremely complex and unknown to us. We must therefore resort to much simpler models that still capture the channel characteristics needed to design a communication system; models including macroscopic constraints – which we are able to find in practice – rather than the microscopic constraints – which would be too complex to use even if they were obtainable.

In Appendix 2.B we describe the linear filter multi-path model in some detail and motivate our modelling choices and assumptions more closely by reference to the literature. The reader unfamiliar with radio channel modelling may consult the appendix at this point; otherwise it used mostly for reference.[3] In short, we will make use of the following macroscopic assumptions regarding the wireless channels.

Assumption 2.5

The multi-path channels between the wireless sensor nodes are frequency non-selective (flat) with a single resolvable multi-path component modelled by a single complex-valued filter tap. ∎

With respect to the linear single-tap model we define

$$\begin{aligned}
\varsigma &\equiv \text{the channel's amplitude gain,} \\
x &\equiv \text{the channel's power gain.}
\end{aligned} \tag{2.21}$$

Regarding the gains ς and x of the channel we consider three types of variations.

- *Fading* through constructive/destructive interference between path contributions.

---

[3]For comprehensive and detailed accounts of radio channel modelling we refer the reader to the works by Parsons (2001) and Vaughan and Andersen (2003). The books by Goldsmith (2005) and Proakis (2001) also treat channel modelling, but with stronger focus on the channel's impact on communication performance.

- *Shadowing* by objects.

- *Propagation loss* due to power dissipation over distance.

The combined effect is modelled as common power gain

$$x = x_l x_s x_f, \tag{2.22}$$

where x_l denotes propagation loss, x_s denotes shadowing and x_f denotes fading. The latter two are normalised to unit average gains $\overline{x}_s = 1$ and $\overline{x}_f = 1$.

Assumption 2.6

The (small-scale) fading is well characterised by the Nakagami-m fading figure m. We use the corresponding gamma distribution

$$p(x_f|m, \overline{x}_f, I) = \mathrm{Gam}(x_f|m, \overline{x}_f)$$
$$\equiv \frac{m^m}{\Gamma(m)\overline{x}_f^m} x^{m-1} e^{-m\frac{x_f}{\overline{x}_f}} \tag{2.23}$$

for x_f, with the normalised average $\overline{x}_f = 1$. ∎

Over ranges where the propagation loss and shadowing effects x_l and x_s are constant, all variations are due to fading and we can use $p(x|m, \overline{x}, I) = \mathrm{Gam}(x|m, \overline{x})$, where $\overline{x}$ is the average channel power gain. The two main reasons for the Nakagami-m assumption are the following:

1. The common Rayleigh fading model has proven insufficient for many environments, see Appendix 2.B.

2. The degree of fading has significant impact on transmission energy, see Appendix 2.C. The Nakagami-m model facilitates assessment of these effects, while a Rayleigh model precludes it.

Observe that the Nakagami-m model includes the Rayleigh model, $m = 1$, and the static channel, $m \to \infty$, as special cases.

Assumption 2.7

The (large scale) shadowing effects are well characterised by their logarithmic standard deviation σ_{dB}. We use a log-normal assignment

$$p(x_s|\overline{x}_s, \sigma_{\mathrm{dB}}, I) = \mathrm{LogN}(x_s|\overline{x}_s, \sigma_{\mathrm{dB}})$$
$$= \frac{1}{x_s} \cdot \frac{\mu}{\sqrt{2\pi}\sigma_{\mathrm{dB}}} e^{-\frac{1}{2\sigma_{\mathrm{dB}}^2}\left(\mu \ln(\overline{x}_s/x_s) - \frac{\sigma_{\mathrm{dB}}^2}{2\mu}\right)^2} \tag{2.24}$$

for x_s, with the normalised average $\overline{x}_s = 1$ and $\mu \equiv 10/\ln(10)$. ∎

Over ranges with constant propagation loss, the average channel gain is denoted $\overline{\overline{x}}$ to signify an average over both fading and shadowing. Then we can use $p(\overline{x}|\overline{\overline{x}}, \sigma_{\mathrm{dB}}, I) = \mathrm{LogN}(x_{\mathrm{s}}|\overline{\overline{x}}, \sigma_{\mathrm{dB}})$ for the shadowing induced variations.

Assumption 2.8

The overall average power gain $x_{\mathrm{l}} = \overline{\overline{x}}$ of the single-tap channel follows a power law attenuation

$$\overline{\overline{x}} \propto (d/d_0)^{-\kappa} \tag{2.25}$$

over distance d, where κ is the propagation loss exponent and d_0 is a reference distance above which the model is presumed valid. ∎

Finally let us introduce the following definition regarding the rate of change in the channel gain with respect to the transmission packet duration.

Definition 2.1 *Let T_{P} be the duration of a packet in seconds, and let T_{IP} be the (average) time between the beginning of each packet; the inter-arrival duration. The channel coherence time is T_{C}.*

1. *If $T_{\mathrm{C}} \to \infty$ the channel is* static

2. *If $T_{\mathrm{IP}} \ll T_{\mathrm{C}} < \infty$ the channel is* quasi-static

3. *If $T_{\mathrm{P}} < T_{\mathrm{C}} < T_{\mathrm{IP}}$ the channel is* slow *(block fading)*

4. *If $T_{\mathrm{C}} < T_{\mathrm{P}}$ the channel is* fast ∎

Communication performance in fading channels

Multi-path channels can have far reaching consequences for the communication performance and the design of communication systems. In Appendix 2.C we review some results relating performance – typically bit error or outage probability – to transmit power, which in turn is important for our analysis of energy efficiency. The reader who is unfamiliar with these matters in general, and Nakagami-m fading in particular, should consult the appendix.

1. The signal-to-noise ratio per bit γ is proportional to the channel gain x, see (2.18), as long as transmit power or interference levels do not change faster than the channel. Hence, the propagation loss, shadowing and fading models (2.23)-(2.25) can be applied to γ directly.

2. Regardless of which communication performance criterion one considers – bit error probability, outage probability or channel capacity –

it is crucially dependent on the received signal-to-noise ratio per bit γ. Hence the performance requirements in terms of the signal-to-noise ratio per bit γ translate into radiated energies $\mathcal{E}_{\mathrm{rad}}$.

3. Quite generally, the probability of error – probability of bit error or outage – will depend on the degree of Nakagami fading as

$$P_{\mathrm{error}} \propto \overline{\gamma}^{-m}, \tag{2.26}$$

where m is the fading figure and $\overline{\gamma}$ is the average signal-to-noise ratio per bit. The energy-wise impact of variations in m can be large, especially close to the Rayleigh case $m = 1$.

2.2.4 Summary of radio models

Our radio communication trade-off model is based on the fixed transmit and receive processing energies $\mathcal{E}_{\mathrm{Pt}}$ and $\mathcal{E}_{\mathrm{Pr}}$ respectively, and the variable transmission energy $\mathcal{E}_{\mathrm{T}}$. The only hardware specific model we use is the Mikami model (2.14) for the efficiency degradation of backed-off power amplifiers.

Our wireless channel model includes propagation loss, log-normal shadowing and frequency flat Nakagami-m fading. Communication performance is very sensitive to the parameter values in these models, and in order to quantify the present uncertainty we now turn to the topic of probability theory as logic.

2.3 Probability theory, uncertainty and decisions

Neither in science or in everyday life can we hope to reach decisions by deductive reasoning more than occasionally; the comfort of absolute certainty is denied to us most of the time. We are forced to reason from incomplete knowledge, uncertain as we are regarding the true state of matters. Without thinking about it, we carry out qualitative plausible reasoning every day. For example, the appearance of clouds makes rain more plausible and we choose to bring the umbrella. Good ability to process incomplete information efficiently has probably been favoured in the evolution of animals and humans – individuals lacking this ability would definitely have a disadvantage in the competition – and we are quite capable of making everyday decisions under uncertainty. But when we face complex problems in science or other professional fields our intuitive abilities are often insufficient. Formal deductive logic helps us to reach the correct conclusion even in immensely complicated chains of reasoning, and a quantitative theory of plausible reasoning would

be of great use. Fortunately, the mathematical framework of deductive logic, within which certain conclusions can be drawn from a set of premises, can be extended to probability theory, the logic of plausible reasoning.

In the present context of energy-efficiency in wireless sensor networks, there are several model parameters whose values are unknown and we should not ignore the present uncertainty when analysing our problems. Sensor networks must generally be designed under considerable uncertainty as to the actual operating conditions they will encounter, and the design requirements demand more than a qualitative plausibility analysis. We will find that probability in the general interpretation presented here provides us with the tools and principles we need, tools that are missing in conventional statistics. The principle of maximum entropy helps us to incorporate our incomplete knowledge in a defensible way by assignment of the most non-committal probability distribution consistent with our information. This use of prior probabilities is denounced in the conventional teachings but is crucial for us here. Moreover, the present theory tells us how to make optimal decisions based on the information we actually include, and also the uncertainty, not over an infinite set of imagined experiments, but for this particular case.

Our exposition of probability theory is to a large extent a condensation of the work of E. T. Jaynes, collected in his posthumously published book (Jaynes, 2003). Interested readers are strongly encouraged to read this book. Because there are many misconceptions regarding the interpretation of probabilities given here, we describe it at some length. The reader who is well acquainted with the theory before us can fast-forward to Section 2.4 on page 48.

2.3.1   Extending deductive logic

We will use the notation of Boolean algebra to handle propositions A, B, C, etc. and the deductive logic applicable to them.

$$
\begin{array}{rl}
AB & \text{The } \textit{logical product} \text{ is true if and only if} \\
   & \text{both } A \text{ and } B \text{ are true.} \\
A + B & \text{The } \textit{logical sum} \text{ is true if at least one of} \\
      & A \text{ and } B \text{ is true.} \\
\overline{A} & \text{The } \textit{negation} \text{ is true if and only if } A \text{ is} \\
             & \text{false.} \\
A \Rightarrow B & \textit{Logical implication} \text{ denotes the truth of} \\
                & \overline{A} + B, \text{ or, equivalently that } A\overline{B} \text{ is false.}
\end{array}
\tag{2.27}
$$

In order to be able to any draw deductive conclusions we must first assume the truth of some basic premise, for instance the truth of $A \Rightarrow B \equiv$ "if A is true, then B is true". Under this assumption[4], the information that B is false lets us deductively conclude that A is also false. However, deductive logic is silent on the issue if we instead learn that B is true; no deductive conclusion can then be drawn. But this is the most common situation. For instance, the medical doctor might know that disease A implies symptom B, but learning that B is true does not generally permit any deductive conclusion regarding A, because of alternative causes for B. On the other hand, it is possible to perform inductive reasoning based on the truth of B. It is precisely this type of plausible reasoning we wish to perform quantitatively by extending deductive logic.

The desiderata for quantitative plausible reasoning

The starting point for the extension of deductive logic to quantitative plausible reasoning is the statement of three fundamental desiderata of the theory to be developed.

I *Degrees of plausibility are represented by real numbers*

II *Qualitative correspondence with common sense*

III *Consistency*

   a) *If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result*

   b) *All relevant evidence is always taken into account. No information is arbitrarily ignored.*

   c) *Equivalent states of knowledge are always represented by equivalent plausibility assignments. That is, if in two problems the reasoner's state of knowledge is the same (except perhaps for labelling of the propositions), then he/she must assign the same plausibilities in both.*

A quantitative theory that fulfills the desiderata must have wide applicability due to the generality of the desiderata.[5] As it turns out, the only quantitative

---

[4]Care should be taken not to interpret *logical* implications as *cause-effect* relationships; $A \Rightarrow B$ does *not* mean that A is the physical cause of B.

[5]Observe that we have not mentioned "relative frequencies", "repetitive experiments", "Venn diagrams" or "betting" one single time.

theory that meets our desiderata is the well-known calculus of probability theory (Jaynes, 2003, Ch. 2).

The rules of probability theory

Let us by

$$P(A|BI) \tag{2.28}$$

denote the probability – that is the *quantitative* plausibility – that proposition A is true, *given* that proposition B and the basic premise I are true. Further, to conform with the standard for probability density functions for continuous variables, let us exercise a slight abuse of the Boolean notation and by

$$p(x|BI)|\mathrm{d}x| \tag{2.29}$$

denote the probability that the value of the continuous *parameter* x is within the range $[x, x+\mathrm{d}x)$; hence $p(x|BI)$ is the probability density function for x's value.[6] By the use of an upper case P we are thus referring to probabilities of discrete propositions – or hypotheses – while we use a lower case p when referring to a probability density functions for continuous parameters – or parameterised hypotheses.

Remark 2.4 *We have explicitly included the truth of the premise I, our background knowledge, in the notation $P(A|BI)$. We include it to stress that all probabilities are conditional on some background information I. The assumption that the premise is true is present also in deductive reasoning, and no deductive conclusion would be reached without it.*

We have stated the desiderata, essentially, in the words of Jaynes (2003), but it was R. T. Cox who provided the seminal derivation of probability theory from an essentially equivalent formulation (Cox, 1946). The refined analysis expounded by Jaynes shows that the product rule and the sum rule of probability theory constitute the only consistent rules for plausible reasoning.[7] As a consequence, anyone who does not adhere to the product

---

[6]If propositions happen to regard the value of a parameter X, we can define a parameterised range of propositions

$$H_x \equiv \text{"the value of } X \text{ is within } [x, x + \Delta x]\text{"} \tag{2.30}$$

and use the proper notation $P(H_x|BI)$. This notation could be defined with any desired accuracy Δx, but we will for simplicity use the more common $p(x|BI)$.

[7]To be precise, the scale is not fixed by the desiderata and any transformation $P' = P^w$, $0 < w < \infty$ fulfills the desiderata. However, it is the P, not the P', that are determined

rule

$$P(AB|I) = P(B|AI)P(A|I) = P(A|BI)P(B|I),$$
$$p(a|b, I) = p(b|a, I)p(a|I) = p(a|b, I)p(b|I),$$

(2.31)

and the sum rule

$$P(A|BI) + P(\overline{A}|BI) = 1,$$
$$\int p(a|b, I)\mathrm{d}a = 1,$$

(2.32)

will inevitably violate the desiderata $(I) - (III)$ in some way.[8]

From the basic rules one can derive the well-known theorem of Bayes,

$$P(H|d, I) = P(H|I)\frac{p(d|HI)}{p(d|I)},$$

(2.33)

were have used H for a hypothesis under consideration, and d for data used in the inference. Our knowledge is updated from the background I to the background *and* knowledge of the data d , and the probability is updated according to Bayes' theorem: it quantifies the learning from data. As a matter of convention one usually calls $P(H|I)$ the *prior* for H, $p(d|HI)$ the *likelihood* for H and $P(H|DI)$ the *posterior* for H. Additionally, we can from the sum rule derive the useful procedure of marginalisation,

$$P(H|d, I) = \sum_{m=1}^{M} P(H|B_m, d, I)P(B_m|d, I),$$
$$p(a|D, I) = \int p(a|b, D, I)p(b|D, I)\mathrm{d}b,$$

(2.34)

by which we can handle so-called *nuisance parameters* B_m and b; propositions/parameters that have influence on our inference but are not of primary interest.

At this point one might ask "So, we have rediscovered the rules of probability theory by extending deductive logic. Why the fuss?" Our short answer is:

---

by the data and the background knowledge, see (Jaynes, 2003, Sec. 2.4). The scale could also be fixed by introducing an extra desiderata on additivity: If three mutually exclusive propositions A_m are equally likely, the compound proposition $A_1 + A_2$ should be twice as plausible as A_3.

[8]Note that when including continuous parameters, we separate them by commas to avoid confusion; $p(a, b|I)$ instead of $p(ab|I)$ since the latter could be perceived as referring to the algebraic product ab.

The Cox-Jaynes formulation shows that probability theory as logic applies to all problems of quantitative reasoning from incomplete knowledge. Probability theory is *not* restricted to relative frequencies in repeated experiments or "physical probabilities". Neither are we required to arbitrarily classify entities as "random" or "deterministic", with allowance to apply probability theory to "random" entities only. Instead, the extended-logic formulation puts the focus where it should be in inference problems, on how to optimally process our incomplete information. With their strong logical foundation, the product and sum rules are applicable in a far more general context than the relative-frequency, random-deterministic, interpretation allows.

Presently, we need to handle nuisance parameters and draw conclusion based on "non-data" priors for them. Conventional methods lacks principles to cope with our situation, principles readily available once we accept probability theory as logic. Let us therefore proceed with the concept of entropy and later the principle of maximum entropy for assigning priors.

2.3.2   Uncertainty and entropy

Once we accept that a probability is a valid quantitative description of the plausibility that *the available information* warrants us to assign for A_m, it stands clear that a probability distribution $P(A_m|I)$ also must represent, quantitatively, our uncertainty regarding the truth of A_m. At first it might seem unnecessary to quantify our uncertainty with an explicit measure, but we will see later that a quantitative uncertainty measure can be useful both in the process of assigning priors and when defining the concept of measurement capacity.

Shannon's desiderata

C. E. Shannon contemplated the issue of uncertainty in the context of communication theory, and his attempt to find a quantitative measure was, indeed, successful (Shannon, 1948a). He started by considering M mutually exclusive and exhaustive propositions A_m; let us use the shorthand notation $P_m = P(A_m|I)$ for convenience. In his search for an uncertainty measure $H(P_1, P_2, \ldots, P_M)$ Shannon invoked the following desiderata:

1. **Continuity.** $H(P_1, P2, \ldots, P_M)$ is continuous in the P_m.

2. **Common-sense correspondence.** When all A_m are equally probable, $P_m = 1/M$, the uncertainty $H(1/M, 1/M, \ldots, 1/M)$ is monotonically increasing with the number of possible outcomes M.

3. **Additivity.** If proposition A_m can be broken down into mutually exclusive sub-propositions $A_{m,1}$ and $A_{m,2}$, with $P_{m,1} + P_{m,2} = P_m$, then

$$
\begin{aligned}
H(P_1, \ldots, P_{m,1}, P_{m,2}, \ldots, P_M) &= H(P_1, \ldots, P_m, \ldots, P_M) \\
&+ P_m H(P_{m,1}/P_m, P_{m,2}/P_m).
\end{aligned}
\tag{2.35}
$$

The resulting uncertainty should be the sum of the original uncertainty and the additional, weighted with probability P_m.

4. **Consistency.** If there are several ways to find $H(P_1, P_2, \ldots, P_M)$, all should lead to the same result.

Shannon found that the uncertainty measure meeting his desiderata must have the form

$$
H(P_1, P_2, \ldots, P_M) = -K \sum_{m=1}^{M} P_m \log(P_m),
\tag{2.36}
$$

where $K > 0$ is an arbitrary constant and the logarithm is taken to any base. For simplicity, we set $K = 1$. Note that

- The uncertainty $H(P_1, P_2, \ldots, P_M) \geq 0$

- For a given M, the uncertainty is maximal when all possibilities are equal, $P_m = 1/M$, and the uncertainty is then $H_{\max} = \log(M)$

- The uncertainty is zero if and only if one possibility must be true, $P_m = 1$, and all other false

The function $H(P_1, P_2, \ldots, P_m) \geq 0$ is called the *entropy* of the probability distribution $P(A_m|I)$.

Entropy and multiplicities

Although the Shannon desiderata seem reasonable and readily acceptable, do we really need to introduce them on top of our desiderata of plausible reasoning in order to assess uncertainty numerically? A more direct approach suggested by Graham Wallis – which also turn out to be very useful when we in Chapter 8 develop our notion of sensor network capacity as the

number of different event distributions the network can respond to – is the following (Jaynes, 2003, Sec. 11.4). Let each possibility A_m correspond to an imaginary box and the let content of box m be n_m indistinguishable quanta of probability that sum to $P(A_m|I)$; the distribution of quanta corresponds to our present probability distribution. Assume that there are $N = \sum n_m$ equal-sized quanta in total, where $n_m/N = P(A_m|I)$; we can choose N large enough to get any accuracy. Now, in how many ways could we rearrange the quanta while maintaining the distribution? The answer is given by the multinomial coefficient

$$W = \frac{N!}{n_1! n_2! \ldots n_M!},$$
(2.37)

see Sivia (1996) for more details. Hence, W is the number of different microstates – the number of rearrangements – that correspond to the given macrostate – the distribution $n_m/N = P(A_m|I)$. By taking the logarithm of (2.37) and applying the Stirling approximation,

$$\ln(n!) = n \ln(n) - n + \ln(\sqrt{2\pi n}) + O(n^{-1}),$$
(2.38)

we find after some straightforward manipulations that

$$\frac{1}{N} \ln(W) = -\sum_{m=1}^{M} \frac{n_m}{N} \ln\left(\frac{n_m}{N}\right) + O\left(\frac{\ln(\sqrt{N})}{N}\right).$$
(2.39)

The quantisation can be made arbitrarily fine by letting $N \to \infty$ and $n_m \to \infty$ in a way that maintains $n_m/N = P(A_m|I)$. In the limit

$$\frac{1}{N} \log(W) \to -\sum_{m=1}^{M} P_m \log\left(P_m\right),$$
(2.40)

and the influence of the P_m on the size of W is through the entropy (2.36) as derived by Shannon. The uncertainty is thus in the present formulation related to the number of microstates that correspond to the macroscopic constraint; for instance, when there is absolute certainty all quanta are in one box and there is only one allowed microstate, $W = 1$. This view will be useful later when we discuss entropy maximisation. Also, it clarifies the connection between information theory and thermodynamics, where the constraints can be, for example, pressure, volume and temperature while the microstates refer to the allowed positions and velocities of the molecules in a gas (Jaynes, 1957a,b).

Continuous variables

Taking the limiting case of a discrete problem formulation – for which the Shannon entropy (2.36) quantifies our uncertainty – in order to approach a continuous problem formulation we must be careful to specify exactly how the limit is approached. Jaynes (2003, p. 375) shows that the entropy expression approaches

$$H'(p(x|I)) \rightarrow - \int p(x|I) \log \left(\frac{p(x|I)}{Mq(x)} \right) \mathrm{d}x \qquad (2.41)$$

when the number of hypotheses $M \rightarrow \infty$. Subtracting off the infinite $\log(M)$ term he arrives at

$$H(p(x|I)) = - \int p(x|I) \log \left(\frac{p(x|I)}{q(x)} \right) \mathrm{d}x, \qquad (2.42)$$

where $q(x)$ is a function of the limiting process that ensures that the continuous entropy is invariant under a change of variables. Note that, due the subtraction of the $\log(M)$ term, $H(p(x|I)) \leq 0$ unlike the always nonnegative entropy $H(P_1, P2, \ldots, P_M)$ for uniform discrete hypothesis spaces.

Where do we stand?

At our disposal is now the quantitative tool for inference – probability theory as logic – and the quantitative description of our uncertainty – the entropy. But how do we get started in a real problem? The theory concerns so far only the manipulation of probabilities and the numerical assessment of uncertainty, but how do we assign the probabilities in the first place? We need tools for this in order to take model parameter uncertainties into account.

2.3.3   Assigning prior probabilities

According to the present theory in which a probability distribution represents incomplete information, the initial assignment of probabilities amounts to a "translation" of the specified prior information into a specific probability distribution. To ensure objectivity we invoke the interface Desiderata $(IIIb)$ and $(IIIc)$. Below we describe invariance (groups) and the principle of maximum entropy as the translational tools for certain types of prior information.

Ignorance priors and invariance arguments

In an information based quantitative theory of inference, the representation of complete ignorance is the natural point of departure – just like zero is a natural starting point for adding numbers. This limiting form of uninformative probability assignment can be achieved by exploiting any invariance that is implicitly defined by lack of constraints in the background information (Jaynes, 2003, Ch. 12).

Discrete hypothesis space. Consider a problem with M mutually exclusive and exhaustive possibilities A_m. The labels A_m are of course arbitrary, and the problem could be restated by some permutation (relabelling) $A_m \rightarrow A'_n$ of the propositions without changing the nature of the problem. Numbering the problems 1 and 2, the permutation $n = \text{Perm}(m)$ gives

$$P(A_m|I)_1 = P(A'_n|I)_2, \qquad \left\{ \begin{array}{rcl} n & = & \text{Perm}(m) \\ m & = & 1, 2, \ldots, M \end{array} \right. . \qquad (2.43)$$

Now, if the background knowledge I is *indifferent* between the propositions A_m – whatever it says about A_i it also says about A_j – the two problems are entirely equivalent. Desideratum ($IIIc$), which requires equal plausibility assignments from equal information, can then be restated as

$$P(A_m|I)_1 = P(A'_m|I)_2. \qquad (2.44)$$

Combining (2.43) and (2.44) we obtain

$$P(A_m|I)_1 = P(A_n|I)_1, \qquad \left\{ \begin{array}{rcl} n & = & \text{Perm}(m) \\ m & = & 1, 2, \ldots, M \end{array} \right. . \qquad (2.45)$$

If invariance under relabelling would not hold, we would always be at the mercy of a lucky choice of labels in the beginning of each inference problem; Desideratum ($IIIc$) protects us against this. Since the propositions are mutually exclusive and exhaustive, the only quantitative choice that does not violate the sum rule (2.32) is

$$P(A_m|I) = \frac{1}{M}. \qquad (2.46)$$

The assignment principle manifested in (2.46) rests on the information being indifferent between propositions, and is consequently referred to as the *principle of indifference*. An important observation is that the probability assignment in (2.46) is precisely the one that maximises the uncertainty given by the Shannon entropy in (2.36); if this was not the case some inconsistency would have been revealed.

Remark 2.5 *Care must be taken in application of the principle of indifference so that it is applied to a hypothesis space where the indifference really applies. Consider for example the transformation of problem 1 above, not through permutation of the A_m but through construction of a composite hypothesis $B_1 = A_1 + A_2$. Then we can no longer assert indifference. We give a more concrete example below, see Example 2.3.*

Continuous hypothesis space. In the discrete case, the uniform assignment (2.46) always holds under indifferent background knowledge. Most of us would be prepared to make this assignment without the line of reasoning above, but in the case of a continuous hypothesis space the problem is more difficult and the uniform assignment is not always the most uninformative one. To see this, consider the entropy (2.42) of a continuous variable x. It can be shown by the method of Lagrange multipliers that the probability distribution corresponding to maximum uncertainty $H(p(x|I))$ is $p(x|I) \propto q(x)$. If the continuous problem was found by a limiting process from a discrete problem $q(x)$ would have arisen in that process, but otherwise we must find it from an invariance property of the continuous problem. Jaynes (2003, Ch. 12) approaches the problem of finding $q(x)$ by the use of transformation groups and invariance. We will not cover the theory in detail but give instead an illustrative example.

EXAMPLE 2.2 The Uniform Prior and Jeffreys' Prior

Consider a signal x, $-\infty < x < \infty$, with unknown mean μ and variance σ^2. Reasonably, if we are completely ignorant about the mean value a shift according to

$$\mu' = \mu + a \tag{2.47}$$

can not make the problem appear different. Invariance under translation (2.47) requires

$$\begin{aligned} p(\mu|I)|\mathrm{d}\mu| &= p(\mu'|I)|\mathrm{d}\mu'| \\ &= p(\mu + a|I)|\mathrm{d}\mu|. \end{aligned} \tag{2.48}$$

The only solution to (2.48) is the uniform probability distribution

$$p(\mu|I) \propto \text{const.} \tag{2.49}$$

It seems that a *location* parameter like μ, for which we want translational invariance, should be assigned a uniform ignorance prior.

Consider now the variance σ^2. Certainly, translational invariance is not suitable – we can not translate the variance to negative values without absurd consequences – but scale invariance could be motivated by the fact the unit (nm or km, pW or MW, etc.) of the measurement should not matter if we are completely ignorant. Therefore, consider

$$\sigma' = b\sigma \tag{2.50}$$

and the invariance requirement

$$
\begin{aligned}
p(\sigma|I)|\mathrm{d}\sigma| &= p(\sigma'|I)|\mathrm{d}\sigma'| \\
&= bp(b\sigma|I)|\mathrm{d}\sigma|.
\end{aligned}
\tag{2.51}
$$

In this case, the only solution is the so-called Jeffreys' prior

$$p(\sigma|I) \propto \frac{1}{\sigma}. \tag{2.52}$$

Noteworthy is that Jeffreys' prior for σ corresponds to a uniform prior for $\log(\sigma)$, that is a uniform prior over the scale of the problem.

We have here illustrated the use of invariance to find the underlying ignorance measure to give an idea of how it can help us get started with an inference problem. More details are given by Jaynes (2003, Ch. 12), who, for instance, resolve Bertrand's famous paradox by the use of invariance arguments.

The principle of maximum entropy (MaxEnt)

As we noted above, the uninformative prior is a natural starting point, but we will not be completely ignorant in all real problems. There should be a way to incorporate cogent background information available at the outset; information that puts constraints on the probability assignment. We noted previously, see (2.46), that the most uninformative discrete probability distribution, the uniform, corresponded exactly to the maximum, unconstrained, Shannon entropy. Jaynes perception of this connection led him to formulate the principle of maximum entropy, MaxEnt for short, here in our choice of verbiage

> Assign the probability distribution which has the maximum entropy permitted by the constraints set by the background information.

The key idea is to include the imposed constraints, but nothing else, through the choice of the maximally non-committal probability assignment. No unwarranted assumptions are to be implicitly encoded into our assignment; no additional restrictions than the ones in I should affect it. This achieved by entropy maximisation as it ensures that the maximum possible range of "microstates" are included in our assignment (Jaynes, 2003).

The general solution. We here give the general solution to the entropy maximisation problem in a continuous setting; the discrete case is readily found by replacing integrals with sums. Let the constraints be expected values

$$\int f_k(x)p(x|I)\mathrm{d}x = F_k, \qquad k = 1, 2, \ldots, K. \tag{2.53}$$

The probability density function $p(x|I)$ which maximises the entropy

$$H(p(x|I)) = -\int p(x|I) \log\left(\frac{p(x|I)}{q(x)}\right) \mathrm{d}x, \tag{2.54}$$

subject to the constraints (2.53) is

$$p(x|I) = \frac{q(x)}{Z(\lambda_1, \lambda_1, \ldots, \lambda_K)} e^{-\lambda_1 f_1(x) - \lambda_2 f_2(x) - \cdots - \lambda_K f_K(x)}, \tag{2.55}$$

where the *partition function*

$$Z(\lambda_1, \lambda_1, \ldots, \lambda_K) = \int q(x) e^{-\lambda_1 f_1(x) - \lambda_2 f_2(x) - \cdots - \lambda_K f_K(x)} \mathrm{d}x, \tag{2.56}$$

and the Lagrange multipliers are found from

$$F_k = -\frac{\partial}{\partial \lambda_k} \log\left(Z(\lambda_1, \lambda_1, \ldots, \lambda_K)\right), \qquad k = 1, 2, \ldots, K. \tag{2.57}$$

Remark 2.6 *The constraints need not be expected values, but the information we include must be* testable*; we must be able to conclude whether a proposed distribution meets the constraint or not. For instance, the constraint $a < 5$ can be formulated as $\int_{-\infty}^{5} p(a|I)\mathrm{d}a = 1$ and we can readily check if a distribution violates this constraint or not.*

EXAMPLE 2.3 Poisson Distribution and Sensor Node Deployment

Assume that sensor nodes are deployed with a density of μ nodes/m^2.
What is the probability $p(n|\mu, a, I)$ to find *exactly* n nodes within an arbitrary region of area a? The average number of nodes within an area a
is μa, and incorporating this into the maximum entropy formalism as a
constraint (2.53) is straightforward,

$$\sum_{n=0}^{\infty} np(n|\mu, a, I) = \mu a = F_1. \tag{2.58}$$

Trickier is the question about the ignorance measure $q(n)$. If we without
further thought would apply the principle of indifference directly to the
nonnegative range of n, $q(n)$ would be uniform and the resulting Max-
Ent distribution would according to (2.55) be an exponential distribution,
$p(n|\mu, a, I) \propto \exp(-n/(\mu a))$. Clearly something is wrong when the most
probable value of n is zero, regardless of how densely we deploy nodes and
how large area we consider.

Consider instead the division of the area into $M \gg n$ conceptual, equal-
sized, boxes and the placement of a particular sensor node. Now, the
background information is indifferent to *which* of the M arbitrarily labelled
boxes one specific node will be placed. It is here, to the hypothesis space
of the position of each node, that the principle of indifference is applicable
in this particular problem. We can transform this space by considering
in how many ways n indistinguishable sensor nodes can be distributed
between M boxes. This reasoning, also expounded by Sivia (1996), yields
the ignorance measure

$$q(n) = \frac{M^n}{n!}, \tag{2.59}$$

the number of possible placements divided by number of sequences consist-
ing of n nodes. Application of the general MaxEnt solution then results in
the Poisson distribution

$$p(n|\mu, a, I) = \frac{(\mu a)^n e^{-\mu a}}{n!}. \tag{2.60}$$

Fortunately, the number M cancels out in the calculation.

Averages and true models. To us, the *principle* of maximum entropy appears perfectly sound, and is also a constructive approach. However, one may argue, reasonably, that our background information never includes the true average value F of a quantity $f(x)$ and that we therefore never can rightfully motivate the MaxEnt approach. Indeed, if we have a measurement sequence $\{x_n\}_{n=1}^N$ and choose to use only the sample average $\overline{x}$ in our probability assignment, we are clearly violating desideratum $(IIIb)$. On the other hand, if we are only given $\overline{x}$ the case for MaxEnt improves, but problems still remain: Should we really fix the mean value based on a small sample average? These objections have spurred criticisms and led some to reject the use of MaxEnt. Our view is that the status of constraints in MaxEnt is the same as the status of the (physical) models we incorporate to conduct inference; in both cases good judgement is required, but we would not get far if we denied ourselves the possibility to use anything but the "true model". For instance, we know that the Newtonian laws of motion are false, but still we use them successfully because they are good enough for everyday velocities. Whether we like it or not, our inferences – deductive or inductive – must be conditioned on the truth of some premise I. This holds also in the present context where we, for example, know that all our radio propagation models in Section 2.2.3 are merely approximations to the laws of physics.

EXAMPLE 2.4 Macroscopic Constraints and Fading Distributions

Consider the sampled received narrow-band low-pass signal

$$
\begin{aligned}
s_n &= a_n + jb_n \\
&= \varsigma_n e^{j\theta_n},
\end{aligned}
\tag{2.61}
$$

where a_n and b_n are the real and imaginary parts respectively – the in-phase I and quadrature Q components – and ς_n and θ_n are the corresponding polar coordinates; the envelope and the phase. Sample time is denoted n. Let us use translational invariance for a_n and b_n and therefore use $q(a_n) = q(b_n) = $ constant as our ignorance measures[9] in (2.54). Assume now that, in addition to our background knowledge that the signal is a narrow-band complex-valued signal, we also know that the mean power in each component is σ^2. We invoke this constraint in the maximum entropy

---

[9]One could here enter a long discussion about this choice, but for brevity we refer the interested reader to Jaynes (2003, Ch. 12).

procedure,

$$\iint a_n^2 p(a_n, b_n | I) \mathrm{d}b_n \mathrm{d}a_n = F_{1a} = \sigma^2,$$
$$\iint b_n^2 p(a_n, b_n | I) \mathrm{d}a_n \mathrm{d}b_n = F_{1b} = \sigma^2. \tag{2.62}$$

From (2.55) we can by the use of $f_{1a}(a_n) = a_n^2$ and $f_{1b}(b_n) = b_n^2$ conclude that the maximum entropy assignment is a two-dimensional, circularly symmetric, Gaussian probability distribution. By performing a change of variables from Cartesian (a_n, b_n) to polar (ς_n, θ_n) we obtain a Rayleigh distribution for the envelope ς_n and a uniform distribution for the phase θ_n. Hence, *the macroscopic constraints on mean power corresponds to assignment of a Rayleigh fading distribution.* For the received power, $x_n = \varsigma_n^2$ we obtain an exponential distribution $p(x_n | I)$, corresponding to a constraint

$$E(x_n^2 | I) = \int x_n^2 p(x_n | I) \mathrm{d}x_n = F_1. \tag{2.63}$$

Now, let us assume that we in a certain environment, for some reason, also know the average order of magnitude[10]

$$E(\ln(x_n) | I) = \int \ln(x_n) p(x_n | I) \mathrm{d}x_n = F_2, \tag{2.64}$$

the expected scale of the power. Together with (2.63) we can now use (2.56) to calculate the partition function Z, and then use (2.57) to solve for the Lagrangian multipliers λ_k. With the aid of Mathematica (Wolfram Research Inc., 2007) we find that

$$p(x_n | F_1, F_2, I) = \frac{m^m}{\Gamma(m) F_1^m} x_n^{m-1} e^{-m \frac{x_n}{F_1}}, \tag{2.65}$$

where the m parameter is found as the solution to

$$\ln(m) - \Psi(m) = \ln(F_1) - F_2. \tag{2.66}$$

Recognising in (2.65) the gamma distribution (2.23) in our Nakagami-m channel assumption, we see that this assumption corresponds to macroscopic constraints on $E(x_n | I)$ and $E(\ln(x_n) | I)$.

---

[10]The author has struggled some time to see whether this constraint can be well motivated, but has not reach a convincing argument.

2.3.4 Optimal decision making under uncertainty

The rules of probability theory prescribes how probabilities should be manipulated and concerns only the *inferential* part of the problem. The posterior $P(A_m|d, I)$, with possible nuisance parameters removed through marginalisation, summarises our knowledge about A_m, but how we should choose our course of action based on this knowledge is not solely determined by the posterior. We need to include, and value, the consequences of acting on correct and erroneous decisions. While the theory of probability rests on logical analysis from a set of nearly inescapable desiderata, decision theory naturally includes value judgements.

Hypothesis testing and parameter estimation

To make a decision A_{d}, and act as if it were correct, we must introduce value judgements that quantify the loss incurred for each combination of decision A_{d} and true state of nature A_m. This value judgement is contained in a so-called loss function $L(A_m, A_{\mathrm{d}})$. Combining the posterior and the loss function we can make the decision which minimises the expected loss,

$$\hat{A} = \arg \min_{A_{\mathrm{d}}} \sum_{m=1}^{M} P(A_m|D, I)L(A_m, A_{\mathrm{d}}). \qquad (2.67)$$

The estimation of a continuous parameter x is equivalently achieved by solving

$$\hat{x} = \arg \min_{x_{\mathrm{d}}} \int L(x, x_{\mathrm{d}})p(x|d, I)\mathrm{d}x. \qquad (2.68)$$

Observe the quite fundamental difference between the present approach, in which all decisions are based on the posterior probability distribution that summarises our knowledge, and the statistical approach, in which we must come up with an estimator without aid from underlying principles. The most common estimates are interpretable as the result of an application of a certain loss function. Let exemplify in the one dimensional case.

Quadratic-error loss. If we want to minimise the expected squared error $L(x, x_{\mathrm{d}}) = (x - x_{\mathrm{d}})^2$, the optimum estimate is the mean value of the posterior for x,

$$\hat{x}_2 = \int xp(x|d, I)\mathrm{d}x. \qquad (2.69)$$

Absolute-error loss. Considering instead the absolute value of the error $L(x, x_{\mathrm{d}}) = |x - x_{\mathrm{d}}|$, we obtain the median $\hat{x}_1$ as our optimum estimate.

It is found by solving

$$\int\limits_{-\infty}^{\hat{x}_1} p(x|d, I)\mathrm{d}x = \frac{1}{2}. \qquad (2.70)$$

Zero-one loss. When the size of the error is irrelevant – we are only satisfied with the true value of x – we can define a zero-one loss

$$L(x, x_{\mathrm{d}}) = \left\{ \begin{array}{ll} 0 & , \quad x_{\mathrm{d}} \equiv x \\ 1 & , \quad x_{\mathrm{d}} \neq x \end{array} \right. . \qquad (2.71)$$

Minimisation of the expected loss then yields the maximum a posteriori estimate; the x corresponding to the highest peak in the posterior $p(x|d, I)$. Note that if our prior $p(x|I)$ for x is uniform, the maximum a posteriori estimate will coincide with the common maximum likelihood estimate.

Minimax. If we are extremely conservative we may consider only the worst possible error and choose $\hat{x}$ so as to minimise that error. The only property of our posterior having influence is then its support; that is to say the set of all propositions that are not ruled out as impossible.

In this thesis we will mainly use the median together with percentiles x_a, defined by

$$a = \int\limits_{-\infty}^{x_a} p(x|I)\mathrm{d}x, \qquad (2.72)$$

when we present our energy efficiency results. Sometimes we complement with the mean value, but because our distributions are non-symmetric we will not display the commonly used standard deviations: sometimes it is outside the range of possible values and can thereby mislead.

2.4   Quantification of modelling uncertainty

In the present work the primary source of uncertainty is the wireless channel; there are extremely large variations in the channel gain over different environments, positions and times. We have in Section 2.2.3 described models for these variations[11], and the assumptions that went into the models,

---

[11]Details can be found in Appendix 2.B.

but our knowledge regarding their parameter values is incomplete. To avoid worst/best case design choices we should try to quantify the present uncertainties and take them into consideration. Unless we consider the uncertainties, we run the risk of making design choices that give excellent performance in a specific scenario but cause large energy losses in the majority of all possible scenarios.

When considering parameter uncertainty in our channel gain model and the sub-models for propagation loss, shadowing and fading – for details, see Appendix 2.B – a conceptual problem is that we know that the models are approximate and that the parameters therefore lack a really good physical interpretation. For this reason, an assignment of probabilities that obey the interface Desideratum (*IIIb*) becomes hard to find because reasoning from first principles is not possible. For example, what is really the physical origin of the propagation loss exponent κ in (2.100), and how should we reason to arrive at a reasonable prior for this empirical parameter? It is sometimes difficult to include much more than empirical averages, and we will use the principle of maximum entropy as our vehicle for this. We strive to include the prior information in the most sensible way, but we are pragmatic when needed and leave it to reader to make her own judgement about our choices. Inference problems are always open ended and we do not claim that we have the right answer, merely a sensible answer given by the information we include in the calculation.

2.4.1 Node position and network density

The positions of the sensor nodes have impact not only on the sensing performance, but also on the choice of transmission scheme. In particular when considering the energy-efficiency of multi-hop communication one must consider the relative position of the relay node: an offset from the optimal location incurs an energy-efficiency degradation. We here assign probability distributions for two scenarios, namely "random" sensor node deployment and non-perfect deployment with target positions.

Unknown or unstructured sensor node deployment

Assume that the position of the nodes are unknown, possibly because nodes were deployed in an uncontrolled fashion, for instance by an air-drop. In any case, we know only the (average) node density λ m^{-2} in the network. The result given in Example 2.3 is then applicable. For a given density λ m^{-2}, the probability for finding exactly k nodes in a region of area A is by the

principle of maximum entropy

$$P(k|A, \lambda, I) = \frac{(\lambda A)^k e^{-\lambda A}}{k!}, \tag{2.73}$$

the classical Poisson distribution. The expected number of nodes within an area A is $E(k|A, \lambda, I) = \lambda A$.

Theorem 2.1 *Assume that, for a two-dimensional network deployment, we have assigned the Poisson distribution (2.73) based on knowledge of the node density λ m^{-2}. Starting from an arbitrary position in the network, more than d m from the network boundary, the distance d to the closest node has a maximum-entropy probability density function*

$$p(d|\lambda, I) = 2\lambda \pi d e^{-\lambda \pi d^2}. \tag{2.74}$$

This Rayleigh distribution has mean value $(4\lambda)^{-1/2}$ and second moment $(\pi\lambda)^{-1}$.

Corollary 2.1 *Let d and ad, $a > 1$, be the semi-minor and semi-major axes of an elliptic area within the wireless sensor network. The shortest semi-minor axis d of such an ellipse that encloses at least one sensor node has a maximum-entropy probability distribution*

$$p(d|a, \lambda, I) = 2a\lambda \pi d e^{-a\lambda \pi d^2}. \tag{2.75}$$

This Rayleigh distribution has mean value $(4a\lambda)^{-1/2}$ and second moment $(a\pi\lambda)^{-1}$.

See Appendix 2.A for proofs of Theorem 2.1 and Corollary 2.1.

Sensor deployment with placement error

Let us assume, in contrast to uncontrolled deployments, that each node n instead has a target position which the placement procedure attempts to attain. We consider an idealised two-dimensional network described in the Cartesian coordinates (x, y), where the target position for node n is (x_n, y_n).

Theorem 2.2 *Assume that node placement is carried out without systematic error $\mu_x = \mu_y = 0$ m, but with a root-mean-square deviation $\sigma_x = \sigma_y = \sigma$ m. The distance $d = \sqrt{(x - x_n)^2 + (y - y_n)^2}$ from the target position has a maximum-entropy probability distribution*

$$p(d|\sigma, \mu = 0, I) = \frac{d}{\sigma^2} e^{-\frac{d^2}{2\sigma^2}}. \tag{2.76}$$

Proof: Entropy maximisation in the (x, y) plane yields a two-dimensional Gaussian probability distribution, (Jaynes, 2003). A subsequent change of variables from Cartesian coordinates (x, y) to polar coordinates (d, θ) results in (2.76). ∎

We see from Theorem 2.1 and Theorem 2.2 that a Rayleigh distribution for d is applicable for both "random" and structured deployments.

2.4.2 Propagation loss exponent

Without a cogent physical explanation of the origin of the propagation loss exponent κ it is hard to set down an entirely convincing line of reasoning regarding how to assign our prior $p(\kappa|I)$. Furthermore, it is also problematic to include numerical estimates $\hat{\kappa}$ from the literature. To process them properly via Bayes' theorem we need a model relating measurements and estimates – including the accuracy of the used instruments, the estimator used, the treatment of nuisance parameters and the length of the measurements – so that we can assign the likelihood $p(\hat{\kappa}|\kappa, I)$. This much information is rarely given in the published papers on measurements. For these reasons we take a pragmatic approach and exploit the MaxEnt tool to condense the most important aspects of the reported estimates into a probability assignment.

To begin with, we know for certain that the exponent κ in the power-law propagation loss model, see Assumption 2.8, must be greater than zero, $\kappa > 1$; otherwise we would have an amplifying channel. Empirically we furthermore know that exponents $\kappa > 6$ are extremely rare, while exponents in the range $[2, 4]$ are common, see Table 2.2 in Appendix 2.B. Our prior $p(\kappa|I)$ must therefore approach zero when $\kappa \to 0$ and when $\kappa > 6$; to achieve this we note that, according to the general MaxEnt solution given in (2.53)-(2.57), maximum-entropy average-constraints on

$$
\begin{aligned}
f_1(\kappa) &= x, \\
f_2(\kappa) &= \ln(x)
\end{aligned}
\tag{2.77}
$$

results in a distribution of the form $\kappa^{-\lambda_2} \exp\left(-\lambda_1 \kappa\right)$, see also Example 2.4. This gamma type of distribution can force probabilities to zero for small and large κ. The sample averages based on the estimates in Table 2.2 on page 62,

$$
\begin{aligned}
F_1 &= \frac{1}{N} \sum_{n=1}^{N} \hat{\kappa}_n = 3.405, \\
F_2 &= \frac{1}{N} \sum_{n=1}^{N} \ln(\hat{\kappa}_n) = 1.185,
\end{aligned}
\tag{2.78}
$$
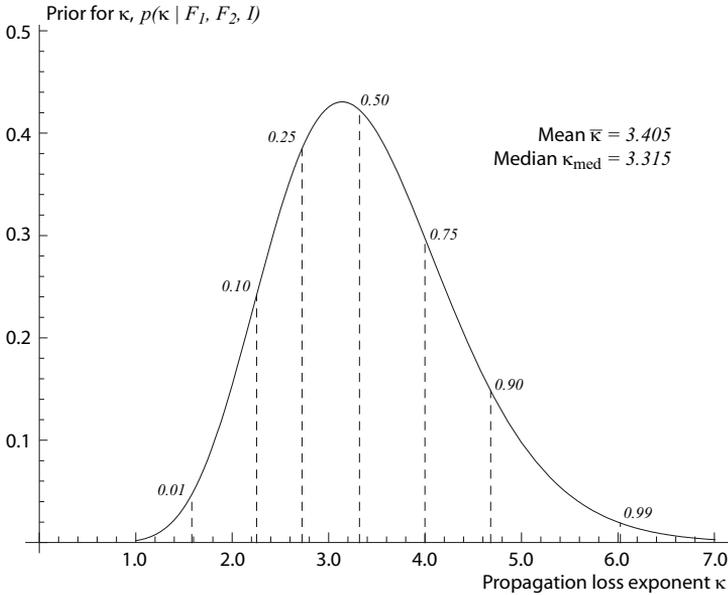
Figure 2.3: Maximum entropy prior for the propagation loss exponent based average constraints on κ and $\ln(\kappa)$ computed from the estimates collected in Table 2.2 in Appendix 2.B. Almost 90 percent of the probability lies in the range $[2, 5]$.

results in the MaxEnt assignment[12]

$$p(\kappa|F_1, F_2, I) = 82.31 \cdot 10^{-3} \kappa^{11.62} e^{-3.707\kappa}. \tag{2.79}$$

We plot this prior in Figure 2.3 together with percentiles.

Remark 2.7 *In the present approach, the choice of constraints is arbitrary: Why not also constrain the averages of κ^2 and $\exp(\kappa)$ while we are at it? The used set of estimates is small and inclusion of to many constraints would lead to over-fitting. The choice of constraints has to be done judicially, just like the choice of propagation loss model. While we admit that our approach is not perfect, we stress that the alternative mostly used in the literature is to simply ignore the uncertainty and pick one single exponent κ and proceed as if it was true. The possible consequences are evident upon comparing Example 1.1 and Example 1.3.*

---

[12]We have not specified the ignorance measure $q(\kappa)$ in (2.55), but it turns out that the constraints in (2.77) have the peculiar property of making the choice between the two most common ignorance measures, $q(\kappa) = $ const. and $q(\kappa) \propto 1/\kappa$, pointless. Regardless of whether we deem translational or scale invariance most appropriate, the result is still the gamma distribution in (2.79).

The probability density function in Figure 2.3 is a reasonable description of the present author's general knowledge about κ, but if we would have considered a more specific type of environment $p(\kappa|F_1, F_2, I)$ would be narrower. For example, in a forest we are not likely encountering $\kappa < 2$ since this implies some kind of wave-guide effect (typically encountered in corridors). As always, the inference problem is open-ended and our prior (2.79) is not *the one and only* prior.

2.4.3 Amount of large-scale shadowing

By taking the same approach here as for the propagation loss exponent, we make use of Table 2.3 in Appendix 2.B and calculate the averages

$$F_1 = \frac{1}{N} \sum_{n=1}^{N} \hat{\sigma}_{\mathrm{dB},n} = 4.48,$$

$$F_2 = \frac{1}{N} \sum_{n=1}^{N} \ln(\hat{\sigma}_{\mathrm{dB},n}) = 1.28,$$

$$(2.80)$$

where σ_{dB} is the decibel standard deviation in the log-normal shadowing model, see (2.101) in Appendix 2.B. The resulting prior for σ_{dB} is

$$p(\sigma_{\mathrm{dB}}|F_1, F_2, I) = 0.180\sigma_{\mathrm{dB}}^{1.40}e^{-0.534\sigma_{\mathrm{dB}}}, \qquad (2.81)$$

and we plot this prior in Figure 2.4.

2.4.4 Degree of small-scale fading

We saw in Example 2.4 that the Nakagami-m *envelope* fading distribution is the maximum entropy solution under the constraints of normalised channel *power* gain $\overline{x} = E(x|I) = 1$, and an average scale of the gain $E(\ln(x)|I) = F_{\mathrm{nak}}$. The gamma distribution for x is

$$p(x|\overline{x}, F_{\mathrm{nak}}, I) = \frac{m^m}{\Gamma(m)}x^{m-1}e^{-mx}, \qquad (2.82)$$

where m is the Nakagami-m fading figure. The role that the fading distribution has in our framework limits the fading figure to $m \geq 1$; that is, we consider Rayleigh fading the worst case.[13] In the entropy maximisation the

---

[13]If the constraint F_{nak} is not included, then the result is the exponential distribution, that is to say Rayleigh fading. Since this represents the worst case of several reflections of approximately equal magnitude (see Appendix 2.B), a fading distribution worse than this implies that some shadowing occurs. We have however modelled shadowing explicitly with the log-normal model and wish to keep their roles separated.
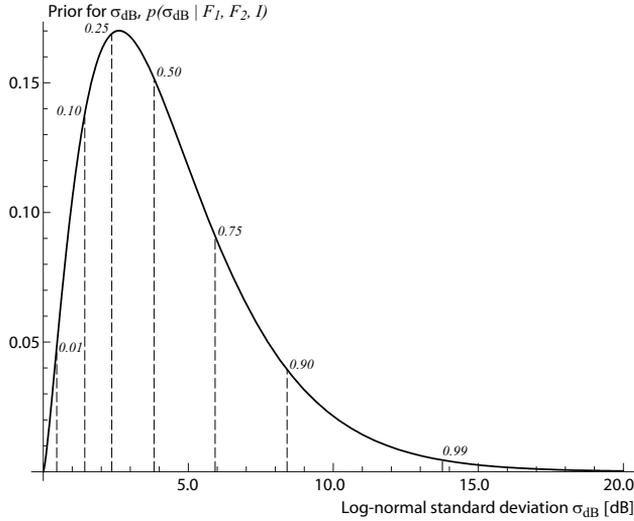
Figure 2.4: *Probability assignment for σ_{dB}, the decibel standard deviation in the log-normal shadowing model, based on Table 2.3 in Appendix 2.B.*

fading figure is determined from

$$F_{\mathrm{nak}} = \Psi(m) - \ln(m), \tag{2.83}$$

see Example 2.4, where Ψ denotes the polygamma function. Now, it is hard to reason about the physical meaning of m, but the situation is slightly better for F_{nak} which essentially constrains the scale of the variations due to fading. If the scale of the variations is completely unknown, then a uniform prior

$$p(F_{\mathrm{nak}}|I) = \mathrm{const.} \tag{2.84}$$

would reflect our ignorance. A change of variables according to (2.83) leads to

$$p(m|I) = \begin{cases} \frac{1}{\gamma_{\mathrm{Euler}}} \left(\Psi(1,m) - \frac{1}{m} \right) & , \quad m \geq 1 \\ 0 & , \quad m < 1 \end{cases} \tag{2.85}$$

where $\Psi(1,m)$ is the first derivative of the polygamma function and γ_{Euler} is Euler's constant. The prior in (2.85) is shown in Figure 2.5. We observe that $p(m|I) \propto 1/m^2$ is a fairly accurate approximation.

We do not include additional constraints, and this is mainly because we do not have enough quantitative information to motivate the same procedure as for the propagation loss exponent κ and the shadowing variance σ_{dB}.

Remark 2.8 *It might seem that our prior (2.85) excludes the possibility of a static channel, $m \to \infty$, and puts all the weight on severely fading*
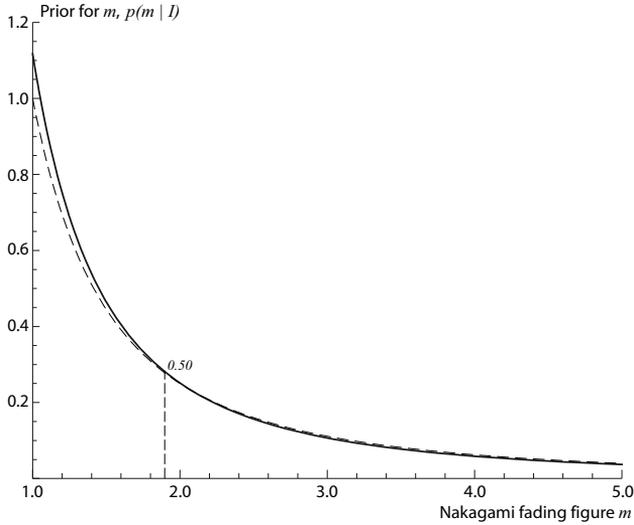
Figure 2.5: Probability assignment for the Nakagami fading figure m (black solid line). Also shown is a $1/m^2$ prior (dashed line) which closely approximates the assignment in (2.85).

channels. However, for practical purposes all channels for which $m > 10$ can be considered static since the communication performance is then quite close to the ideal static channel, see Appendix 2.C. In that sense, our prior says that there is a chance of about 10 percent to encounter a static channel.

Appendix 2.A   Proof of Theorem 2.1

Proof: Consider an in-network region R of area A, and let k be the number of nodes within R. The probability that it will contain at least one node is by the sum rule of probability theory, and the Poisson assignment in (2.73),

$$
\begin{aligned}
P(\text{“Area } A \text{ encloses at least one node”}|\lambda, I) &= 1 - P(k = 0|A, \lambda, I) \\
&= 1 - e^{-\lambda A}. \qquad (2.86)
\end{aligned}
$$

We can thus write, with a slight abuse of notation,

$$
P(A_1|\lambda, I) = 1 - e^{-\lambda A_1}, \qquad (2.87)
$$

where A_1 is the "parameterised proposition" that an area A_1 is not empty. By differentiation we obtain the probability density function

$$
p(A_1|\lambda, I) = \lambda e^{-\lambda A_1}. \qquad (2.88)
$$

Put differently: the area A_1 for which the region R becomes non-empty is assigned an exponential probability distribution with mean λ^{-1}.

Specialising the region to a circle of radius r, we carry out a change of variable according to $A_1 = \pi r^2$, $|\mathrm{d}A_1| = |2\pi r \mathrm{d}r|$, in (2.88) and arrive at

$$
p(r|\lambda, I) = 2\pi\lambda r e^{-\lambda \pi r^2}, \qquad (2.89)
$$

which is (2.74). Likewise, considering an ellipse of area $A_1 = \pi a r^2$, the change of variable leads to (2.75). ∎

Appendix 2.B   Linear-filter channel model

By modelling the low-pass channel between the transmitter and receiver as consisting of different paths, each with a corresponding delay $\tau_n(t)$ and gain $a_n(t)$, we can readily express the channel's properties as the (time-variant) impulse response of a linear filter

$$
h(\tau, t) = \sum_{n=1}^{N(t)} a_n(t) e^{-j2\pi f_c \tau_n(t)} \delta\left(\tau - \tau_n(t)\right). \qquad (2.90)
$$

Here, $a_n(t)$ is the amplitude gain of the nth path at time t, and the delay corresponding to this path is denoted $\tau_n(t)$ while $N(t)$ is the number of paths between transmitter and receiver. The carrier frequency is denoted f_c and we use $\delta(\cdot)$ for the Dirac delta function. This type of channel is naturally called a *multi-path* channel, and is essentially a plane wave approximation of the full solution of Maxwell's equations.

Appendix 2.B.1   Delay spread and frequency-selectivity

The length of the channel's impulse response is commonly termed delay spread, $T_h(t) \equiv \max(\tau_n(t))$. The delay spread is practically always non-zero and the channel will cause inter symbol interference. The interesting question is whether this inter symbol interference is negligible or not. When $T_h(t)$ is considerably smaller than the symbol duration T there will not be any significant interference between the transmitted symbols at the receiving antenna. We can then simplify the model to a single-tap filter that models all path contributions collectively.[14]   We thus obtain

$$h_0(\tau, t) \equiv a_0(t)\delta\left(\tau - \tau_0\right), \tag{2.91}$$

where

$$a_0(t) \equiv \left(\sum_{n=1}^{N(t)} a_n(t)e^{-j2\pi f_c \tau_n(t)}\right). \tag{2.92}$$

Over the bandwidth of the transmitted signal the single-tap channel will have a constant (flat) frequency response and is therefore called frequency non-selective, or frequency flat. So, what delay spreads are to be expected for the data gathering sensor networks we are considering?

Frequency Non-Selective Sensor Network Channels   Sensor nodes will typically use low power radios and communicate over relatively short distances. Nodes placed indoors or close on the ground outdoors are more likely to encounter small delay spreads than large (a large delay spread requires, in addition to the direct path, open paths via far away objects). A short survey of the literature on channel measurements reveals that root-mean-square (RMS) delay spreads[15] in relevant scenarios typically range from 10 ns to 100 ns, with means around 40 ns. Worst case root-mean-square delays of around 300 ns are reported but seem very

---

[14]Most easily this is illustrated by an infinite symbol duration for which all delay spreads are small. The transmitted low-pass signal in this extreme case is $s_T(t) = A\exp(-j\phi)$, where ϕ is the carrier phase and A is the carrier amplitude. The received low-pass signal $s_R(t)$ is by the use of (2.90)

$$s_R(t) = \sum_{n=1}^{N(t)} a_n(t)e^{-j2\pi f_c \tau_n(t)} s_T(t - \tau_n(t)) = \left(\sum_{n=1}^{N(t)} a_n(t)e^{-j2\pi f_c \tau_n(t)}\right) Ae^{-j\phi}.$$

The receiver can not resolve the different paths – it sees only a pure carrier – and the channel model is effectively reduced to the single (complex-valued) tap within the brackets.

[15]The maximum delay spread $T_h(t)\max(\tau_n(t))$ is quite impractical when characterising a given type of transmission environment. By assuming that different paths have uncorrelated behaviour and that the *average* behaviour of the channel does not depend on t one can define the multi-path intensity profile (or power delay profile) $q(\tau) \equiv E\left(|h(\tau, t)|^2\right)$, where $E(\cdot)$ denotes expectation (average). The profile $q(\tau)$ gives the average power gain corresponding to a certain delay. It is common to use the root-mean-square (RMS) delay

rare. In Table 2.1 we summarise the references we have consulted. We have deliber-
ately disregarded from non-representative scenarios with respect to wireless sensor
networks (for instance long-distance rooftop-to-rooftop transmissions). As a conse-
quence, our list is by no means exhaustive and one can surely find measurements
showing more extreme values.

   We adopt a reasonably conservative definition of "flat" by considering a channel
with $T_{\mathrm{RMS}} < 10T$ as flat (Goldsmith, 2005, Sec. 3.3.1). Our conclusion based
on Table 2.1 is then that symbol durations down to $T = 10 \cdot 40$ ns $= 400$ ns
(corresponding to a symbol rate of 2.5 MS/s) will in most cases lead to negligible
inter symbol interference, while symbol durations down to 1 μs (symbol rate 1 MS/s)
almost surely will be on the safe side. As we consider 1 MS/s a high rate in the
sensor network context, we will henceforth use Assumption 2.5 on page 28, here
restated,

> The multi-path channels between the wireless sensor nodes are fre-
> quency non-selective (flat)with a single resolvable multi-path compo-
> nent modelled by a single complex-valued filter tap.[16]

Resting on this assumption and (2.91) we make the following definitions for the
frequency non-selective channel.

Definition 2.2 *Given a single-tap model* $h_0(t, \tau) = a_0(t)\delta(\tau - \tau_0)$ *the channel's
(time varying) envelope gain is*

$$r(t) \equiv |a_0(t)|, \tag{2.93}$$

and its power gain *is*

$$x(t) \equiv |a_0(t)|^2 = r^2(t). \tag{2.94}$$

■

---

spread without reference to the shape of $q(\tau)$:

$$T_{\mathrm{RMS}} \equiv \left(\int\limits_0^\infty \left(\tau - \overline{T}_h \right)^2 \breve{q}(\tau)\mathrm{d}\tau \right)^{1/2},$$

where $\breve{q}(\tau) \equiv q(\tau)/\int q(\tau)\mathrm{d}\tau$ is the normalised intensity profile, and $\overline{T}_h$ is the average
delay spread.

[16]It is of course conceivable that networks must communicate under frequency-selective
circumstances and our analysis in this part of the thesis would in that case have to be
modified. We do not believe that such a generalisation would alter the results regarding
energy-efficiency to a noticeable extent. The most prominent difference would be the
extra energy consumption of a channel equaliser. On the other hand, the energy effects
of frequency selectivity should be taken into account if the comparison happens to be
between narrow-band and wide-band alternatives since the latter would have to cope with
frequency selectivity through an energy consuming equaliser.

Table 2.1: *Short survey of root-mean-square (RMS) delay spreads T_{RMS} reported in the literature. Average values are typically about 40 ns, but averages up to 100 ns are found in some cases.*

Reference	Frequency	Environment	RMS delay spread T_{RMS}		
			Mean	St. dev.	Worst
	GHz		ns	ns	ns
Bohdanowicz et al. (1999)	17	Corridor	36	16	
		Canteen	36	12	
		Office	16	4	
		Parking	53	12	
Zhao et al. (2002)	5.3	Urban	44		
		Suburban	66		
		Rural	22		
Win et al. (1997)	(UWB)	Forest	38	7	
Ghassemzadeh et al. (2004)	5.0	Indoor (residential)	<12		
Poon and Ho (2003)	2.0–8.0	Office	20-30		
Bultitude et al. (1989)	0.9	Office floor	26	8	
	1.7	Office floor	28	17	
Saleh and Valenzuela (1987)	1.5	Office floor	25		50
Rappaport (1989)	1.3	Factory	100		300
Janssen et al. (1996)	2.4	Indoor lab/office	1–17		
	4.75		1–5		
	11.5		1-10		

Appendix 2.B.2   Propagation loss, shadowing and fading

We have now, by virtue of Assumption 2.5, reduced our linear multi-path model to a single complex-valued tap and the remaining question is how to model its behaviour. In the following we conform to the standard approach in which the channel gain, modelled by the filter tap, is studied on three different spatial scales: overall average attenuation over distance (propagation loss), large scale attenuation (shadowing), and small-scale attenuation (fading). Albeit on different scales, the three superimposed models are all on the macroscopic level since they avoid modelling of the fine details in the environment and rather collect the average behaviour into a

small number of model parameters.

More specifically, we decompose the channel's *power gain x* into three factors,

$$x = x_\mathrm{l} x_\mathrm{s} x_\mathrm{f}, \tag{2.95}$$

where the right hand side factors are ordered according to scale of modelling and represent propagation loss (l), shadowing (s) and fading (f) respectively (we have for simplicity dropped the explicit time-dependence notation (t)). We will use the following normalisation of x_s and x_f.

Definition 2.3 *Let x_s and x_f denote the the contributions from shadowing and fading to the channel gain in (2.95). Their* average *power gains are normalised to unity,*

$$\begin{aligned} \overline{x}_\mathrm{s} &= E(x_\mathrm{s}) \equiv 1, \\ \overline{x}_\mathrm{f} &= E(x_\mathrm{f}) \equiv 1. \end{aligned} \tag{2.96}$$

■

From a small scale perspective x_l and x_s are approximately constant and the changes in x follow the changes in x_f. Moving up one step, we can average over the small scale changes to obtain

$$\overline{x} = x_\mathrm{l} x_\mathrm{s}, \tag{2.97}$$

which over the intermediate (shadow) scale follows x_s since x_l is almost constant. Finally, on the overall scale we have

$$\overline{\overline{x}} = x_\mathrm{l}. \tag{2.98}$$

The division into three different scales is of course not clear-cut and there generally tends to be some overlap between the effects from each.

Distance dependent propagation loss

If transmission takes place in free space by the use of an isotropically radiating antenna the channel is, for any fixed transmitter and receiver positions, truly described by one complex-valued tap with a time-independent power gain $x(t) \equiv x$. This gain will however be distance dependent: The transmitted power will be spread over a sphere of area $4\pi d^2$ resulting in a free-space propagation gain

$$x_\mathrm{free-space}(d) \propto d^{-2}, \tag{2.99}$$

where d is the distance separating transmitter and receiver. For practical transmission environments there is no corresponding general formula, but empirical evidence and the need for simple models has led researches to generalise the free-space formula to the average power law

$$x_\mathrm{l}(d) \propto \left(\frac{d}{d_0} \right)^{-\kappa} \tag{2.100}$$

which refers to the channel's average power gain, and where d_0 is a reference distance depending on the environment.[17] For some special cases one can motivate a particular value of the propagation loss exponent κ. For instance, by the use of a two-path model one can show that the propagation loss exponent under some conditions converges to $\kappa = 4$ for large d (Parsons, 2001, Sec. 2.3.3). Generally, however, the power-law model is a fairly crude empirical model used for overall trade-off analyses, and there are several other proposed models that are more specific and accurate (Parsons, 2001, Sec. 3.6). Specialisation of the propagation loss model to specific scenarios leads to a loss of generality, and we choose to use the power law model in (2.100), see Assumption 2.8,

> The overall average power gain $x_1 = \overline{\overline{x}}$ of the single-tap channel follows a power law attenuation $\overline{\overline{x}} \propto (d/d_0)^{-\kappa}$ over distance d, where κ is the propagation loss exponent.

Remark 2.9 *The more accurate a model is the better,* provided *that we possess reasonable knowledge of the model parameters. If we were to replace (2.100) with a more accurate model we would have to cope with more parameters. In this case our judgement is that additional parameters would not improve our general trade-off analysis enough to motivate their use, but we acknowledge that (2.100) is not very accurate in all scenarios.*

In Table 2.2 we give representative values for a few different environments. It seems that propagation loss exponents κ typically reside in the range $[2, 4]$, but values slightly below 2 and in the range $[4, 5]$ are not too uncommon. Values above 5 and below 1.5 are however rare.

Large scale shadowing

The propagation loss model in (2.100) accounts only for the average loss over distance. No real environment is smooth, or isotropic, enough to be well described by this model only, so the next step is to include large obstacles (with respect to the wavelength of radio carrier) that blocks or attenuates the signal. As such, the model should capture variations around the prediction given by (2.100). One may reason as follows (Goldsmith, 2005, p. 50). A blocking object of thickness d_t attenuates the signal power by a factor $\exp(-cd_t)$ where c is a material-dependent attenuation constant. Several consecutive blocking objects of different thickness, but identical attenuator constants, give an attenuation $\exp\left(-c \sum d_t\right)$. Assuming a large number of objects we may invoke the central limit-theorem and assign a Gaussian distribution for $\sum d_t$. Consequently, when a change of variables according to $x_s \propto \exp(-c \sum d_t)$ is carried out, the channel power gain $x \propto x_s$ is assigned a log-normal distribution. While the theoretical motivation just given is not fully

---

[17]The average is over all positions separated by a distance d, and over time if the environment is changing; it is assumed that the distance does not change as this would obviously introduce a time dependence.

Table 2.2: Measured propagation loss exponents in published papers based on the empirical model (2.100). A selection has been made based on relevance to the topic of wireless sensor networks (for instance, no highly elevated, mast mounted antennas). Results are given for a range of data sets, see the Range column, and sometimes for a the total of all data sets, see the Overall column. Consult the references for more details.

| Reference | Frequency | Environment | Loss exponent κ | |
	GHz		Range	Overall
Durgin et al. (1998)	5.85	Indoor (residential)	3.3–3.5	3.4
		Outdoor (garden)	2.9–3.0	2.9
Seidel and Rappaport (1992)	0.91	Grocery store	1.8	1.8
		Retail store	2.2	2.2
		Office 1	2.7–4.0	3.5
		Office 2	3.3–(-)	4.3
		All locations		3.1
Sohrabi et al. (1999)	0.8–1.0	Office building	1.4–2.2	1.9
		Hallway	1.9–2.2	2.0
		Parking structure	2.7–3.4	3.0
		Corridor	1.4–2.4	1.9
		Patio	2.8–3.8	3.2
		Concrete canyon	2.1–3.0	2.7
		Plant fence	4.6–5.1	4.9
		Small boulders	3.3–3.7	3.5
		Sandy beach	3.8–4.6	4.2
		Dense bamboo	4.5–5.4	5.0
		Tall under-bush	3.0–3.9	3.6
Janssen et al. (1996)	2.4	Indoor lab/office	1.9–3.3	
	4.75		2.0–3.8	
	11.5		1.9–4.5	
Di Renzo et al.	(UWB)	Forest	2.5–2.7	
		Suburban	2.6–3.5	
		Hilly	2.2–2.3	

satisfactory – it seems a little too specialised – there is good empirical support for the log-normal assignment, see for example the work by Salo et al. (2007) and the references therein.

By convention, the severity of the log-normal shadowing is given by the standard deviation σ_{dB} of the Gaussian distribution in the dB domain.[18] Based on this macroscopic modelling parameter, our probability assignment for the channel gain $\overline{x}$ (averaged over the small scale effects in x_{f}) is

$$p(\overline{x}|\overline{\overline{x}}, \sigma_{\mathrm{dB}}) = \frac{1}{\overline{x}} \cdot \frac{\mu}{\sqrt{2\pi}\sigma_{\mathrm{dB}}} e^{-\frac{1}{2\sigma_{\mathrm{dB}}^2}\left(\mu \ln\left(\overline{\overline{x}}/\overline{x}\right) - \frac{\sigma_{\mathrm{dB}}^2}{2\mu}\right)^2} \qquad (2.101)$$

where $\mu \equiv 10/\ln(10)$ and $\overline{\overline{x}} = x_{\mathrm{l}}$ is the overall average. So, given the overall average $\overline{\overline{x}}$ from the propagation loss model (2.100) and the environment characteristic σ_{dB} we summarise our uncertainty with the probability assignment in (2.101). As shorthand we will use $\mathrm{LogN}(\overline{\overline{x}}, \sigma_{\mathrm{dB}})$ to denote the log-normal distribution with mean $\overline{\overline{x}}$ and dB-variance σ_{dB}^2. Consequently, through a change of variables we assign a $\mathrm{LogN}(x_{\mathrm{s}}|1, \sigma_{\mathrm{dB}})$ distribution for x_{s}.

The log-normal *sample standard deviation* is the value quoted in analyses of channel measurement data. In Table 2.3 we give a selected collection published results which we believe are relevant in our context. Although the log-normal model enjoys good empirical support and is widely used, other models have been proposed. Abdi and Kaveh (1999), for instance, do argue that the Gamma model in many cases works just as well with the additional advantage of computational tractability. Fully aware of other proposed alternatives we will use a log-normal assignment as stated in Assumption 2.7 on page 29;

> The (large scale) shadowing effects are well characterised by their standard deviation σ_{dB}. We use a log-normal assignment $p(x_{\mathrm{s}}|\overline{x}_{\mathrm{s}}, \sigma_{\mathrm{dB}}, I) = \mathrm{LogN}(x_{\mathrm{s}}|\overline{x}_{\mathrm{s}}, \sigma_{\mathrm{dB}})$ for x_{s}, with the normalised average $\overline{x}_{\mathrm{s}} = 1$.

Small scale fading

We note from (2.92) that the filter tap in our linear channel model is a sum of complex-valued multi-path contributions. So far we have accounted for average attenuation through the propagation loss model and the large-scale shadowing effects through the log-normal assignment, but from (2.92) it is clear that small-scale changes to individual path delays $\tau_n(t)$ can influence the channel power gain $x(t)$ even if the individual paths are not resolvable. Constructive/destructive combining of paths depend on the delays, and the common term for these multi-path induced variations is *fading*.

When the number of multi-path components $N(t)$ is large we can invoke the central limit-theorem for the sum of complex-valued path contributions and assign a two-dimensional zero-mean Gaussian distribution for the filter tap $a_0(t)$. As

---

[18]Observe that σ_{dB}^2 is *not* the variance of the *log-normal* distribution, but the variance of the *Gaussian* distribution.

Table 2.3: *Measured sample standard deviations σ_{dB} for fluctuations around the local mean. As far as we know, all included studies have averaged out small scale fading, but it is not always clear were the line between the large and the small scales are drawn.*

Reference	Frequency	Environment	Standard deviation	
	GHz		Range dB	Mean dB
Durgin et al. (1998)	5.85	Indoor (residential)	7.3–8.3	8.0
		Outdoor (garden)	6.4–9.0	7.9
Seidel and Rappaport (1992)	0.91	Grocery store	5.2	5.2
		Retail store	8.7	8.7
		Office 1	4.3–11	13
		Office 2	5.2–(-)	13
		All locations		16
Sohrabi et al. (1999)	0.8–1.0	Office building	2.4–3.6	2.4
		Hallway	1.7–3.3	2.8
		Parking structure	1.5–4.1	2.8
		Corridor	2.0–4.0	2.8
		Patio	1.0–3.0	1.9
		Concrete canyon	2.2–4.5	3.2
		Plant fence	1.7–2.2	3.1
		Small boulders	3.0–4.3	3.6
		Sandy beach	1.5–3.2	2.0
		Dense bamboo	0.6–6.8	3.4
		Tall under-bush	2.0–4.0	2.9
Healey et al. (2000)	2.4	Residential campus		9

consequences we assign a Rayleigh distribution for the amplitude gain ς and an exponential distribution for the power gain x. Undoubtedly, the Rayleigh distribution is the most common small-scale fading distribution assignment found in the literature. But, the use of the central limit theorem relies on equal (or nearly equal) path gains $a_n(t)$, and knowledge about deviations from this assumption has been incorporated by both S. O. Rice and M. Nakagami (Hashemi, 1993).

The more general approach of the two was taken by Nakagami (1960) who made allowance for arbitrary path gains $a_n(t)$. His assignment, based on an approximate solution to his original sum-of-random-vectors problem, is now called the Nakagami-

m distribution and is for the envelope gain

$$p(\varsigma|m,\overline{x}) = \frac{m^m}{\overline{x}^m\Gamma(m)}\varsigma^{2m-1}e^{-\frac{m}{\overline{x}}\varsigma^2}. \tag{2.102}$$

Here $m \in (0,\infty)$ is called the fading figure and $\overline{x} = E(\varsigma^2) = x_l x_s$ is the average power gain.[19] The corresponding distribution for x is the gamma distribution

$$p(x|m,\overline{x}) = \frac{m^m}{\overline{x}^m\Gamma(m)}x^{m-1}e^{-\frac{m}{\overline{x}}x}. \tag{2.103}$$

The Nakagami-m (gamma) distribution includes the Rayleigh (exponential) distribution as a special case when $m = 1$.[20] We can also represent the non-fading channel by letting $m \to \infty$. The Nakagami-m assignment use two macroscopic parameters, m and $\overline{x}$, where the former captures the severeness of the fading.

We will in this thesis make extensive use of Assumption 2.6 on page 29;

> The (small-scale) fading is well characterised by the Nakagami-m fading figure m. We use a gamma distribution $p(x_f|m,\overline{x}_f,I) = \mathrm{Gam}(x_f|m,\overline{x}_f)$ for x_f, with the normalised average $\overline{x}_f = 1$.

Our motivation of the Nakagami assumption is fourfold.

- The Rayleigh distribution is in many cases inadequate; it can often not characterise the channel's small scale behaviour well enough. Several measurements campaigns – including our own, see Chapter 3 – have lead to this conclusion, which is also theoretically supported by the fact that the Rayleigh distribution cannot properly account for different path gains. Hashemi et al. (1994) made an extensive measurement campaign at 1.1 GHz in an office environment and tested different distributions, and they found that "[T]he Weibull and Nakagami distributions provide the best fit for most cases. Rayleigh fit is poor for almost all cases." Sheikh et al. (1993) came to a similar conclusion "[T]here is a considerable deviation from the Rayleigh slope suggesting unsuitability of the Rayleigh model. ...Our finding suggests that for most of the measurements, the Nakagami model fits the best." Further support is given by explicit estimation of the Nakagami-m parameter; see examples in Table 2.4. As opposed to the Rayleigh distribution, the Nakagami distribution captures all different degrees of fading through the m parameter.

- The degree of fading has a large impact on the attainable wireless communication performance, see Section 2.2.3. This is especially true for small m, and relatively small deviations from the Rayleigh assumption $m = 1$ can cause

---

[19]The approximation used by Nakagami holds for $m > 1/2$ but the probability distribution is normalisable and well behaved for all $m > 0$.

[20]Additionally, it closely resembles the Rice distribution which assumes *one* strong path accompanied by several weak, approximately equal, paths. For $m = (K+1)^2/(2K+1)$, where $K \geq 0$ is the Rice factor, the Nakagami-m distribution provides a good fit to the Rice distribution (Hashemi, 1993).

Table 2.4: Published estimates of Nakagami-m fading figures.

Reference	Frequency GHz	Environment	Fading figure m		
			Range	Mean	St.d.
Beauvarlet and Virga (2002)	30	Office corridor	1.0–4.5		
Abouraddy and Elnoubi (2000)	10	Office floor	1.4–12.5		
Rubio et al. (2007)	0.9	Urban		1.5	0.34
Wennström (2002)	1.8	Office floor	1.1–6.3		

large changes in the required transmit energy. For this reason, the degree of fading is important in the topic of energy efficient communication under processing cost. We need to somehow incorporate it, be it in terms of the Nakagami-m, the Weibull, or the Rice distributions.

- The m parameter has numerically the same impact on average bit error rate performance as an mth order diversity system has in Rayleigh fading (Proakis, 2001, Sec. 14.4.1). Hence, the Nakagami-m distribution is convenient for analysis of diversity systems under different degrees of fading. In our context, this has importance when we study the possible energy-wise gains of utilising diversity in sensor network communication.

- The Nakagami-m distribution lends itself to tractable performance calculations in terms of bit error rates and outage capacity, see Section 2.2.3.

Remark 2.10 *Observe that we here find that the extra model parameter m of the Nakagami-m distribution is well motivated. This is in contrast with the propagation loss modelling (2.100) when we preferred to keep the model simple, see Remark 2.9*

Appendix 2.B.3   Temporal characteristics

Shadowing and fading have been described as time-dependent variations in the channel gain x, but we have said nothing about how fast changes occur. It is common to quantify the speed of change by use of the coherence time T_C as defined by the autocorrelation function of the impulse response in (2.91) (Proakis, 2001, Sec. 14.1.1). In essence, the coherence time quantifies the time over which the channel impulse response is practically unchanged. Let us define four types of temporal channel characteristics, here restated from Definition 2.1 on page 30;

Let T_P be the duration of a packet in seconds, and let T_{IP} be the (average) time between the beginning of each packet; the inter-packet duration. The channel coherence time is T_C.

1. If $T_\mathrm{C} \to \infty$, then the channel is *static*

2. If $T_\mathrm{IP} \ll T_\mathrm{C} < \infty$, then the channel is *quasi-static*

3. If $T_\mathrm{P} < T_\mathrm{C} < T_\mathrm{IP}$, then the channel is *slow*

4. If $T_\mathrm{C} < T_\mathrm{P}$, then the channel is *fast*

Appendix 2.C   Communication performance in shadowing and fading

Receiver performance is fundamentally dependent on the received signal-to-noise ratio *per bit*

$$\gamma \equiv \frac{\mathcal{E}_\mathrm{rec}}{N_0}, \tag{2.104}$$

where $\mathcal{E}_\mathrm{rec}$ is the received energy per bit and N_0 is the noise power spectral density (assumed constant over the receiver bandwidth).[21]   The received energy per bit is directly proportional to the energy per bit $\mathcal{E}_\mathrm{rad}$ radiated from the transmitter's antenna and the power gain x of the channel,

$$\gamma \propto x \frac{\mathcal{E}_\mathrm{rad}}{N_0}. \tag{2.105}$$

Due to this proportionality we can apply our propagation loss, shadowing and fading models also for γ.

Appendix 2.C.1   Average bit error rate

The probability of a bit error is in a static Gaussian channel determined by the constant signal-to-noise ratio γ per bit and, of course, the transmission scheme used. Although differences exists between transmission schemes, the overall impact of the per-bit signal-to-noise ratio is similar for most schemes, at least to first order. We here illustrate the degradation in average bit error rate performance induced by fading by the use of differential binary phase shift keying (DBPSK). Over a static Gaussian channel the expected bit error rate $B(\gamma)$ (or equivalently, probability of bit error) is

$$B(\gamma) = \frac{1}{2} e^{-\gamma}, \tag{2.106}$$

see (Proakis, 2001, Eq. (5.2-69)). Under the assumption that the receiver can obtain perfect estimates of the channel taps while the transmitter transmits at constant rate and power, we find the average bit error rate $\overline{B}(m, \overline{\gamma})$ over a Nakagami-m

---

[21]The signal-to-noise power ratio $\gamma_\mathcal{P} = \mathcal{P}_\mathrm{rec}/\mathcal{P}_\mathrm{noise}$ is related to the signal-to-noise ratio per bit by $\gamma_\mathrm{P} = \gamma R_b/W$, where R_b is the bit rate and W is the bandwidth of the receiver. This relation rests on the assumption of a matched filter receiver.
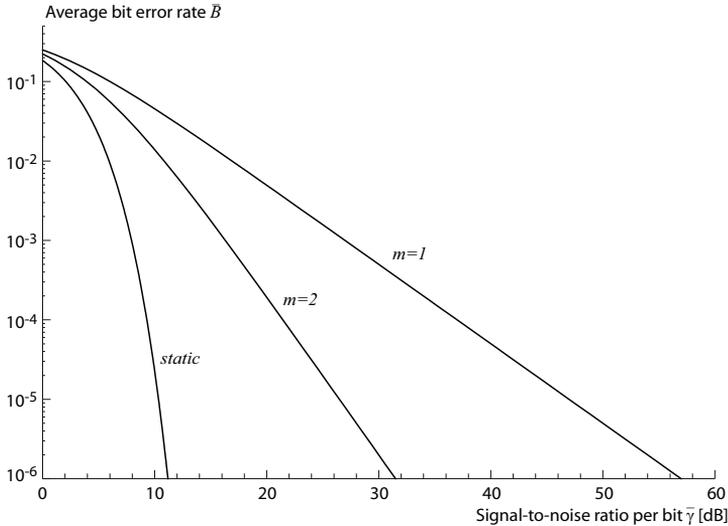
Figure 2.6: Average bit error rates $\overline{B}$ for differential binary phase shift keying in three different degrees of fading. The leftmost curve corresponds to the static Gaussian channel, the middle curve to Nakagami fading with fading figure $m = 2$, the rightmost curve to Rayleigh fading ($m = 1$).

channel as follows,

$$\overline{B}(m,\overline{\gamma}) = \int\limits_0^\infty B(\gamma)p(\gamma|m,\overline{\gamma})\mathrm{d}\gamma = \frac{1}{2}\left(\frac{m}{m+\overline{\gamma}}\right)^m. \qquad (2.107)$$

Henceforth, we simplify the notation by using $\overline{B}$ without explicitly expressing its dependence on the channel parameters. For large average signal-to-noise ratios $\overline{\gamma}$ we obtain

$$\overline{B} \approx \frac{m^m}{2\overline{\gamma}^m} \propto \overline{\gamma}^{-m}, \qquad (2.108)$$

which signifies the impact of m on the performance. We find with the aid of Mathematica 6 (Wolfram Research Inc., 2007) that the same scaling behaviour $\overline{B} \propto \overline{\gamma}^{-m}$ holds also for coherent phase shift keying (PSK), coherent frequency shift keying (FSK), non-coherently detected frequency shift keying (NCFSK) and M-ary quadrature amplitude modulation (MQAM). In conjecture, all modulations with exponentially decreasing bit error rates with increasing signal-to-noise ratio per bit will display the $\overline{\gamma}^{-m}$ error behaviour in Nakagami-m fading. In Figure 2.6 we show the DBPSK bit error rates for static and Nakagami-m fading channels. Observe that

- Without doubt the fading can incur a large increase in transmission energy, especially for low bit error rates in Rayleigh fading ($m = 1$). For instance, to

maintain $\overline{B} = 10^{-6}$ in Rayleigh fading the transmitter must use an output power level which is more than 30 000 times higher than in the static channel case $(m \to \infty)$.

- However, a change in the Nakagami parameter from $m = 1$ to $m = 2$ yields a reduction in output power of more than a factor 300. Consequently, even relatively small deviations from $m = 1$ can be important from an energy perspective.

We argue that the strong sensitivity to changes in m renders a pure Rayleigh assumption $m \equiv 1$ misleading in assessments of energy efficiency.

EXAMPLE 2.5 Node Density and the Nakagami-m Parameter

Assume that we are to deploy a wireless sensor network in an area where the propagation loss exponent $\kappa = 3.5$. Given the nodes' maximum transmit power and a required average bit error rate $\overline{B} = 10^{-6}$ we have found that the inter node distance in Rayleigh fading should be $d_{m=1}$. But what if the small-scale fading is instead best characterised by $m = 2$? From (2.108) we find that the required signal-to-noise ratio per bit then decreases by a factor of 354. By the use of the power law propagation loss model in (2.100) we obtain an increase in inter-node distance to $d_{m=2} = (354)^{1/\kappa} d_{m=1} = 5.35 d_{m=1}$. Consequently, by increasing the inter-node distance by a factor 5.35, the number of nodes required to ensure communication at the given level is reduced to 5.35^{-2}, or 3.5 percent, of the Rayleigh calculation. In other words, an over-deployment by a factor of $5.35^2 = 28.6$ has been avoided.[22]

The criterion of average bit error rate will in most cases be enough for assessing the energy efficiency of different transmission approaches, and we will make use of it several times in this thesis. But, in some applications it can be more appropriate to replace, or complement, the average bit error rate with the so-called outage criterion as discussed next.

Appendix 2.C.2   Outage performance

A wireless connection is said to be in *outage* if the received signal-to-noise ratio is below a certain level γ_{out}. By classifying channel condition as good or bad, and accepting loss of connectivity with a specified probability P_{out}, we suspend the impact of the worst conditions on the average performance. Likewise, under good conditions, we disregard from the fact that there are different degrees of "good". The outage criterion is thus of quite different character than the average criterion.

---

[22]Here, we have implicitly assumed that it is the communication that limits the node spacing, not the sensing application. This may not hold in some cases.

For the Nakagami channel we obtain the outage probability $P_{\text{out}} = P(\gamma < \gamma_{\text{out}}|m, \overline{\gamma})$ by integration over the gamma distribution,

$$P_{\text{out}} = \int_0^{\gamma_{\text{out}}} \frac{m^m}{\Gamma(m)\overline{\gamma}^m} \gamma^{m-1} e^{-\frac{m}{\overline{\gamma}}\gamma} \mathrm{d}\gamma = 1 - \frac{\Gamma\left(m, m\frac{\gamma_{\text{out}}}{\overline{\gamma}}\right)}{\Gamma(m)}, \qquad (2.109)$$

where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function. For small outage probabilities we can expand the expression around $\gamma_{\text{out}}/\overline{\gamma} = 0$ to find that

$$P_{\text{out}} = \frac{m^{m-1}}{\Gamma(m)}\left(\frac{\gamma_{\text{out}}}{\overline{\gamma}}\right)^m + O\left(\left(\frac{\gamma_{\text{out}}}{\overline{\gamma}}\right)^{m+1}\right). \qquad (2.110)$$

Hence, we have the same qualitative behaviour as displayed in (2.108) for the average bit error rate; the error probability decays like $\overline{\gamma}^{-m}$. Neglecting higher order terms we can use

$$\overline{\gamma} = m\gamma_{\text{out}}\left(m\Gamma(m)P_{\text{out}}\right)^{-1/m}. \qquad (2.111)$$

For instance, assume that the channel is quasi-static as given by Definition 2.1. Assume further that the application-specific goals can be met if i) less than one in a hundred packets are lost and ii) in the accepted packets the average bit error rate $B \leq 10^{-6}$. Consequently, γ_{out} is determined by the performance (2.106) over a Gaussian static channel; $\gamma_{\text{out}} \geq 13.12 = 11.2$ dB is required for $B = 10^{-6}$. In Rayleigh fading, we obtain from (2.111) that $\overline{\gamma} \geq 1300 = 31$ dB to achieve $P_{\text{out}} = 0.01$. Compare this with the required $\overline{\gamma} = 5 \cdot 10^5 = 57$ dB to achieve an average bit error rate $\overline{B} = 10^{-6}$ in Rayleigh fading. Which of the criteria that should be used–average, outage, or a combination–is determined by the application.

Considering shadowing effects, we can specify an acceptable outage probability $P_{\text{out}} = P(\overline{\gamma} < \overline{\gamma}_{\text{out}}|\overline{\overline{\gamma}}, \sigma_{\text{dB}})$ and find from the log-normal distribution in (2.101) that

$$\begin{aligned} P_{\text{out}} &= \int_0^{\overline{\gamma}_{\text{out}}} \frac{1}{\overline{x}} \cdot \frac{\mu}{\sqrt{2\pi}\sigma_{\text{dB}}} e^{-\frac{1}{2\sigma_{\text{dB}}^2}\left(\mu \ln(\overline{\overline{x}}/\overline{x}) - \frac{\sigma_{\text{dB}}^2}{2\mu}\right)^2} \mathrm{d}\overline{\gamma} \\ &= \frac{1}{2}\left[1 + \text{Erf}\left(\frac{\sigma_{\text{dB}}}{2\sqrt{2}\mu} - \frac{\mu}{\sqrt{2}\sigma_{\text{dB}}}\ln\left(\frac{\overline{\overline{\gamma}}}{\overline{\gamma}_{\text{out}}}\right)\right)\right]. \end{aligned} \qquad (2.112)$$

Here, $\mu = 10/\ln(10)$ and $\text{Erf}(\cdot)$ is the error function. We note that to maintain a certain outage probability for different shadowing degrees σ_{dB} roughly corresponds to a constant value of the second term inside the error function; $\ln(\overline{\overline{\gamma}}/\overline{\gamma}_{\text{out}})/\sigma_{\text{dB}} = $ const. Therefore, the average signal-to-noise ratio $\overline{\overline{\gamma}}$ must scale approximately exponentially with σ_{dB}, and we see in Figure 2.7 that the increase is in fact super-exponential. As a consequence, just like in the Rayleigh-Nakagami small scale case, care must be taken as worst case designs will grossly over-estimate the need for large transmit powers.
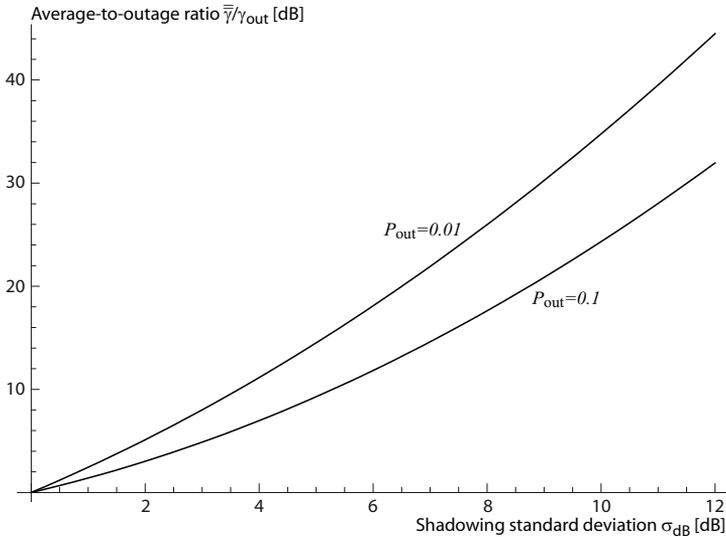
Figure 2.7: The ratio of the required average signal-to-noise ratio $\overline{\overline{\gamma}}$ to the outage signal-to-noise ratio γ_{out}. When the standard deviation σ_{dB} of the shadowing effects increase there is a super-exponential growth in the required margin $\overline{\overline{\gamma}}/\gamma_{\mathrm{out}}$. The top curve corresponds to an accepted outage probability $P_{\mathrm{out}} = 0.01$ while the lower curve corresponds to $P_{\mathrm{out}} = 0.1$.

Appendix 2.C.3   Channel capacity

Shannon (1948a,b) found the ultimate performance limit for error-free communication over a bandlimited static channel corrupted by Gaussian noise. His limit states that the maximum bit-rate C is

$$C = W \log_2 \left(1 + \gamma_{\mathcal{P}}\right) \qquad [\text{bit/s}], \qquad (2.113)$$

where W is the bandwidth and $\gamma_{\mathcal{P}} = \gamma C/W$ is the constant received signal-to-noise ratio for bit rate $R_b = C$. He also showed that there exists error correcting codes that can achieve this limit. Over fading channels, there is not one single, all-encompassing, capacity result to quote since the capacity depends on what is known about the channel behaviour and also the information about the channel available at the transmitter and the receiver. In almost all cases however, the capacity is degraded by fading[23].

For wireless sensor networks with long idle times and short packets, the most

---

[23]One exception is the capacity achieved for small signal-to-noise ratios when both the transmitter and the receiver have perfect information about the channel state at each moment (Goldsmith, 2005, p. 109). So-called water filling can then be applied by the transmitter; it makes use of good channel conditions to increase the bit rate, and the gain from this is larger than the loss during bad channel conditions.

appropriate capacity measure is probably that of *outage capacity*.[24] It applies to
quasi-static fading when there are long periods of fairly static channel conditions,
but long-term fading. The outage probability-capacity relation in Nakagami fading
is found from (2.109) and (2.113), and we obtain

$$
P_{\text{out}} = 1 - \frac{\Gamma\left(m, m\frac{2^{C'}-1}{\overline{\gamma}C'}\right)}{\Gamma(m)} = \frac{\left(\frac{2^{C'}-1}{\overline{\gamma}C'}\right)^m}{m\Gamma(m)}\overline{\gamma}^{-m} + O\left(\overline{\gamma}^{-m-1}\right), \tag{2.114}
$$

where $C' = C_{\text{out}}/W$ is the outage spectral efficiency and $P_{\text{out}} = P(C < C_{\text{out}}|m, \overline{\gamma})$.
It is impossible to achieve a non-zero capacity at zero outage probability for any
finite m. For finite m there is always a non-zero probability of a channel zero, and
error-free communication can not be guaranteed. At all non-zero outage probabili-
ties, however, we observe the same error behaviour as before, namely $P_{\text{out}} \propto \overline{\gamma}^{-m}$.

---

[24]So-called ergodic capacity is achieved by extremely long codewords that make sure that
each codeword can cover all fading states and thereby correct all errors without transmit
adaptivity (Goldsmith, 2005, p. 104). Under this assumption the ergodic capacity is

$$
C_{\text{erg}} = W \int\limits_0^\infty \log_2(1 + \gamma_{\text{P}})p(\gamma_{\text{P}}|m, \overline{\gamma}_{\text{P}})\mathrm{d}\gamma_{\text{P}}
$$

The presupposed use of very long codewords makes ergodic capacity less well suited than
outage capacity for assessing communication limits in data gathering networks.

Chapter 3

Fading and Polarisation Diversity, Channel Measurements and Analysis

D UE to destructive and constructive interference between radio signals that travel different paths from transmitter to receiver, there can be spatial and temporal variations of considerable magnitude in the received signal strength. This so-called multi-path fading can cause differences in the received signal-to-noise ratio of several orders of magnitude, even for displacements of less than half of the carrier wavelength. Deep fading dips disturb the wireless communication, and to ensure sufficient performance even during the rare, unpredicted, dips the transmitter has to use an output power that is much larger than a non-fading channel would require. For example, Figure 2.6 in Appendix 2.C shows that the output power required to guard against Rayleigh fading at a bit error rate $B = 10^{-6}$ is more than 38000 times (46 dB) larger than required for a non-fading channel.[1] Between these extremes there is a range of fading degrees, and as a consequence a range of required transmission energies, that we must consider in the design of energy efficient wireless sensor networks (see Example 2.5 for an illustration).

Although there are many published papers regarding channel measurements and modelling in general, there is a lack of sensor-network specific channel measurements with low-lying antennas or close-to-wall types of ar-

---

[1]The result pertains to differential binary phase shift keying with its bit error rate averaged over a Rayleigh distribution for the received signal-to-noise ratio per bit. Perfect channel state information at the receiver is assumed.

rangements. Those that exist, for instance the work by Sohrabi et al. (1999),
are to the best of our knowledge focussed on propagation loss and/or shad-
owing. The lack of relevant measurements of the degree of fading – which we
have chosen to include via the Nakagami fading figure m – in combination
with its large impact on the transmission energy is our first motivation to
perform sensor-network specific channel measurements.

From an energy perspective, the main problem with fading multi-path
channels is that it takes a lot of extra transmit power to overcome rare but
very severe channel conditions. Guarding against the occasional deep fading
dip by adjusting a constant power level wastes significant amounts of trans-
mission energy. To counter the negative effects of fading more efficiently
we can make sure that each information bit is "spread" over several channel
states and utilise this *diversity* to improve performance. Error correcting
codes utilise time diversity against noise and fast fading, but does not work
well in slow fading. Frequency diversity can be achieved by spreading or
multi-carrier communication, and hence requires larger bandwidth than the
original signal. Array antennas facilitates the use of spatial diversity, spread-
ing information over different spatial channels, but this is seldom an option
for sensor nodes due to their limited size. Therefore, *polarisation* diversity
suggests itself as an attractive possibility thanks to its compact configura-
tion.

According to Almers et al. (2007) surprisingly little attention has been
paid to polarisation properties in channel modelling, and we agree. Vaughan
(1990) studied the viability of polarisation diversity for mobile cellular com-
munications, but we have found no characterisation of the polarisation prop-
erties of wireless sensor network channels. Polarisation diversity will be of
practical use only if the polarisations vary with sufficient independence from
each other. Otherwise, if polarisations are strongly correlated, fading dips
will occur simultaneously and the performance is not improved. This has
motivated us to investigate polarisation diversity, and this chapter contains
an analysis of the correlation properties and diversity gains for measured
channels.

3.1   Polarisation diversity

Before we describe our measurements, let us define gain metrics for diversity
systems and discuss how energy gains from diversity is affected by fading
degree and branch correlations. In the analysis of antenna-diversity perfor-
mance, there are two general type of metrics commonly referred to as the

array gain and the *diversity gain*. The first refers to an increase in the average signal-to-noise ratio achieved by receiving multiple copies of the signal, while the latter refers to the decreased variability in the resulting signal. Below we make the meanings more precise.

Definition 3.1 *The* array gain A_{D} *of a diversity system is the increase in average signal-to-noise ratio achieved,*

$$A_{\mathrm{D}} \equiv \frac{\overline{\gamma}_{\mathcal{P},\mathrm{c}}}{\overline{\gamma}_{\mathcal{P},\mathrm{branch}}}, \tag{3.1}$$

where $\overline{\gamma}_{\mathcal{P},\mathrm{c}}$ is the average signal-to-noise (power) ratio of the combined signal, and $\overline{\gamma}_{\mathcal{P},\mathrm{branch}} = (\sum_{n=1}^{M} \overline{\gamma}_{\mathcal{P},n})/M$ is the average signal-to-noise ratio over all branches $n = 1, 2, \ldots, M$. ∎

Note the difference between average signal-to-noise ratio over all branches and the branch-wise average signal-to-noise ratios. The array gain with respect to a specific branch n will be denoted

$$A_{\mathrm{D},n} \equiv \frac{\overline{\gamma}_{\mathcal{P},\mathrm{c}}}{\overline{\gamma}_{\mathcal{P},n}}, \tag{3.2}$$

where $\overline{\gamma}_{\mathcal{P},n}$ is the average signal-to-noise ratio for the nth branch.

Definition 3.2 *If the average-error behaviour for large average branch signal-to-noise ratios can be asymptotically expressed*

$$P_{\mathrm{error}} \propto \left(\overline{\gamma}_{\mathcal{P},\mathrm{branch}}\right)^{-M}, \tag{3.3}$$

the diversity order *of the system is M. Here, P_{error} is the probability of error under consideration, that is bit error, symbol error or outage.* ∎

Observe that Definition 3.2 of diversity order regards the joint properties of the channel and the diversity scheme. For instance, the bit error rate for non-diversity differential binary shift keying over a Nakagami-m channel is $B \propto \overline{\gamma}_{\mathrm{P}}^{-m}$, see Appendix 2.C, showing that the inherent diversity order of the Nakagami-m channel is m.

Definition 3.3 *The* inherent diversity order *of a Nakagami-m channel is m.* ∎

This does not mean that the channel has any inherent *diversity*, but rather that the communication performance behaves as if we had a diversity system communicating over a Rayleigh fading channel.

Definition 3.4 *The* diversity order gain *of a diversity scheme is* Δ *if the branch-wise average-error slopes are given by (3.3) and the diversity scheme improves this to*

$$P_{\text{error}} \propto \left(\overline{\gamma}_{\mathcal{P},\text{branch}} \right)^{-\Delta \cdot M} . \tag{3.4}$$

That is, the diversity order is improved by a factor Δ *from* M *to* ΔM. ∎

A diversity scheme with diversity order gain $\Delta = 2$, say, will in Nakagami-m fading achieve an asymptotic bit error rate $B \propto \overline{\gamma}_{\mathcal{P},\text{branch}}^{-2m}$. [2]

3.1.1 Transmission energy gains in Nakagami fading

In the present context of energy efficient transmissions, we recognise two important facts about diversity order and degree of fading.

First, we see that if the required output power scales like $P_{\text{error}}^{-1/M}$, there are diminishing returns from an increased diversity order from Δ to $\Delta + 1$. For example, with $P_{\text{error}} = 0.01$ the transmission energy gain in going from $M = 1$ to $\Delta \cdot M = 2$ is about $(0.01^{-1})/(0.01^{-1/2}) = 10$, while the gain in going from order $\Delta \cdot M = 2$ to $(\Delta + 1)M = 3$ is $(0.01^{-1/2})/(0.01^{-1/3}) = 2.15$. The main gain is achieved when introducing diversity.

Second, remembering that the Nakagami-m channel has an inherent diversity order m, we conclude that the transmission energy gains from employment of diversity systems are strongly dependent on the fading figure. A scheme which in Rayleigh fading achieves a transmission energy gain G_{D} from the error slope, will in Nakagami fading achieve $(G_{\text{D}})^{1/m}$. The effect becomes more accentuated as the error probabilities get lower: The harder requirements we pose on performance, the larger the gains from diversity. But the larger the gains from diversity are, the more sensitive they are to the Rayleigh fading assumption. For assessment of diversity transmission energy gains it is important to consider the degree of fading, not only the order gain Δ of the scheme.

3.1.2 Unequal branch quality and branch correlations

It is important to consider the effects of correlations and uneven branch qualities because these effects lower the achievable diversity gains. Simon and Alouini (1999) have performed a quite thorough investigation of the

---

[2]Note that the definition of Δ does not hold for branches with different m. Combining two branches with different fading figures $m_1 \neq m_2$ results in diversity orders gains that are generally not expressible in a simple closed form, because combing two gamma distribution does not generally result in a new gamma distribution (Wolfram Research Inc., 2007).

matters for dual branch receiver selection-diversity systems[3] in Nakagami-m fading, and find that the diversity order gain of 2 is maintained even for correlated branches of uneven quality. However, a branch correlation coefficient $c = 0.9$ incurs a signal-to-noise penalty of about 5 dB, while $c = 0.5$ incurs approximately 2 dB in penalty. We will assess the branch dependence in our measurements by use of the sample correlation coefficient.

Definition 3.5 *The branch sample correlation coefficient c for a dual polarisation scheme is*

$$c \equiv \frac{\widehat{\mathrm{Cov}}\,(\xi_\mathrm{V}, \xi_\mathrm{H})}{\hat{\sigma}_\mathrm{V}\hat{\sigma}_\mathrm{H}}, \tag{3.5}$$

where ξ_V and ξ_H are the channel envelope gains vertical (V) and horizontal (H) polarisations respectively, while $\widehat{\mathrm{Cov}}$ and $\hat{\sigma}$ denote envelope sample covariance and envelope sample standard deviations respectively. ■

Even if branches are independent, the practical diversity order gain can be smaller than expected if the branches' signal-to-noise ratios are very different. This is because the branch of low quality seldom can contribute to the combined signal quality; in the limit of infinitely different branch qualities the dual branch system is effectively reduced to a single branch system. However, we still have the possibility to use the best branch and thereby gain signal quality even if we can not reduce the fading in that branch's signal. Regarding branch quality, Simon and Alouini (1999) found, specifically for selection diversity, that a 10 dB difference in average branch signal-to-noise ratios incurs approximately a 5 dB penalty.[4]

3.2 Measurement description

We chose two different environments where sensor networks are likely to be deployed; the forest and the office. Our investigations included both transmit and receive polarities – vertical and horizontal. All measurements were carried out using a pure carrier wave, so no information on delay spread was obtained. The signal generator provided a frequency reference for the vector signal analyser and they were thus always locked to the same reference oscillator. For each location and transmitter-receiver polarisation combination a 1.5 m sweep along a linear guide was performed on the receiver side while

---

[3]A receiver that uses only one branch at a time, but always the best one.

[4]To be strict, the branch quality is not merely the average signal-to-noise ratio $\overline{\gamma}_{\mathcal{P},n}$ but includes also the fading figures m_n (which can be different for different polarisations). Simon and Alouini (1999) assumed equal m for both branches.

Table 3.1: Measurement setup and parameters for the conducted measurements. All sweeps were performed for all four transmitter-receiver polarisation pairs VV, VH, HV and HH, where V denotes vertical and H horizontal. Non-line-of-sight (NLOS) and line-of-sight (LOS) conditions are indicated.

Environment	Frequency MHz	Distances m	Sweeps
Forest	868	[8, 16, 24, 32, 40]	1
	2400	[8, 16, 24, 32, 40]	1
Office floor (LOS)	434	[8, 16, 24, 32, 40, 48]	1
	868	20	9
	2400	[8, 16, 24, 32, 40, 48]	1
Office floor (NLOS)	434	20	15
		[8, 24, 32, 48]	1
		[8 16 24, 32, 40, 48]	2
	868	20	30
	2400	20	30

recording the in-phase and quadrature samples. All sweeps included in the following analysis are summarised in Table 3.1.[5]

3.2.1   Forest

Measurements were taken in a forest area with medium to tall trees accompanied by light understorey. The terrain is slightly undulating, see Figure 3.1. During the measurements the transmitting antenna was stationary, placed 15 cm above the ground. The receiving antenna was moving at a constant speed of roughly 0.1 m/s along a linear guide. The average distance to the ground was about 25 cm, but variations caused by ground roughness were present. Data were collected for two carrier frequencies, 868 MHz and 2400 MHz respectively.

---

[5]A couple of sweeps had to be discarded due to haphazard quantisation errors in the vector signal analyser. The quantisation in these measurements were severe and would have distorted the result considerably if included. The exclusion of some sweeps explains why the measurement setups in Table 3.1 are not complete.

Figure 3.1: Example picture from the forest in which measurements where performed. There were greater variation in the undulation and the density of the understorey than displayed in the figure.

3.2.2   Office floor

Office measurements were taken at the second floor in house 7 at the Ångström Laboratory, Uppsala University. The building has concrete outer walls and floors, but most inner partitions are plaster walls (with metal support). The antennas were kept close to the floor and/or close to a wall in order to achieve sensor network like transmission conditions. Figure 3.2 shows the floor and the measurement locations corresponding to Table 3.1.
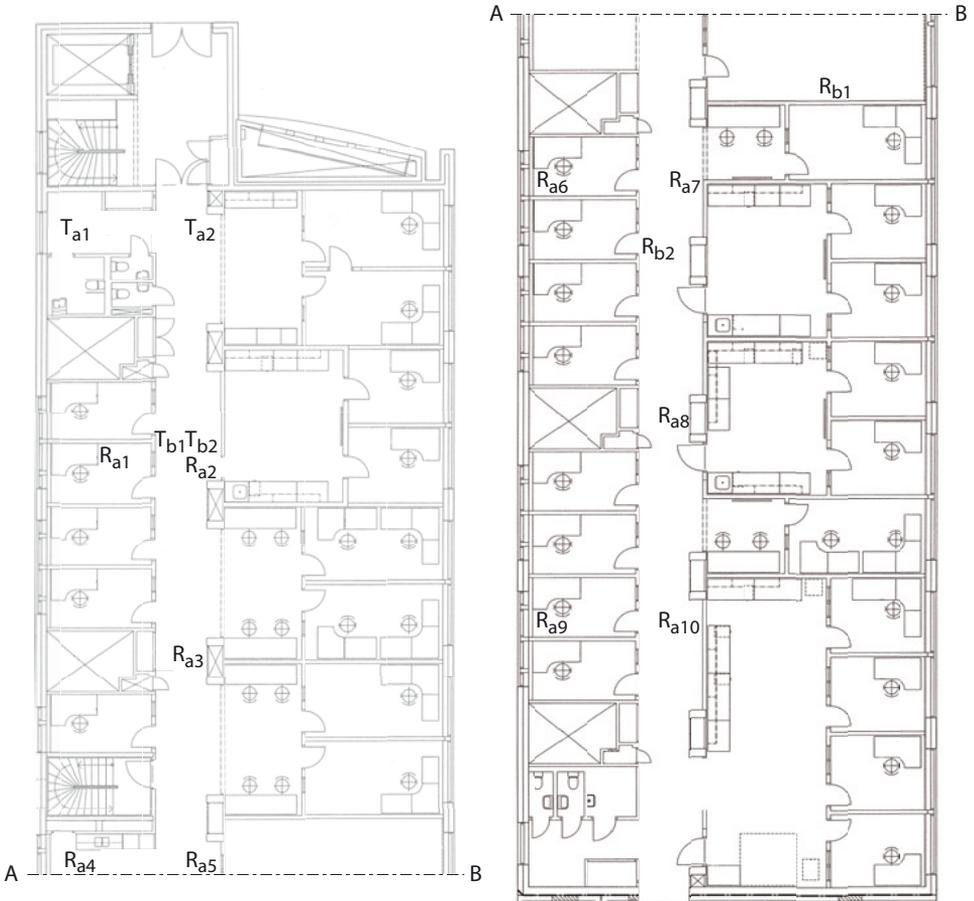
Figure 3.2: Floor 2, house 7 at the Ångström Laboratory, Uppsala University. Transmitter (T) and receiver (R) positions during measurements are labelled (a) or (b) according which batch they belong to. Data were taken for all combinations of transmitter-receiver locations within each batch. The data sets were during the analysis grouped into line-of-sight and non-line-of-sight groups, see Table 3.1.

3.3   Estimation of the m parameter

Adopting probability theory as the logic of inference, our posterior probability distribution for the fading figure m summarises our state of knowledge. The posterior can by application of the sum rule (2.32) and the product rule

(2.31) be expressed

$$p(m|D, I) = \int_0^\infty p(m|\overline{x}, D, I)p(\overline{x}|D, I)\mathrm{d}\overline{x} \qquad (3.6)$$

$$= \int_0^\infty p(m|\overline{x}, I)\frac{p(D|m, \overline{x}, I)}{p(D|\overline{x}, I)}p(\overline{x}|D, I)\mathrm{d}\overline{x}, \qquad (3.7)$$

where I denotes all background information we include (essentially the previous part of this thesis), $\overline{x}$ is the mean channel power gain, and D denotes the N collected data points $D \equiv \{d_1, d_2, ..., d_N\}$. The index n of each complex-valued data point d_n corresponds to a certain position along the sweep track used for the receiving antenna. By successive application of the product rule we can expand the likelihood $p(D|m, \overline{x}, I)$ for m to

$$p(D|m, \overline{x}, I) = \prod_{n=1}^N p(d_n|D_{n+1}^N, m, \overline{x}, I), \qquad (3.8)$$

where D_n^N is the partial data set $\{d_n, d_{n+1}, ..., d_N\}$. The measurements are corrupted by noise e_n,

$$d_n = A\sqrt{x_n}\exp(-j\theta_n) + e_n \qquad (3.9)$$

where $\sqrt{x_n} = \xi_n$ is the channel's amplitude gain for position n, θ_n is the corresponding phase shift and A is a known scaling constant determined by settings on the signal generator and the analyser respectively. For ξ_n we have assigned the Nakagami-m prior, and $p(x_n|m, \overline{x}, I)$ is hence the gamma distribution $\mathrm{Gam}(m, \overline{x})$, see Assumption 2.6 on page 29. For θ_n we assign from translational invariance a uniform prior, while we assign a Gaussian prior for e_n by application of MaxEnt under an average power constraint.

We could now obtain $p(d_N|m, \overline{x}, I)$, the last factor in (3.8), by a change of variables according to (3.9). However, the issue of assigning the remaining factors $p(d_n|D_{n+1}^N, m, \overline{x}, I)$ introduces the whole new problem of modelling the dependence between samples: What does the data D_{n+1}^N tell us about sample d_n? Samples between positions $n+1$ and N definitely tells us something about the sample at position n – this is indeed the basis for all channel prediction methods developed to increase wireless system capacity – but requires the introduction of more detailed, temporal/spatial, channel models.[6]

---

[6]For more information on channel estimation based on probability theory as extended logic, the reader is referred to the thesis by Aronsson (2007).

To avoid the complexity that temporal/spatial channel modelling would introduce we will adopt a simplified estimation approach which includes an assumption on logical independence between samples: $p(d_n, d_{n+1}|, m, \overline{x}, I) = p(d_n|, m, \overline{x}, I)p(d_{n+1}|, m, \overline{x}, I)$. The most prominent loss caused by this simplification is that we cannot obtain a correct measure of the uncertainty in our estimates, as would have been provided by the posterior $p(m|D, I)$. An independence assumption will correspond to a certain amount of "sample reuse" – samples are correlated and carry partly the same information, but this fact is not recognised under our independence assumption – yielding an overly optimistic accuracy. However, we conjecture that the estimate itself will not be changed much as long as the prior $p(m|\overline{x}, I)$ is vague with respect to the amount of data used.[7] This is indeed the case here if we use the prior in Figure 2.5 on page 55.

3.3.1   Maximum likelihood estimation with independence

Below we make a couple of simplifying assumptions which we incorporate in the background information, henceforth denoted $I' \equiv \{I, \text{assumptions}\}$.

True estimate. The marginalisation over the average power gain $\overline{x}$ in (3.6) is avoided by assuming that a maximum likelihood (ML) estimate $\overline{x}_{\text{ML}}$ is indeed the true value; $p(\overline{x}|D, I') = \delta(\overline{x} - \overline{x}_{\text{ML}})$. The marginalisation integral in (3.6) then collapses and we obtain the approximation

$$p(m|D, I') = p(m|\overline{x}_{\text{ML}}, D, I'). \tag{3.10}$$

No prior information. We omit the prior knowledge about m, which we indeed have quantified in (2.85), with the motivation that it will, due to its vagueness, be overwhelmed by the data anyway.[8] Let $p(m|\overline{x}, I) = $ const. for simplicity, and hence (3.10) becomes

$$p(m|D, I') \propto p(D|m, \overline{x}_{\text{ML}}, I'). \tag{3.11}$$

Independent samples. We neglect the logical dependence between measurements d_n and let $p(d_n|D_{n+1}^N, m, \overline{x}, I') = p(d_n|m, \overline{x}, I')$. We thereby

---

[7]The estimate will correspond to the situation of a "randomised" data set.

[8]Observe that if we had very accurate prior information about m, the prior probability density $p(m|\overline{x}, I)$ would be sharply peaked and have strong influence on the estimate. The prior in (2.85) is fairly non-informative, so here it has little influence, but later in the thesis we must use it because it then represents our state of knowledge in lack of measurements.

introduce the common assumption of "independent and identically distributed data" allowing us to rewrite (3.11) to

$$p(m|D, I') \propto \prod_{n=1}^{N} p(d_n|m, \overline{x}_{\mathrm{ML}}, I').$$
(3.12)

No noise. The envelope measurements are virtually noise free – the signal-to-noise ratio in the envelope measurements is typically larger than 20 dB – so we can effectively set

$$\begin{aligned} e_n &\equiv 0, \\ A &\equiv 1, \end{aligned}$$
(3.13)

in (3.9). We further assign

$$\begin{aligned} p(d_n|m, \overline{x}, I') &= p(\sqrt{x_n}|\theta_n, m, \overline{x}, I')p(\theta_n|m, \overline{x}, I') \\ &= p(\xi_n|m, \overline{x}, I')p(\theta_n|I'), \end{aligned}$$
(3.14)

because we cannot infer anything about the channel gain x_n from the phase θ_n, whose value in turn can not be inferred from $(m, \overline{x})$.

By Assumption 2.6 on page 29, the likelihood $p(\xi_n|m, \overline{x}_{\mathrm{ML}}, I')$ for m is the Nakagami-m distribution

$$p(\xi_n|m, \overline{x}_{\mathrm{ML}}, I') = \frac{m^m}{\Gamma(m)\overline{x}_{\mathrm{ML}}^m}\xi_n^{2m-1}e^{-\frac{m}{\overline{x}_{\mathrm{ML}}}\xi_n^2}.$$
(3.15)

We can now express our simplified posterior (3.12) – corresponding to the full solution in (3.7) – as follows

$$p(m|D, I') \propto \prod_{n=1}^{N} \frac{m^m}{\Gamma(m)\overline{x}_{\mathrm{ML}}^m}|d_n|^{2m-1}e^{-\frac{m}{\overline{x}_{\mathrm{ML}}}|d_n|^2}.$$
(3.16)

Here, we have suppressed all probabilities that are independent of m, for instance $p(\theta_n|I')$, because they do not affect inferences about m. To find the most probable value m_{ML} we take the logarithm of the right hand side of (3.16), differentiate with respect to m and set to zero. As shown by Cheng and Beaulieu (2001), the resulting equation to solve is

$$\ln(m_{\mathrm{ML}}) - \Psi(m_{\mathrm{ML}}) = \ln\left(\overline{x}_{\mathrm{ML}}\right) - \frac{1}{N}\sum_{n=1}^{N}\ln\left(|d_n|^2\right),$$
(3.17)

where the maximum likelihood estimate of $\overline{x}$ is

$$\overline{x}_{\mathrm{ML}} = \frac{1}{N} \sum_{n=1}^{N} |d_n|^2, \tag{3.18}$$

and $\Psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ is the psi, or digamma, function. We can not solve (3.17) for the estimate m_{ML} analytically, but the solution is readily obtained numerically.

Remark 3.1 *Our long, and to some extent agonising, route via the original posterior (3.7) and all simplifications is taken to make sure that no assumptions are swept under the rug. Direct application of (3.18) conceals the simplifications made, but now we know the assumptions behind the estimate in (3.17).*

In Section 3.3.3 we present the numerical maximum likelihood estimates calculated from our measurement data, but we first make a digression on sufficient statistics and maximum entropy. We hereby try to clarify the connection between the m and $\overline{x}$ parameters as *macroscopic constraints* on the fading gain, and all the simplifying assumptions we have made to arrive at (3.17). Employing only these constraints in entropy maximisation amounts quantitatively to the present assumptions on noise and prior information.

3.3.2   Maximum likelihood and maximum entropy

We identify two sufficient statistics for m in the foregoing derivation of the estimate m_{ML}, namely the sample averages

$$
\begin{aligned}
T_1(D) &= \frac{1}{N} \sum_{n=1}^{N} |d_n|^2 \\
T_2(D) &= \frac{1}{N} \sum_{n=1}^{N} \ln\left(|d_n|^2\right),
\end{aligned}
\tag{3.19}
$$

which are found in (3.18) and (3.17) respectively. As sufficient statistics, $T_1(D)$ and $T_2(D)$ are the only functions of the data that appear in our posterior[9] (3.16) for m. This means that out of all the properties of the data set D, only $T_1(D)$ and $T_2(D)$ are relevant to the question we are asking.

Recall Example 2.4 in which we employed the principle of maximum entropy to assign a fading distribution under constraints on the average

---

[9]The statistic $T_2(D)$ is present through the relation $\ln\left(\prod |d_n|^2\right) = \sum \ln\left(|d_n|^2\right)$.

channel power gain $E(x|I) \equiv F_1$ and the average scale of the variations $E(\ln(x)|I) \equiv F_2$. The maximisation of entropy resulted in a gamma distribution

$$p(x|F_1, F_2, I) = \frac{m^m}{\Gamma(m)F_1^m} x^{m-1} e^{-m\frac{x}{F_1}}, \qquad (3.20)$$

where the m parameter is found as the solution to

$$\ln(m) - \Psi(m) = \ln(F_1) - F_2. \qquad (3.21)$$

Upon comparing (3.21) with the maximum likelihood estimate in (3.17) we immediately see the similarity. By replacing the averages in the maximum entropy solution with their corresponding sample averages $T_1(D)$ and $T_2(D)$, the sufficient statistics, we arrive at the maximum likelihood equation (3.17).

The connection with the sufficient statistics in (3.19) is by no means a lucky coincidence of our particular problem, but comes from the fact that the class of all maximum entropy distributions is exactly the class of all distributions with sufficient statistics (Kullback, 1968). The coincidence between maximum likelihood and maximum entropy is the consequence of two assumptions :

1. No noise,

2. No prior information,

see (Jaynes, 1982). Remember, however, that $T_1(D)$ and $T_2(D)$ are not sufficient statistics in the more complete formulation of the problem which is conditioned on I instead of I' which includes the simplifying assumptions (a discussion of sufficiency and prior information is given by Jaynes (2003, Ch. 8)). When noise, prior information and nuisance parameters are present only the full solution as provided by probability theory as logic will be optimal.

By including only the macroscopic constraints on $E(x|I)$ and $E(\ln(x)|I)$ in our channel model, we do not give the maximum entropy principle any reason to introduce correlations and therefore it automatically introduces the "independent and identically distributed" assumption. Any correlation would reduce the entropy in a way the constraints did not warrant.

3.3.3 Numerical maximum likelihood estimates

Here we present the maximum likelihood estimates found from application of (3.17) to our measurement data. The data sets obtained for different

Table 3.2: Nakagami fading figures for vertical (V) and horizontal (H) transmit and receive polarisations respectively in a partly obstructed forest environment. The left column gives the transmit-receive combination, for instance VH for vertical transmit polarity and horizontal receive polarity.

Polarisation	Fading figure	
[−]	m_{ML}	
	868 MHz	2400 MHz
VV	1.49	1.75
VH	1.89	1.32
HV	1.59	1.57
HH	1.03	1.57

distances were combined into one set after normalising the received average power to one.[10]

Remark 3.2 *In our modelling framework we have modelled fading and shadowing separately, with the m parameter quantifying the severeness of the fading only. We here use the data as if no shadowing was present during the relatively short measurement sweeps. However, shadowing effects do sometimes occur on the same spatial scale as fading and in such cases the m parameter will be under-estimated since we here consider the fading model solely – the m parameter becomes smaller in order to account for low received power in shadowed segments. In that sense, our estimates are conservative.*

The results in Table 3.2 pertain to the forest measurements grouped according to the transmitter-receiver polarisations. At least for a carrier frequency of 868 MHz there are clearly noticeable, but not extreme, differences between the combinations of polarities. Horizontal-to-horizontal polarisation is the worst choice with respect to the degree of fading. It is also interesting to note that the best polarisation for 868 MHz is the worst for 2400 MHz, indicating that no given polarisation pair is uniformly the best even in a given environment.

Fading figures for line-of-sight measurements taken indoors are collected in Table 3.3. As we see in Table 3.3 there are good reasons to believe that

---

[10]From (3.17) we find that this does not affect the estimate of m because the equation can be written

$$\ln(m_{\mathrm{ML}}) - \Psi(m_{\mathrm{ML}}) = -\frac{1}{N} \sum_{n=1}^{N} \ln\left(\frac{|d_n|^2}{\overline{x}_{\mathrm{ML}}}\right).$$

Normalisation is inherent in the estimate and this is intuitively pleasing as m is perceived as a "shape" parameter.

Table 3.3: *Indoor line-of-sight measurements. Impact of polarisation on the degree of fading. V denotes vertical and H horizontal, and are given in transmit-receive order.*

Polarisation	Fading figure		
[−]	m_{ML}		
	434 MHz	868 MHz	2400 MHz
VV	6.07	1.80	2.12
VH	8.69	3.14	1.32
HV	3.85	1.81	3.09
HH	16.2	18.5	0.97

polarisation has quite large impact on the fading degree; some polarisations are more "richly scattered" than other. It is remarkable that the HH combination exhibits such extreme differences between low and high frequencies. On the whole, there is – as expected for a line-of-sight scenario – a distinct deviation from the Rayleigh distribution $m = 1$, most pronounced for 434 MHz.

The non-line-of-sight results are given in Table 3.4. We note the significant decrease in the fading figure m as compared to the line-of-sight case. For 868 MHz and, especially, 2400 MHz we are now much closer to the Rayleigh fading case. For these two frequencies there are in addition no obvious polarisation advantages with respect to the severeness of the fading. On the other hand, in the 434 MHz case there remains a clear distinction from the Rayleigh fading model as well as a difference between polarisations. A plausible explanation is that the lower frequencies better penetrate the indoor walls and create "partial line-of-sight" – which is precisely what the Nakagami-m model handles better than the rayleigh model. Horizontal-to-vertical polarisation seems markedly better than the other choices.

3.4 Polarisation diversity: measured correlations and diversity gains

In this section we analyse our forest and office measurements with respect to receiver branch correlations, array gains and the diversity order gain, or "fading reduction". Sample correlation coefficients c are calculated according to (3.5). Individual array gains for each respective receive polarisation, $A_{\mathrm{D},V}$ and $A_{\mathrm{D},H}$, are found by application of (3.2), and the common array gain A_{D} is computed as defined in (3.1).

Table 3.4: *Indoor non-line-of-sight measurements. Impact of polarisation on the degree of fading. V denotes vertical and H horizontal, and are given in transmit-receive order.*

Polarisation	Fading figure		
[−]	m_{ML}		
	434 MHz	868 MHz	2400 MHz
VV	2.11	1.12	1.03
VH	2.54	1.13	1.08
HV	7.13	1.15	1.01
HH	2.04	1.20	1.15

Since different diversity schemes – such as selection diversity, switched diversity and maximum ratio combining – have different performance we cannot present entirely general gains. Our reference diversity technique in our numerical analysis is maximum ratio combining (MRC); other simpler schemes are analysed in Chapter 5. We estimate the Nakagami fading figure m for the combined signal to study the reduction in fading severeness that maximum ratio combining can achieve. The estimates are obtained by application of (3.17).

3.4.1 Numerical results for polarisation diversity

In Table 3.5 we summarise the sample correlation coefficients calculated from our measurements with vertical and horizontal receive polarisations. The results are encouraging as it seems that the two branches are, for all practical purposes, uncorrelated; most coefficients are smaller than 0.1. If these results carry over to other environments, the use of polarisation diversity should definitely be considered in node radio design.

The array gains achieved by a maximum ratio combining scheme are presented in Table 3.6. For reference we observe that the theoretical array gain A_{D} for independent branches of equal quality is 3 dB (Goldsmith, 2005). It is noteworthy that occasionally the gains from the vertical and horizontal polarisations differ significantly (for instance the horizontal transmit polarisation in the line-of-sight (LOS) case at 434 MHz) and the reason is that one branch has significantly better average quality than the other. Overall, our conclusion from Table 3.6 is that polarisation diversity delivers gains close to what can be predicted from theory.

We give in Table 3.7 the estimated fading figures, which are relevant to the issue of diversity order. Diversity order is often related to Rayleigh fading

Table 3.5: *Sample correlation coefficients between received polarisations, estimated as in (3.5). Results are given for each transmit polarity.*

Environment	Transmit polarity	Correlation coefficient c		
		434 MHz	868 MHz	2400 MHz
Forest	V	(-)	0.03	-0.14
	H	(-)	-0.10	-0.08
Office	V	0.08	0.04	-0.23
(LOS)	H	-0.35	-0.09	0.07
Office	V	0.16	-0.08	0.05
(NLOS)	H	0.36	-0.08	0.06

Table 3.6: *Maximum ratio combining gains for each combination of transmit-receive polarisation. Both branch-wise array gains $A_{D,n}$, see (3.2), and the array gain A_D, see Definition 3.1, are given.*

Environment	Transmit-Receive polarisations	MRC array gain [dB]					
		434 MHz		868 MHz		2400 MHz	
		$A_{D,n}$	A_D	$A_{D,n}$	A_D	$A_{D,n}$	A_D
Forest	VV	(-)		0.5		4.5	
	VH	(-)	(-)	12	2.8	2.6	2.6
	HV	(-)		2.6		4.5	
	HH	(-)	(-)	5.3	2.3	2.9	2.4
Office	VV	1.5		1.1		4.4	
(LOS)	VH	8.5	2.3	9.6	2.5	2.4	2.7
	HV	15		16		4.1	
	HH	0.44	2.7	0.2	2.9	2.9	2.6
Office	VV	6.5		2.1		5.6	
(NLOS)	VH	2.2	2.2	5.4	2.6	1.7	2.8
	HV	3.2		6.8		4.3	
	HH	6.6	1.4	2.0	2.3	2.2	2.9

Table 3.7: *Estimated Nakagami-m fading figures for the maximum ratio combined branches. Also displayed is the increase $\overline{\Delta}$ in average fading figure, an ad-hoc approximation to the diversity order gain in (3.4).*

Environment	Transmit polarity	MRC fading figure / order gain					
		434 MHz		868 MHz		2400 MHz	
		m_{ML}	$\overline{\Delta}$	m_{ML}	$\overline{\Delta}$	m_{ML}	$\overline{\Delta}$
Forest	V	(-)	(-)	1.84	1.09	3.46	2.25
	H	(-)	(-)	3.58	2.73	2.77	1.76
Office	V	8.06	2.18	3.50	1.42	4.00	2.33
(LOS)	H	31.6	3.15	20.0	1.97	3.40	1.67
Office	V	2.59	1.11	2.20	1.96	1.97	1.87
(NLOS)	H	3.56	0.776	2.56	2.18	2.14	1.98

and independent branches, and under these circumstances the maximum ratio combining approach should attain a dual-branch post-combining fading figure $m = M = 2$. Although the order gain Δ is not defined for the case of different branch diversity orders or different branch qualities, we can get an idea of the order gain by use of the *ad-hoc* metric

$$\overline{\Delta} \equiv 2 \frac{m_{\mathrm{ML,D}}}{m_{\mathrm{ML,V}} + m_{\mathrm{ML,H}}} \qquad (3.22)$$

which compares the average pre-combining fading figure with the post-combining fading figure. We see in Table 3.7 that with only two exceptions $m > 2$ and we can again conclude that polarisation branches seem sufficiently uncorrelated to meet theoretical predictions. Regarding the exceptions $m < 2$ we observe that they pertain to cases with highly unbalanced branch quality combined with severe branch fading. There are in Table 3.7 also a few exceptionally large fading figures which correspond to line-of-sight measurements. Even though maximum ratio combining still attains a noticeable increase in the fading figure, it will have minor impact on the performance because $m = 18$, say, is favourable enough to almost disregard from the fading effects altogether.

The more severe the fading is, the more we need diversity techniques to alleviate the negative effects on transmission energy consumption. Now it seems, both from theoretical arguments and our measurements, that the polarisation diversity improves as the channel variability increases; the richer the scattering is, the more severe the fading will be, and the better the polarisation diversity becomes. We deem it highly likely that rich scattering

causing severe (Rayleigh) fading also erases the correlation between the polarisations; otherwise we must face an environment with the curious property that all reflections are almost identical for the two polarisations. Polarisation diversity is available when it is most needed.

3.5 Conclusions

Based on our channel measurements carried out in forest and office environments, under line-of-sight and non-line-of-sight conditions, we draw the following conclusions.

- The Nakagami-m model is preferable to the more limited Rayleigh model; our data support that significant deviations do occur in practice.[11]

- Polarisation branches are uncorrelated, a fact that encourages the consideration of polarisation diversity as a means of saving transmission energy in fading environments. Additionally, the measurements support the theory that the more severe the fading is, the smaller are the correlations between branches.

- Polarisation branches can be of considerably different quality. This opens up for simple schemes relying on branch selection instead of complicated coherent combining. We discuss diversity schemes further in Chapter 5.

3.6 Future work

We have studied two relevant environments, a forest and an office area, but good channel characterisations for other sensor network environments would be informative and useful. Our planned future work is broadly the following:

1. Investigate other environments: urban areas, traffic related scenarios, water surfaces, open fields, etc.

2. Study temporal characteristics in more detail.

3. Combine propagation loss, shadowing and fading in joint estimation. Perform proper model selection on a larger set of candidate models.

---

[11]A future task is to perform a thorough model-selection analysis based on more data and explicit temporal/spatial channel modelling.

Chapter 4

Power Control and Rate Adaption

IN this chapter we investigate the impact of circuit processing energy on wireless node-to-node (point-to-point) communication. Although one must eventually consider the whole network aspect of a wireless sensor network, he who has a good understanding of energy efficient node-to-node communication will be in a better position to address the network issues than he who does not. For this reason we start with a slightly idealised energy optimisation based on the Shannon capacity limit and then move on to study transmit power control under feedback costs, (adaptive) error correcting codes and adaptive modulation. In all cases it is the trade-off between transmission and processing costs that is in focus.

We find that some existing radios would loose, or at least not benefit, energy-wise from the use of transmit power control and/or error correcting codes. Sensor nodes equipped with these radios could very well be set to use fixed output power and uncoded data transmissions. The reason is that these nodes' processing energy costs dominate and unless the transmission-to-total-processing ratio ρ', see (2.8), can exceed $1/3$, their total energy consumption will, in spite of reduced transmission energy, increase if the mentioned techniques are employed. However, there are radios that do benefit from power control and coding, at least in the upper part of their transmit power range where $\rho' > 0.4$. Here, the energy savings increase fairly rapidly and when the transmission energy becomes dominant, $\rho' > 1$, the benefits from power control and coding are clear. Our general recommendation is then to use a low-feedback, coarse grained power control scheme to counter slow fading, in combination with simple, possibly slowly adaptive, block codes to counter noise and possibly fast fading channels.

Considering adaptive quadrature amplitude modulation (QAM), which offers the possibility to reduce transmission energy costs by faster packet transmissions, we find that there are indeed large savings to be made for transmission-to-total-processing ratios below 1/10. However, we show that the complex, processing intensive, hardware platform required for adaptive QAM outweighs the attainable benefits; the initial processing energy – which can be reduced by adaptive QAM – is too large compared to simpler radio architectures.

4.1 Shannon-limit rate optimisation

Consider a radio transmission taking place with a transmission energy per bit $\mathcal{E}_{\mathrm{T},0}$ J, and at a spectral efficiency of C_0 bits/s/Hz. Assume further that the spectral efficiency can, at the cost of increased transmit energy per bit, be increased so that the transmission time per bit is decreased (and vice versa, that the spectral efficiency can be decreased to save transmission energy per bit, but at the cost of increased transmission time per bit). In order to find the energy-optimal tradeoff between transmission energy and transmission time, which translates to processing energy, through adaption of the spectral efficiency we must know how the energy and time scales with the spectral efficiency. Let us for this purpose use the Shannon capacity limit for static Gaussian channels.

The lowest achievable received energy per bit that can guarantee error-free reception at a given spectral efficiency C_{e} is for the static Gaussian channel given by the Shannon limit (Shannon, 1948a)[1]

$$\mathcal{E}_{\mathrm{rec}} = N_0 \frac{2^{C_{\mathrm{e}}} - 1}{C_{\mathrm{e}}}. \tag{4.1}$$

Here N_0 is the noise power spectral density and $C_{\mathrm{e}} \equiv C/W$, where C is the Shannon rate in bits per second and W is the channel bandwidth in hertz (Hz). Consequently, the minimum energy per bit in (4.1) is always increasing with the spectral efficiency C_{e} and the "traditional" wisdom is that systems are most energy-efficient at very low spectral efficiency; the limit is $\mathcal{E}_{\mathrm{rec}} \to \ln(2) N_0$ when $C_{\mathrm{e}} \to 0$. This limit can be approached by reducing the bit rate for a given bandwidth – coding over time – and/or by increasing the bandwidth for a given bit rate – coding over frequency. But the conclusion that low spectral efficiency saves energy is not generally valid

---

[1]See also (2.113) on page 71; by replacing the signal-to-noise ratio according to $\gamma_{\mathcal{P}} = \mathcal{E}_{\mathrm{rec}} C/(N_0 W)$ we arrive at the present form when we solve for C/W.

for our present problem in which the radio circuits' processing energy is not negligible.[2]

The idealised energy model

Consider now an idealised node radio system which can communicate at the Shannon-limit energy levels in (4.1), assuming that communication can be undertaken at any spectral efficiency $C_e \geq 0$, but is constrained to a fixed bandwidth W. It has an ideal power amplifier which has the same efficiency at all output power levels; its degradation exponent in (2.14) is $g = 1$. Consequently, the total transmission energy per bit is by the use of (2.14)

$$
\begin{aligned}
\mathcal{E}_\mathrm{T} &= K \left(\mathcal{E}_\mathrm{rec}\right)^{1/g}, \\
&= [g = 1] \\
&= K\mathcal{E}_\mathrm{rec},
\end{aligned}
\tag{4.2}
$$

where K is a propagation and system dependent constant which will cancel out in the following calculations. By combining the Shannon limit (4.1) and (4.2) we can relate the initial operating point $(\mathcal{E}_\mathrm{T,0}, C_0)$ to any other operating point $(\mathcal{E}_\mathrm{T}, C_e)$ as

$$
\mathcal{E}_\mathrm{T} = \mathcal{E}_\mathrm{T,0} \frac{C_0}{2^{C_0} - 1} \frac{2^{C_e} - 1}{C_e}.
\tag{4.3}
$$

The expression in (4.3) shows how the transmission energy cost per bit scales with the spectral efficiency, and we can quantitatively assess the change in going from C_0 to C_e.

The radio circuitry draws a constant supply current; it has constant processing power consumption $\mathcal{P}_\mathrm{RP}$. Under the assumption that it can perform ideal start-up and shut-down, from and to an ideal sleep mode, it will consume a processing energy per-bit which scales linearly with the per bit transmit duration T_bit; $\mathcal{E}_\mathrm{RP} = \mathcal{P}_\mathrm{RP} T_\mathrm{bit}$. In turn the bit duration is inversely proportional to the rate $C = WC_e$, and due to the fixed bandwidth W we can express the processing energy

$$
\mathcal{E}_\mathrm{RP} = \mathcal{E}_\mathrm{RP,0} \frac{C_0}{C_e},
\tag{4.4}
$$

---

[2]Verdu (2002) has in addition shown that conclusions drawn from the infinite bandwidth limit can be misguided for finite bandwidths, even when processing energy is not included: "Indeed,... low spectral efficiency does not imply disregard for the bandwidth required by the system". Consequently, the infinite bandwidth limit should be used with some care.

where $\mathcal{E}_{\mathrm{RP},0}$ is the per-bit processing energy at efficiency C_0.

We have here a typical trade-off situation between the transmission energy (4.3), which increases with C_{e}, and the processing energy (4.4), which decreases with C_{e}. The total per-bit energy is

$$
\begin{aligned}
\mathcal{E}_{\mathrm{tot}} &= \mathcal{E}_{\mathrm{RP}} + \mathcal{E}_{\mathrm{T}} \\
&= \mathcal{E}_{\mathrm{RP},0}\frac{C_0}{C_{\mathrm{e}}} + \mathcal{E}_{\mathrm{T},0}\frac{C_0}{2^{C_0}-1}\frac{2^{C_{\mathrm{e}}}-1}{C_{\mathrm{e}}}.
\end{aligned}
\tag{4.5}
$$

The absolute values of $\mathcal{E}_{\mathrm{RP}}$ and $\mathcal{E}_{\mathrm{RP},0}$ will not determine the energy-optimal spectral efficiency, it is their relative size that is important. Therefore, to study the transmission-processing tradeoff, the use of the transmission-to-processing ratio is adequate and simplifies the analysis to one parameter. Rewriting (4.5) by the use of the transmission-to-total-processing ratio $\rho' = \mathcal{E}_{\mathrm{T}}/(\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pt}})$ defined in (2.8) on page 21, we obtain

$$
\breve{\mathcal{E}}_{\mathrm{tot}} = \frac{C_0}{C_{\mathrm{e}}} + \rho'_0\frac{C_0}{2^{C_0}-1}\frac{2^{C_{\mathrm{e}}}-1}{C_{\mathrm{e}}}.
\tag{4.6}
$$

Here, the breve symbol $\breve{\mathcal{E}}_{\mathrm{tot}}$ shows that we have normalised the expression, this time with respect to $\mathcal{E}_{\mathrm{RP},0}$. Adjusting the spectral efficiency from C_0 to C_{e} results in a new transmission-to-total-processing ratio ρ'_{e}, which from (4.6) is found to be

$$
\rho'_{\mathrm{e}} = \rho'_0\frac{2^{C_{\mathrm{e}}}-1}{2^{C_0}-1}.
\tag{4.7}
$$

Observe that an increase in the spectral efficiency increases the transmission-to-total-processing ratio through decreased processing energy, not only through increased transmission energy.

4.1.1   The optimum transmission-processing tradeoff

Let us now optimise the spectral efficiency C_{e} with respect to the total energy per bit in (4.6).

Lemma 4.1 *The energy-optimal spectral efficiency C_{opt} for a system whose energy consumption is described by (4.6) is*

$$
C_{\mathrm{opt}} = \frac{1}{\ln(2)}\left[1 + W_{\mathrm{L}}\left(\xi\right)\right],
\tag{4.8}
$$

where

$$
\xi = \frac{2^{C_0}-1-\rho'_0}{e\rho'_0}
\tag{4.9}
$$

and $W_{\mathrm{L}}(\cdot)$ is the Lambert W function.[3]

 Proof: See Appendix 4.A.   ■

Depending on the initial transmission-to-total-processing ratio ρ_0', the optimum trade-off can be achieved either by an increase or a decrease in the spectral efficiency from C_0.

- When $\rho' \to \infty$, the Lambert W function's argument in (4.8), $\xi \to -1/e$, and $W_{\mathrm{L}}(\xi) \to -1$. Therefore, $C_{\mathrm{opt}} \to 0$.

- When $\rho' \to 0$, $\xi \to \infty$ and the optimum spectral efficiency C_0 grows without bound.

Lemma 4.1 provides the optimum spectral efficiency C_{opt} as a function of the initial operating point (ρ_0', C_0). A change from C_0 to C_{opt} will however affect also the transmission-to-total-processing ratio as described by (4.7).

Theorem 4.1 *Let the transmission-to-total-processing ratio ρ' be defined as in (2.8) on page 21. Energy-wise rate-optimisation of a communication system with a total energy consumption per bit given by (4.5) results in a transmission-to-total-processing ratio*

$$\rho_{\mathrm{opt}}' = \frac{2^{C_{\mathrm{opt}}} - 1}{\ln(2)C_{\mathrm{opt}}2^{C_{\mathrm{opt}}} - (2^{C_{\mathrm{opt}}} - 1)}, \tag{4.10}$$

with the solution for the optimum spectral efficiency

$$C_{\mathrm{opt}} = \frac{1}{\ln(2)} \left[1 + \frac{1}{\rho_{\mathrm{opt}}'} + W_{\mathrm{L}}(\zeta) \right]. \tag{4.11}$$

Here,

$$\zeta = -\left(1 + \frac{1}{\rho_{\mathrm{opt}}'} \right) e^{-\left(1 + \frac{1}{\rho_{\mathrm{opt}}'} \right)}. \tag{4.12}$$

 Proof: See Appendix 4.B.   ■

The results in Theorem 4.1 define a curve in the $(\rho_{\mathrm{opt}}', C_{\mathrm{opt}})$ plane that an energy optimal transmission scheme must follow. Unfortunately, the expressions do not lend themselves to easy interpretation but we can gain insight into the general behaviour by the following corollary.

---

[3]The Lambert W function is described by Corless et al. (1996) and it satisfies $x = W_{\mathrm{L}}(x)\exp(W_{\mathrm{L}}(x))$. For $x > -1/e$ the function is monotonic, increasing, real-valued and $W_{\mathrm{L}}(x) > -1$. The function is sometimes referred to as the Product-Log function or the Omega function.

Corollary 4.1 *An energy-wise rate-optimised system is bounded from below and from above by*

$$\frac{1}{\ln(2)} \leq \rho'_{\text{opt}} C_{\text{opt}} \leq \frac{2}{\ln(2)}. \tag{4.13}$$

Proof: See Appendix 4.C.   ■

Corollary 4.1 shows that the optimum spectral efficiency is approximately inversely proportional to the transmission-to-total-processing ratio.[4]

In Figure 4.1 we show curves corresponding to Lemma 4.1, Theorem 4.1 and Corollary 4.1. We note that small transmission-to-total-processing ratios $\rho'_{\text{opt}} < 0.1$ correspond to excessively high spectral efficiencies $C_{\text{opt}} > 15$. This indicates that the optimum point of operation is in practice out of reach if $\rho'_{\text{opt}} < 0.1$. At the other end of low spectral efficiencies $C_{\text{opt}} < 0.1$, we see that $\rho'_{\text{opt}} > 28$; more than 96 percent of the energy must be transmission energy. The traditional wisdom that low spectral efficiency results in the smallest energy per bit is motivated only when $\rho'_{\text{opt}} \to \infty$.

A guiding rule of thumb. In spite of the idealising assumptions that go into the current calculation, Corollary 4.1 suggest that the simple rule of thumb

$$\rho'_{\text{opt}} C_{\text{opt}} \approx 2 \tag{4.14}$$

can reveal when systems are operating far from the optimum balance, for instance by the use of excessively powerful error correcting codes. We stress that (4.14) is a rule of thumb, not a result that *dictates* what is a good and a bad design.

The Transmission-to-Processing Concentration Effect. By the use of (4.10) in Theorem 4.1 we find that the range $C_{\text{opt}} \in [0.5, 6]$ corresponds to the range $\rho'_{\text{opt}} \in [0.3, 5.5]$; the range of common spectral efficiencies correspond to a fairly narrow range of transmission-to-processing ratios. Figure 4.1 shows that as soon as either processing or transmission costs become dominant, the optimisation will tend to balance them by adjusting the spectral efficiency; in the given examples we go from $\rho' = 10$ to $\rho' = 2.9$, and from $\rho' = 0.01$ to $\rho' = 0.28$. This concentration effect is something we will

---

[4]It should be observed that by optimising the transmission rate according to (4.6) we have *not* found the *capacity per unit energy* as achieved by maximising mutual information under a cost constraint (Verdu, 2002). Rather, we use the Shannon limit – which is the capacity per unit time – as an analytical tool to gain insights into the transmission-processing trade-offs we are currently interested in.
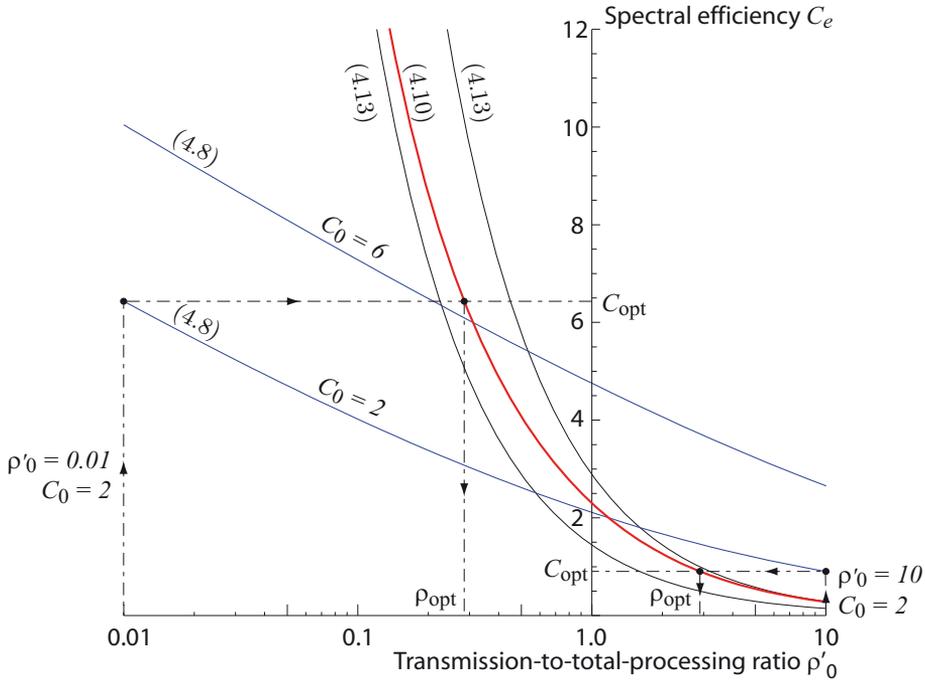
Figure 4.1: The blue solid curves correspond to (4.8) and show the optimum spectral efficiency C_{opt} for two different initial efficiencies $C_0 = 2$ and $C_0 = 6$. The red solid curve shows the optimum transmission-processing trade-off given by (4.10). It is accompanied by the upper and lower limits in (4.13). Two examples are given. First, a system operating with $\rho_0' = 0.01$ and $C_0 = 2$ is severely unbalanced towards excessive processing costs. Following the dash-dotted curve on the left upwards until it intersects (4.8), we find the optimum rate $C_{\mathrm{opt}} = 6.4$ on the vertical axis. The resulting transmission-to-total processing ratio ρ_{opt}' is obtained from the red solid line; $\rho_{\mathrm{opt}}' = 0.28$. Second, a system with $\rho_0' = 10$ operates with excessive transmission costs and the optimisation results in $C_{\mathrm{opt}} = 0.9$ and $\rho_{\mathrm{opt}}' = 2.9$.

observe repeatedly in this thesis: as soon as one type of cost starts to dominate there will be a more energy-efficient strategy that equalises the costs to some extent.[5]

---

[5]There are of course conceivable exceptions to this rule. For example, if ρ' is very large *and* $C_{\mathrm{e}} < C_{\mathrm{opt}}$, the result is an increased ρ'. These cases are however so extreme that they are unlikely to be encountered in practice.

4.1.2   Distance to the Shannon limit

An apparent objection to the above optimisation based on the Shannon limit is that no practical systems will operate at this limit. We can however include a distance v to the limit and replace ρ_0' in (4.6) with $\rho_{0,v}' = v\rho_0'$. In this way we retain the qualitative features of the rate optimisation as long as the minimum received energy per bit $\mathcal{E}_{\mathrm{rec}}$ scales like the Shannon limit in (4.1).

Fading channels

The optimisation results we have now presented applies to the stationary additive white Gaussian noise channel, but let us now briefly discuss the impact of time-variability.

In many sensor network scenarios a quasi-static, or block fading, model is appropriate because of the supposedly low activity (duty cycle) of fixed sensor nodes. For such scenarios we resort to the concept of outage capacity (Goldsmith, 2005). The outage spectral efficiency C_{out} can be achieved with a given probability $1 - P_{\mathrm{out}}$, where P_{out} is the outage probability. For the Nakagami-m channel we can use (2.114) with $\overline{\gamma} = \mathcal{E}_{\mathrm{rec}}/N_0$ and $C' = C_{\mathrm{out}}$ to obtain

$$P_{\mathrm{out}} = 1 - \frac{\Gamma\left(m, m\mathcal{E}_{\mathrm{rec}}^{-1}N_0\frac{2^{C_{\mathrm{out}}}-1}{C_{\mathrm{out}}}\right)}{\Gamma(m)}. \tag{4.15}$$

There is no analytic solution for C_{out}, but let us reason as follows. Define

$$X \equiv \mathcal{E}_{\mathrm{rec}}^{-1}N_0\frac{2^{C_{\mathrm{out}}}-1}{C_{\mathrm{out}}}, \tag{4.16}$$

and rewrite (4.15)

$$P_{\mathrm{out}} = 1 - \frac{\Gamma\left(m, mX\right)}{\Gamma(m)}. \tag{4.17}$$

Solving (4.17) for X, we see that the solution can only depend on m and P_{out},

$$X = f(P_{\mathrm{out}}, m). \tag{4.18}$$

Now we conclude from (4.16) and (4.18) that the minimum received energy per bit $\mathcal{E}_{\mathrm{rec}}$ is proportional to its static channel counterpart in (4.1),

$$\mathcal{E}_{\mathrm{rec}} = f(P_{\mathrm{out}}, m)N_0\frac{2^{C_{\mathrm{out}}}-1}{C_{\mathrm{out}}}. \tag{4.19}$$

The previous rate optimisation results for the static channel are thus applicable also to the quasi-static fading channel. The modification lies in that

the distance $\upsilon = f(P_{\text{out}}, m)$ to the Shannon limit must be accounted for in (4.6).

4.1.3 Practical relevance of idealised rate optimisation

On the one hand, the above calculations highlight the importance of the transmission-processing trade-off and shows the possible impact of transmission optimisation. They show that if spectral efficiency is decreased, supposedly to save transmission energy, the penalty in terms of circuit processing costs can be large; systems operating at very low spectral efficiencies should have very small processing costs. This confirms the intuition that low spectral efficiency is beneficial energy-wise only when the transmission energy dominates, typically at large distances.

On the other hand, the model ignores several real-world considerations needed for a fair assessment. The above results share with Shannon's theory the attractive feature of independence from practical implementation details, but this is presently also the limitation of the model; implementation details will matter. It is some of these practical details that the rest of this chapter concerns. First of all, the benefits from optimised transmission rates presupposes the use of power control, that is a transmitter adapting its output power level to the current conditions. This adaption requires feedback and we study that particular transmission-processing trade-off in Section 4.2. Following that study, we turn to error correcting codes and their possible energy savings in Section 4.3. Observe that the introduction of such codes can only decrease the rate. The final topic is adaptive modulation in Section 4.4, where we focus on quadrature amplitude modulation (QAM).

4.2 Power control or fixed link margins

Considering the energy efficiency of wireless transmissions, one of the first things that comes to mind is that the transmission power should be adjusted to the present conditions so that no extraneous energy is radiated in vain. It seems obvious that if sensor nodes adjust their transmit power level to the current channel conditions, instead of applying a fixed link safety margin, they will save energy. We must however include the energy cost of the necessary feedback from the receiver node, and it turns out that it can in fact be less energy-consuming to apply a fixed link margin than to perform power control. Johansson et al. (2006) analyse fast and slow power control over fading channels and find that fixed margins are preferable over the propagation losses expected in most wireless sensor networks. Here we extend the analysis

and provide insights obtained by our use of the transmission-to-processing ratio ρ.

Channel inversion. Our analysis is carried out for unconstrained and constrained channel inversion, that is when the transmitter adapts its output power with the aim of a constant signal-to-noise ratio at the receiver. Channel inversion relies on the feedback of accurate channel state information, which under our assumption of a single-tap channel amounts to the channel power gain x. From a channel-capacity viewpoint, the channel inversion scheme is known to be sub-optimal, and the unconstrained version can even result in zero capacity for Rayleigh fading. On the other hand, the results depend strongly on the degree of fading and it is also seen that so-called truncated schemes perform well even in Rayleigh fading (Goldsmith and Varaiya, 1997). As we are here only concerned with power adaptivity, no adaptive modulation or coding is assumed, channel inversion is the straight-forward alternative.

Nakagami-m fading, DBPSK and bit error rates. We here concentrate on the fading effects x_f for which we have assigned a gamma probability distribution through our Nakagami fading model, see Assumption 2.6 on page 29. We assume that nothing but the channel varies[6] and can therefore express the received signal-to-noise ratio per bit

$$\gamma_{\text{rec}} = K x_f \mathcal{E}_{\text{rad}}, \tag{4.20}$$

where K is a system and propagation dependent constant whose exact value is of no present interest as it will cancel in the subsequent calculations. The channel fading gain x_f is assigned the distribution

$$p(x_f | m, I) = \frac{m^m}{\Gamma(m)} x_f^{m-1} e^{-m x_f}, \tag{4.21}$$

where m is the Nakagami-m fading figure. For simplicity we will assume, when it is required, that differential binary phase shift keying (DBPSK) is used.[7]

---

[6]It would in principle be straightforward to include all types of variations, not only the fading. Variations in interference and noise at the receiver are examples of phenomena that we could include.

[7]Apart from a constant offset in signal-to-noise ratio per bit, the error behaviour – bit error or outage – is very similar for most modulation schemes. The offset plays a minor role as we are using relative measures such as the transmission-to-processing ratio; it is

4.2.1 Fixed link margin approach

If we adopt the link margin (LM) approach, in which the smallest fixed output power that achieves the given performance criteria is chosen, the receiver will see a signal-to-noise ratio per bit γ_{LM} that follows the fading gain x_{f}. Due to the channel variations the average signal-to-noise ratio per bit $\overline{\gamma}_{\text{LM}}$ must be larger than for a static channel.

Definition 4.1 *The link (fade) margin M_{L} is defined with respect to a static Gaussian channel,*

$$M_{\text{L}} \equiv \frac{\mathcal{E}_{\text{LM}}}{\mathcal{E}_{\text{sta}}} = \frac{\overline{\gamma}_{\text{LM}}}{\gamma_{\text{sta}}}, \tag{4.22}$$

where $\overline{\gamma}_{\text{LM}}$ and γ_{sta} are the signal-to-noise ratios per bit required to meet specified performance goals for fading and static channels respectively.[8] ∎

The size of the link margin is determined by the degree of fading and the performance requirements. To exemplify, consider differential binary phase shift keying for which the average bit error rates in static and Nakagami-m fading channels are given by (2.106) and (2.107) respectively. The link margin required for a bit error rate $\overline{B}$ becomes

$$M_{\text{L,DBPSK}} = m\frac{1 - (2\overline{B})^{-1/m}}{\ln(2\overline{B})}, \tag{4.23}$$

which is strongly affected by the values of $\overline{B}$ and the fading figure m; deep and frequent fading dips require a large safety margin if errors are to be avoided. For instance, $\overline{B} = 10-3$ and $m = 2$ results in $M_{\text{L,DBPSK}} = 6.87$, while $\overline{B} = 10-6$ and $m = 1$ results in $M_{\text{L,DBPSK}} = 36.2 \cdot 10^3$.

4.2.2 Ideal channel inversion through power control

Let us for a moment pretend that the transmitter has no power limitation, which in our present notation means that the radiated energy per bit $\mathcal{E}_{\text{rad}}$ has no upper bound. Given that the transmitter also has perfect channel state information, that is to say it knows the exact value of the fading gain x_{f} at each instant, it can apply the ideal channel inversion (ICI) scheme

$$\mathcal{E}_{\text{ICI}} = \frac{\mathcal{E}_{\text{ICI,0}}}{x_{\text{f}}} \tag{4.24}$$

---

the error slope that matters the most. For the quantitative analysis we use the criterion of average bit error rate, but point out the connection to the combined use of outage probability and bit error rate.

[8]We here refer only to the *radiated* energies per bit $\mathcal{E}_{\text{rad,LM}}$ and $\mathcal{E}_{\text{rad,sta}}$ by the simplified notation $\mathcal{E}_{\text{LM}}$ and $\mathcal{E}_{\text{sta}}$ respectively.

to provide the receiver with a constant signal-to-noise ratio γ_{ICI}. Here $\mathcal{E}_{\text{ICI},0}$ is a constant which is to be chosen to satisfy the performance requirement; a constant received signal-to-noise ratio per bit $\gamma_{\text{ICI}} = \gamma_{\text{sta}}$ which results in the targeted bit error rate B. In the long-term energy perspective it is however the average energy per bit that is of interest.

Theorem 4.2 *Consider the ideal channel inversion scheme (4.24). The average radiated energy per bit, $\overline{\mathcal{E}}_{\text{ICI}}$, is for Nakagami-m channel inversion given by*

$$\overline{\mathcal{E}}_{\text{ICI}} = \frac{m}{m-1}\mathcal{E}_{\text{sta}}, \qquad m > 1, \tag{4.25}$$

where $\mathcal{E}_{\text{sta}}$ is the energy per bit required for a static channel. For all fading figures $m \leq 1$ the average energy is infinite.

Proof: See Appendix 4.D. ∎

Corollary 4.2 *The radiated-energy gain G_{ICI} relative to the fixed link margin scheme is for ideal channel inversion*

$$G_{\text{ICI}} \equiv \frac{\mathcal{E}_{\text{LM}}}{\overline{\mathcal{E}}_{\text{ICI}}} = \frac{m-1}{m}M_{\text{L}}. \tag{4.26}$$

Proof: The result in (4.26) follows readily from Theorem 4.2 and Definition 4.1. ∎

Figure 4.2 shows the energy gain G_{ICI} for differential binary phase shift keying. It is evident that large gains are achievable by avoiding the waste of energy that a fixed link margin scheme causes during good channel conditions. However, we also see that the severest conditions are too costly to invert and the gain decreases to zero for small m.

The problem of infinite average output power in the Rayleigh case, evidenced in Theorem 4.2, is well recognised in the literature, see for instance the work by Goldsmith and Varaiya (1997) and references therein. Loosely speaking, the probability that $x_{\text{f}} = 0$ is not small enough in the Rayleigh case to let us disregard from that possibility, and since it takes (mathematically) an infinite amount of power to invert a channel zero we also face an infinite average power. There exists an entirely practical and unavoidable radio limitation which alleviates our infinite-power concerns: All radios have a maximum output power level and the infinite peaks are just mathematical artifacts.
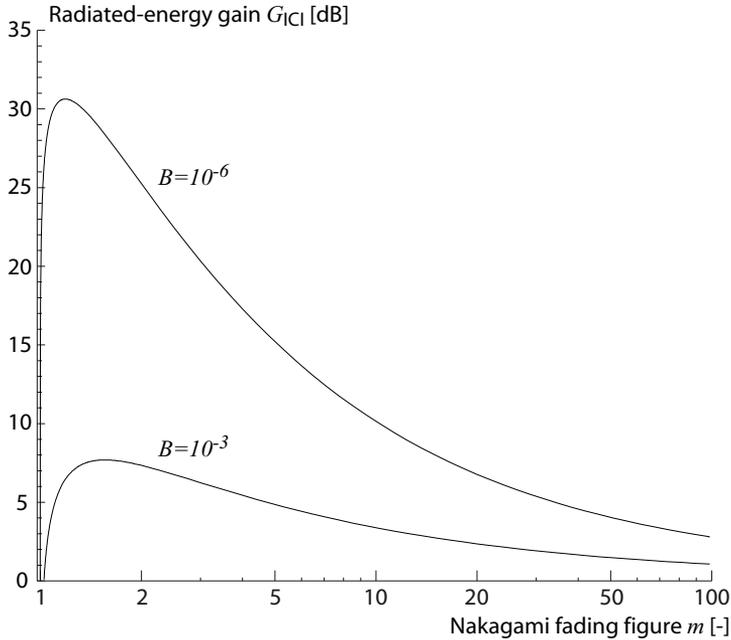
Figure 4.2: Transmission energy gain G_{ICI} achieved by the ideal inversion scheme for differential binary phase shift keying. Over a Rayleigh fading channel, $m = 1$, the transmitter spends so much effort inverting the deepest fading deeps that the gain becomes zero ($-\infty$ dB). The large difference between the two bit error rates is due to the size of the link margin M_{L} which is much larger for $B = 10^{-6}$ than for $B = 10^{-3}$.

4.2.3 Power-limited (truncated) channel inversion

Although the obvious reason to invoke a power limitation is that it is practically inevitable, one should note that the limitation is actually beneficial in the energy context we are now considering. The expected energy cost of inverting a Rayleigh channel is infinite, but since the link margin approach can meet the presently used performance requirement at a constant and finite energy per bit one must conclude that it can be better to leave some fluctuations unconsidered – even at the price of wasting some energy when the channel is good.[9]

Given that the transmitter manages a maximum radiated energy per bit

---

[9]This resembles the truncated channel inversion considered by Goldsmith and Varaiya (1997) in their channel capacity analysis; no transmissions take place below a certain channel threshold and enough energy is thereby saved to make the capacity in Rayleigh fading non-zero, which it otherwise is (under the assumption of receiver channel state information only).

$\mathcal{E}_{\max}$, which corresponds to its maximum output power, we obtain a power-limited channel inversion (CI) scheme that operates according to the rule

$$\mathcal{E}_{\mathrm{CI}} = \min\left\{\mathcal{E}_{\max}, \frac{\mathcal{E}_{\mathrm{CI},0}}{x_{\mathrm{f}}}\right\}, \qquad (4.27)$$

where $\mathcal{E}_{\mathrm{sta}} \leq \mathcal{E}_{\mathrm{CI},0} \leq \mathcal{E}_{\mathrm{LM}}$ is a constant that must be chosen to satisfy the performance requirement. Naturally, if the comparison with the link margin approach is to be meaningful we require that $\mathcal{E}_{\max} \geq \mathcal{E}_{\mathrm{LM}}$.

Lemma 4.3 *Let the bit error rate over a static channel be denoted $B_{\mathrm{sta}}(\gamma)$ where γ is the received signal-to-noise ratio per bit. Assuming perfect receiver channel state information, the transmission constant $\mathcal{E}_{\mathrm{CI},0}$ in the truncated channel inversion scheme (4.27) is found by solving*

$$\overline{B}_{\mathrm{CI}} = P\left(x_{\mathrm{f}} \geq \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\max}}|I\right) B_{\mathrm{sta}}\left(\overline{\gamma}_{\mathrm{LM}}\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}}\right)$$
$$+ \int_0^{\overline{\gamma}_{\mathrm{LM}}\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}}} p\left(\gamma|\gamma = \overline{\gamma}_{\mathrm{LM}}\frac{\mathcal{E}_{\max}}{\mathcal{E}_{\mathrm{LM}}}x_{\mathrm{f}}, I\right) B_{\mathrm{sta}}(\gamma)\mathrm{d}\gamma \qquad (4.28)$$

for $\mathcal{E}_{\mathrm{CI},0}$. Here, $\overline{\gamma}_{\mathrm{LM}}$ and $\mathcal{E}_{\mathrm{LM}}$ respectively denote the received average signal-to-noise ratio per bit and the radiated energy per bit required to achieve the bit error rate $\overline{B}_{\mathrm{LM}} = \overline{B}_{\mathrm{CI}}$ with a fixed link margin (LM). For differential binary phase shift keying in Nakagami-m fading the average bit error rate is given by

$$\overline{B}_{\mathrm{CI}} = \frac{\Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\max}}\right)}{\Gamma(m)} \cdot \frac{1}{2}e^{-\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}}\gamma_{\mathrm{LM}}}$$
$$+ \frac{\Gamma(m) - \Gamma\left(m, \mathcal{E}_{\mathrm{CI},0}\left(\frac{m}{\mathcal{E}_{\max}} + \frac{\gamma_{\mathrm{LM}}}{\mathcal{E}_{\mathrm{LM}}}\right)\right)}{\Gamma(m)} \cdot \frac{1}{2}\left(\frac{m\mathcal{E}_{\mathrm{LM}}}{m\mathcal{E}_{\mathrm{LM}} + \mathcal{E}_{\max}\gamma_{\mathrm{LM}}}\right)^m.$$
$$\qquad (4.29)$$

> *Proof:* See Appendix 4.E.   ■

The average bit error rate $\overline{B}_{\mathrm{CI}}$ in (4.28) consists of two contributions. First, the static-channel bit error rate weighted by the probability that the channel can be inverted. Second, the bit error rate averaged over the remaining channel fluctuations.

Theorem 4.3 *Consider the power limited (truncated) channel inversion in (4.27). The average radiated energy per bit, $\overline{\mathcal{E}}_{\mathrm{CI}}$, required for power limited Nakagami-m channel inversion is given by*

$$\overline{\mathcal{E}}_{\mathrm{CI}} = \mathcal{E}_{\mathrm{max}} \frac{\Gamma\left(m\right) - \Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)} + m\mathcal{E}_{\mathrm{CI},0} \frac{\Gamma\left(m-1, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)}, \quad (4.30)$$

where $\mathcal{E}_{\mathrm{CI},0}$ is determined by solving (4.28) for Nakagami-m fading (or by solving (4.29) if differential binary phase shift keying is used).

Proof: See Appendix 4.F. ■

The use of truncated channel inversion will result in a reduction in the average radiated energy per bit, with respect to the fixed link margin (LM) approach. How large the reduction is is shown by the following Corollary.

Corollary 4.3 *The achieved radiated-energy gain, with respect to a fixed link margin (LM), is for truncated channel inversion*

$$G_{\mathrm{CI}} \equiv \frac{\mathcal{E}_{\mathrm{LM}}}{\overline{\mathcal{E}}_{\mathrm{CI}}}$$

$$= \frac{\mathcal{E}_{\mathrm{LM}}}{\mathcal{E}_{\mathrm{max}}} \left(\frac{\Gamma\left(m\right) - \Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)} + m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}} \frac{\Gamma\left(m-1, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)} \right)^{-1}.$$

$$(4.31)$$

Proof: Straightforward from Theorem 4.3. ■

The gain in (4.31) is difficult to interpret due to the nonlinear and nontrivial influence of the modulation and bit error rate on the $\mathcal{E}_{\mathrm{CI},0}$ parameter. Essentially, the result of power-limited channel inversion is a received signal-to-noise ratio which is mostly constant and slightly larger than for a completely static channel in order to compensate for the occasional un-invertible fading dip. However, the remaining dips are shallower than for the fixed link margin approach. In Figure 4.3 we illustrate the result of truncated channel inversion on a measured channel.

An outage criterion amounts to truncation

For some applications it may be more reasonable to use an outage criterion accepting a certain probability for lost packets, given that the channel otherwise is good enough to support successful packet reception. Referring to

Energy per bit relative to the static channel case [dB]

Received signal-to-noise ratio per bit [dB]

Time or position [arbitrary units]
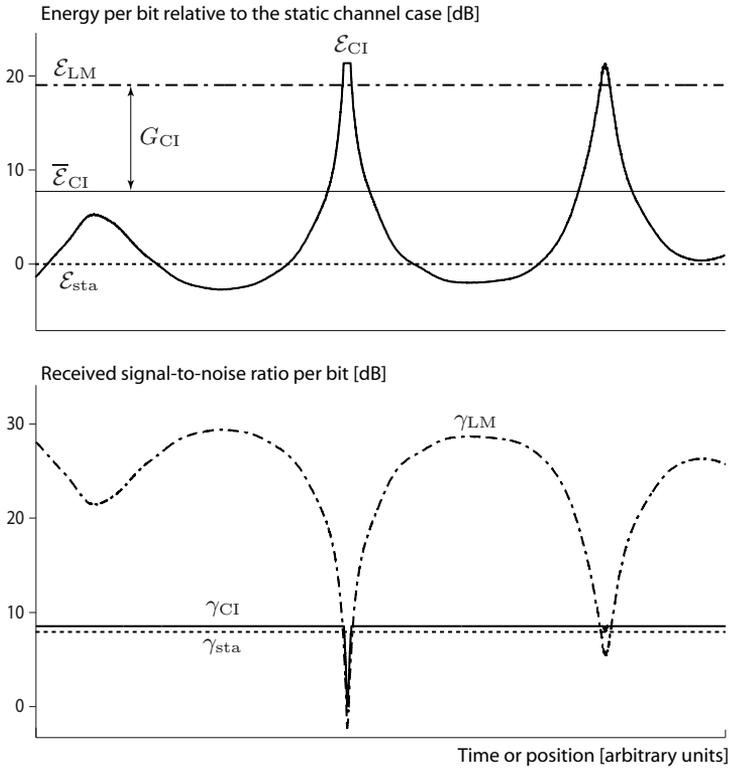
Figure 4.3: Power-limited (truncated) channel inversion. Observe that the received signal-to-noise ratio per bit is larger than for the static channel, even during the periods the transmitter manages to keep it constant. This is to compensate for the residual fading when the transmitter saturates its power amplifier. By contrast to the link margin scheme, the power control scheme utilises good channel conditions to save energy; it only exceeds the link margin level during the deepest fades.

Figure 4.3, we see that since an outage criterion accepts a certain amount of fading dips that cause packet loss it will not force the transmitter to invert all dips. Consequently, the outage criterion will amount to a certain level of transmit power limitation, or truncation, which is not imposed by the maximum output power. Combining outage and bit error rate criteria for the fading channel thus corresponds to a an artificial output power limit and a bit error rate criterion.

4.2.4 Energy saving through truncated channel inversion

The price paid for adapting the transmission power is the cost of feeding back channel state information to the transmitter. We omit the details regarding the exact procedure and employ the simple model used by Johansson et al. (2006) which collects all additional overhead in r, the fraction of extra overhead transmissions required. Put differently, the impact of the feedback overhead is modelled to be energy-wise equivalent to sending $1 + r$ bits per bit sent in the fixed power scheme.[10] The numerical results we present here pertain to differential binary phase shift keying, but we believe that they are fairly general in applicability.

The power control scheme saves a factor G_{CI} in radiated energy per bit $\mathcal{E}_{\text{rad}}$, see (4.31), but due to the degradation in amplifier efficiency the actual gain in transmission energy $\mathcal{E}_{\text{T}}$ is less. According to the Mikami model in (2.14), we have

$$\mathcal{E}_{\text{T,CI}} = \frac{\mathcal{E}_{\text{T,LM}}}{G_{\text{CI}}^{1/g}}, \tag{4.32}$$

where $g > 1$ is the degradation exponent. Channel inversion (CI) will save a fraction w of energy with respect to the link margin (LM) approach;

$$
\begin{aligned}
w &= \frac{\mathcal{E}_{\text{tot,LM}} - \mathcal{E}_{\text{tot,CI}}}{\mathcal{E}_{\text{tot,LM}}} \\
&= 1 - \frac{(1+r)\left(\mathcal{E}_{\text{RP}} + \frac{\mathcal{E}_{\text{T,LM}}}{G_{\text{CI}}^{1/g}}\right)}{\mathcal{E}_{\text{RP}} + \mathcal{E}_{\text{T,LM}}} \\
&= 1 - \frac{(1+r)\left(1 + \frac{\rho'_{\text{LM}}}{G_{\text{CI}}^{1/g}}\right)}{1 + \rho'_{\text{LM}}} \\
&= \frac{\rho'_{\text{LM}}\left(1 - \frac{1+r}{G_{\text{CI}}^{1/g}}\right) - r}{1 + \rho'_{\text{LM}}},
\end{aligned}
\tag{4.33}
$$

where $\mathcal{E}_{\text{RP}} = \mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr}}$ is the total processing energy per bit, see (2.5), and ρ'_{LM} is the transmission-to-total-processing ratio, see (2.8). As we previously

---

[10]This simplified model neglects that, in addition to transmission and reception of extra bits, there will be energy expended by node circuitry, most notably local oscillators and frequency synthesisers, during the switches between receive and transmit modes. Especially if the adaption must be conducted on a shorter time scale than the packet duration. Moreover there will be processing costs related to the adaption. We lump all this into an equivalent fraction r of overhead. Johansson et al. (2006) assumed two-way communication and thereby simplified the analysis somewhat.

mentioned, the size of the gain G_{CI} is determined by the degree of fading and the performance requirement; the harder the conditions and requirements, the larger the gain. Taking our uncertainty regarding the fading figure m into account, we take the expected energy saving

$$E(w|\overline{B}, g, I) = \frac{\rho'_{\text{LM}}\left(1 - (1+r)E\left(\frac{1}{G_{\text{CI}}^{1/g}}\Big|\overline{B}, g, I\right)\right) - r}{1 + \rho'_{\text{LM}}}, \qquad (4.34)$$

as our mean-square optimal prediction given a target bit error rate $\overline{B}$ and an amplifier degradation exponent g. The expectation is presently found by (numerical) averaging over our probability distribution for m, see (2.85). This prior for m has the median $m_{\text{med}} = 1.88$, the 10th percentile $m_{10} = 1.10$ and the 90th percentile $m_{90} = 8.83$. By the use of these values and (4.33) we find the corresponding median and percentiles for w.

In Figure 4.4 and Figure 4.5 we show the saving w for target bit error rates $\overline{B} = 10^{-3}$ and $\overline{B} = 10^{-6}$ respectively, together with examples of maximum transmission-to-total-processing ratios from Section 2.2.2. It is evident that the bit error rate has significant impact on the outcome. This also holds true for the feedback fraction r; if kept small the benefits from power control increase.

Maximum allowed feedback fraction

Let us now study the maximum allowable feedback fraction r, since this parameter plays a key role in the present trade-off between transmission energy and processing energy. For ease of exposition we focus on the expected energy saving in (4.34), and omit the percentiles for w. To attain the expected energy saving $E(w|B, g, I) = \overline{w}$ we must achieve the expected radiated-energy gain $\overline{G_{\text{CI}}^{-1/g}} = E(G_{\text{CI}}^{-1/g}|B, g, I)$ using a feedback fraction of at most

$$r_{\text{max}} = \frac{\rho'_{\text{LM}}\left(1 - \overline{w} - \overline{G_{\text{CI}}^{-1/g}}\right) - \overline{w}}{1 + \rho'_{\text{LM}}\overline{G_{\text{CI}}^{-1/g}}}, \qquad (4.35)$$

found by solving (4.34) for r. We plot the feedback fraction together with node characteristics in Figure 4.6. Note the rapid increase in r_{max} with ρ'_{LM}; substantial feedback is allowed for $\rho'_{\text{LM}} > 1$, that is when transmission costs starts dominating processing costs.
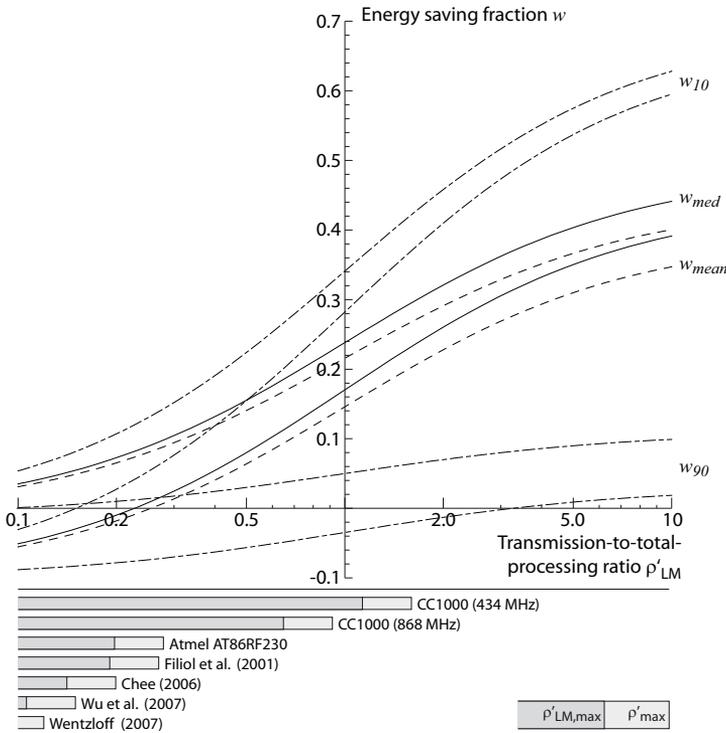
Figure 4.4: Saved fraction w of energy for a target bit error rate $B = 10^{-3}$. The upper curve of each type (solid, dashed, dash-dotted) corresponds to a feedback fraction $r = 0.01$, while the lower curve corresponds to $r = 0.1$. It is assumed that $\mathcal{E}_{\mathrm{rad,max}} = 2\mathcal{E}_{\mathrm{LM}}$ so that truncated inversion is feasible – see the dark grey area of the bars (described closer in item 1. on page 111) – and that $g = 2$.

Observations

From Figure 4.4, Figure 4.5 and Figure 4.6 we observe the following:

1. It is required that $\mathcal{E}_{\mathrm{rad,max}} > \mathcal{E}_{\mathrm{rad,LM}}$ for the inversion scheme to work; we have used $\mathcal{E}_{\mathrm{rad,max}} = 2\mathcal{E}_{\mathrm{rad,LM}}$ in our calculations. Therefore the node specific maximum transmission-to-total-processing ratios ρ'_{max} are not fully applicable. They should be lowered by a factor $2^{1/g} = 1.41$ to be representative, (we have indicated this as dark grey bars in the figures).

2. The saving w increases with the transmission-to-processing ratio, but so does the uncertainty regarding this gain (stemming from the uncertainty regarding the degree of fading). At any rate, it seems that
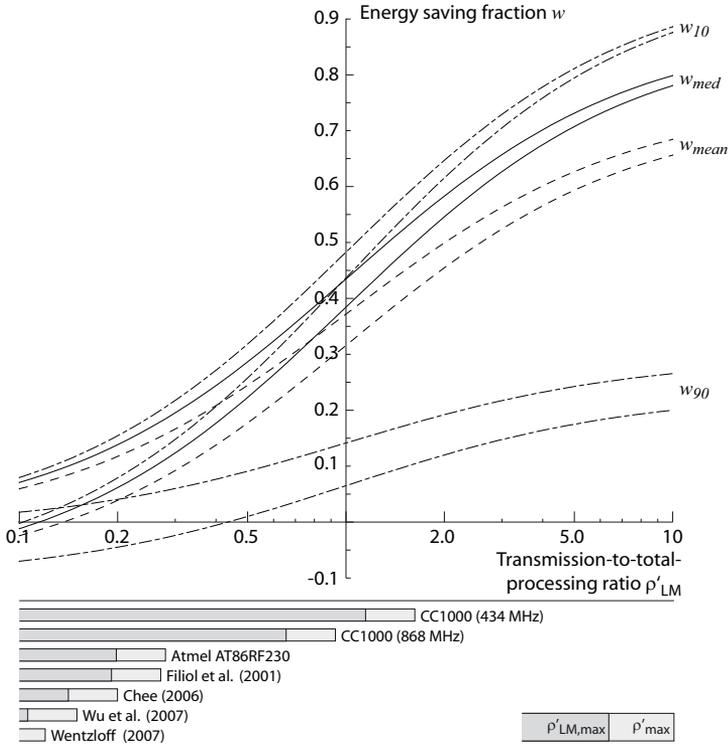
Figure 4.5: Saved fraction w of energy for a target bit error rate $B = 10^{-6}$. The upper curve of each type (solid, dashed, dash-dotted) corresponds to a feedback fraction $r = 0.01$, while the lower curve corresponds to $r = 0.1$. It is assumed that $\mathcal{E}_{rad,max} = 2\mathcal{E}_{LM}$ so that truncated inversion is feasible – see the dark grey area of the bars (described closer in item 1. on page 111) – and that $g = 2$.

power control becomes strongly motivated for $\rho'_{LM} > 0.25$ if $\overline{B} = 10^{-3}$, and for $\rho'_{LM} > 0.15$ if $\overline{B} = 10^{-6}$.

3. If we want to *save* energy – not just break even – we see a significant threshold effect in Figure 4.6. No feedback should be used below a threshold transmission-to-total-processing ratio. Above the threshold the allowed amount of feedback increases quite rapidly and $r = 0.1$ is no problem at $\rho'_{LM} = 1$.

4. Not all exemplified radios would benefit from power control with certainty, although it goes without saying that the CC1000 would if operating close to its ρ'_{max}. The likely energy savings for the other radios are somewhat more modest but still positive for the most part.
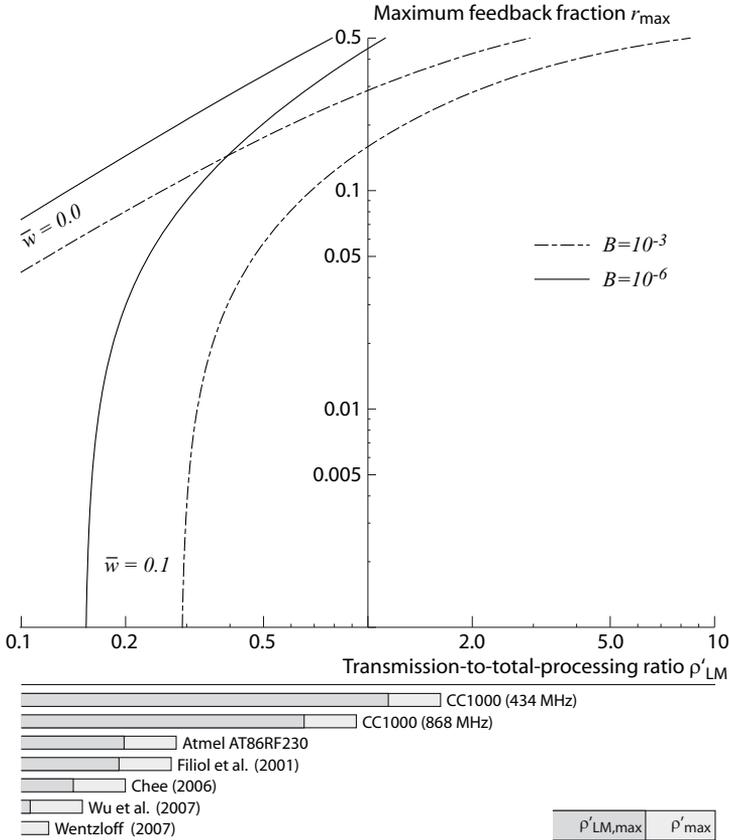
Figure 4.6: Maximum feedback fraction allowed according to (4.35) for expected energy savings of $\overline{w} = 0$ and $\overline{w} = 0.1$ respectively.

5. The maximum energy *loss* from power control is simply r; in case no transmission gain is attainable we spend the overhead in vain.

4.2.5 Slow and fast power control

If the channel fading is quasi-static or slow, see Definition 2.1 on page 30, the power control overhead can easily be incorporated into an existing packet structure. How large r becomes is of course highly dependent on the lengths of the packets, but since some applications may use very short packets – for instance reporting a single sensor reading – the power control information can be of non-negligible size even for slow and quasi-static channels. The use of fast power control introduces significantly larger amounts of overhead due to the need to shorten packets to allow the transmitter to follow the

fading. Then not only power control bits are introduced, but also the cost
of transmit-receive switching and increased synchronisation overhead.

4.2.6   Summary power control

We believe that power control should be employed under the following con-
ditions:

- Only invert slow channel variations so that a small overhead cost can
  be achieved.

- The radios' maximum transmission-to-processing ratio $\rho'_{\max} > 0.3$.[11]

It will probably suffice with a coarse power control scheme that operates with
few feedback bits – maybe just a single bit – and a few power level settings
(this amounts to a combination of link margins and coarse power control).
The important thing is to avoid excessive power levels during good channel
conditions.

4.3   Error correcting codes in static channels

The use of error correcting codes can be viewed as a limited variant of rate
adaption in the sense that we can only trade increased circuit processing
energy for reduced transmission energy. Assuming that the system is initially
uncoded, error correcting codes can reduce the transmit energy, not the
processing energy.

Howard et al. (2006) have studied the energy consumption of coding and
decoding, and compared them to the transmission costs. They conclude
that codes save energy already at very short distances, but they have not ac-
counted for the increase in transmission time, only the energy consumed by
the coder and decoder. Similarly, Zhong et al. (2005) neglect the increased
transmission time when they propose an asymmetric one-hop network struc-
ture based on powerful error correcting codes. Their idea is to exploit the
fact that the encoding process is much less energy consuming than the de-
coding process and therefore use powerful codes when transmitting from

---

[11]We have hereby revealed a "hidden assumption" in our own published study where
we did not work with transmission-to-processing ratios explicitly (Johansson et al., 2006).
However, we implicitly assumed – by the use of energy consumption numbers published
by others – that the maximum transmission-to-processing ratio was considerably smaller
than one. In the case then considered, our conclusion had to be that power control was
inefficient energy wise.

the simple sensor nodes to the central node.[12] A. Chandrakasan has led research which includes error correcting codes and his group has suggested the concept of "energy-agile coding" which amounts to adapting the code rate to achieve the lowest overall energy consumption; see the work by Min et al. (2002) and Shih et al. (2004). Their study of error correcting codes is based on convolutional codes at a fixed propagation loss of 70 dB and slow Rayleigh fading, and their conclusion is that coding – more precisely a rate 3/4 convolutional code – in this case should be applied if the target bit error rate is below $B = 10^{-5}$. The special case is illustrative but lacks generality regarding the trade-off between transmission energy and processing energy. But more importantly, the calculation does not include the receiver's total processing costs but only the decoding circuitry.[13]

Sartipi and Fekri (2004) include processing costs for both the coding and the increased transmission time in their study, but use a fixed transmission-to-processing ratio when they compare conventional BCH codes (Bose, Chaudhuri and Hocquenghem) with low density parity check codes. Zorzi and Rao (2004) and Tralli (2005) both study combined multi-hop and error correcting codes – which we will do in Chapter 6 – and it is not possible to isolate the possible energy gains from the coding in these studies.

We will in this section study error correcting codes with special focus on the transmission-to-total-processing ratio ρ' to gain insights into the fundamental energy trade-off.

4.3.1 Energy consumption for coded transmissions

Let us here sort out how the use of error correcting codes affects the total energy consumption, consisting of transmission energy, radio processing energy and code processing energy. By the use of an error correcting code with

$$\text{code rate } R_{\mathrm{c}} \leq 1 \tag{4.36}$$

and

$$\text{coding gain } G_{\mathrm{c}} \equiv \frac{\gamma_{\mathrm{u}}}{\gamma_{\mathrm{c}}}, \tag{4.37}$$

---

[12]Zhong et al. (2005) have focused on the possibility to implement a one-hop network structure and compared it with a multihop structure. The possible energy savings due to coding has not been studied for a single hop as we do here.

[13]The nodes used consume 82 mW during transmission with an output power of 1 mW, and 180 mW during reception (Shih et al., 2004, p. 82). At a rate of 1 Mbit/s this amounts to a total processing cost of 262 nJ/bit, but the displayed results start at roughly 82 nJ/bit (Shih et al., 2004, Figure 13). This indicates that only the transmitter's energy consumption has been accounted for.

the required signal-to-noise ratio per bit γ at the receiver can be reduced from the uncoded γ_u to the coded γ_c, where subscripts u and c denote uncoded and coded schemes respectively (observe that γ_c is the signal-to-noise ratio per *information bit*, just like γ_u). According to the Mikami power amplifier degradation model (2.14), which states that the actual power consumption decreases less than the radiated power, we achieve the following reduction of the transmission energy per bit:

$$\mathcal{E}_{\text{T,c}} = \frac{\mathcal{E}_{\text{T,u}}}{G_{\text{c}}^{1/g}}. \tag{4.38}$$

The efficiency degradation effect is quantified by the degradation exponent g, which is a property of the transmitter's power amplifier.

The total energy per bit consumed by the coded scheme is affected by an increased transmission time caused by the code redundancy $1/R_c$, effectively increasing the processing energy. Also taking the per-bit energy consumption $\mathcal{E}_{\text{CD}}$ of the coding and decoding processes into account, we find by use of (4.38) that

$$\mathcal{E}_{\text{tot,u}} = \mathcal{E}_{\text{RP}} + \mathcal{E}_{\text{T,u}} \tag{4.39}$$

$$\mathcal{E}_{\text{tot,c}} = \frac{\mathcal{E}_{\text{RP}}}{R_{\text{c}}} + \frac{\mathcal{E}_{\text{T,u}}}{G_{\text{c}}^{1/g}} + \mathcal{E}_{\text{CD}}, \tag{4.40}$$

where $\mathcal{E}_{\text{RP}}$ is the total processing energy per bit as given by (2.5). By normalising (4.39) and (4.40) with respect to the radio processing energy $\mathcal{E}_{\text{RP}}$ we obtain

$$\breve{\mathcal{E}}_{\text{tot,u}} = 1 + \rho_{\text{u}}', \tag{4.41}$$

$$\breve{\mathcal{E}}_{\text{tot,c}} = \frac{1}{R_{\text{c}}} + \rho_{\text{u}}' \frac{1}{G_{\text{c}}^{1/g}} + \breve{\mathcal{E}}_{\text{CD}}, \tag{4.42}$$

where ρ_{u}' is the transmission-to-total-processing ratio of the uncoded transmission ($R = 1$), see (2.8). Above, we have suppressed the parameters dependence of G_{c} and $\mathcal{E}_{\text{CD}}$ on R_{c} for notational convenience.

Optimisation of the code rate R_{c} to achieve the minimum energy consumption (4.42) would now be straight-forward had we explicit models of $G_{\text{c}}(R_{\text{c}})$ and $\breve{\mathcal{E}}_{\text{CD}}(R_{\text{c}})$. Unfortunately there are, to the best of our knowledge, no such general models with adequate accuracy. Therefore, we will first make reasonable simplifications of $G_{\text{c}}(R_{\text{c}})$ and $\breve{\mathcal{E}}_{\text{CD}}(R_{\text{c}})$ that allow us to gain insights into the energy-optimal choice of code rate. Second, we specialise our study to adaptive coding within a class of BCH codes (Bose, Chaudhuri and Hocquenghem).

Neglected coding and decoding energy consumptions. Howard et al. (2006) assess several practical coding schemes with respect to the trade-off between the coding gain and the energy cost of decoding, and based on their analysis we make the following simplifying approximation.

Assumption 4.1 *The processing energy consumption for coding and decoding is negligible in relation to other processing energy costs,*

$$\mathcal{E}_{CD} = 0. \tag{4.43}$$

∎

According to Howard et al. (2006), the energy consumption can be as small as tens of pJ per bit, and this can be compared to ultra-low power radios, such as the one developed by Chee (2006), which at best achieve communication energies of a couple of nJ.[14] Also the work by Shih et al. (2004, Figure 12) supports this simplification.

An upper bound on the coding gain. We make use of an upper bound $G_{c,max}$ applicable for block codes given in (Anderson and Svensson, 2002, p. 83),

$$G_c \leq G_{c,max} = R_c(t+1), \tag{4.44}$$

where t is the number of errors the code can correct (of course, for any useful code we have $R_c(t+1) > 1$).[15] The error-correcting capability t is dependent on R_c, $t = t(R_c)$, so we have not arrived at a analytical expression which we can optimise directly. However, we can now more easily calculate a conservative limit for specific (R_c, t)-codes because values of R_c and t are readily available in the literature for many codes.

EXAMPLE 4.1 Coding Gains From BCH Codes

A common and well-known set of block codes is due to Bose, Chaudhuri and Hocquenghem, and are consequently termed BCH codes. Proakis (2001,

---

[14]Howard et al. (2006) do not include the energy cost of the increased transmission time – due to radio baseband processing of the redundancy bits of the code – and therefore reach the conclusion that error correcting codes are motivated already at very short distances, on the order of metres. Our present study concerns precisely the trade-off between the extra radio processing of redundant code bits and the transmission energy gain.

[15]For convolutional codes, the upper bound is given by $R_c d_{free}$ where d_{free} is the minimum free distance of the code (Proakis, 2001, p. 510). Both this limit and the one in (4.44) are derived for the binary symmetric channel with Gaussian noise.

Table 8.1-6) give t and R_c for a range of different BCH codes with block lengths from 7 to 255. We find that the best of these codes in terms of $G_{c,max}$ is the (255,115) code which has $G_{c,max} = 9.92$. Application of this code will consequently give at most a 10 dB reduction in transmission energy per bit. This limit is however quite optimistic with respect to practical cases. Let us illustrate by comparing the actual gains of two other BCH codes with their upper bounds. The actual gains (Goldsmith, 2005, Figure 8.4) pertain to bit error rates $B = 10^{-4}$ and $B = 10^{-6}$ for binary phase shift keying and are displayed in Table 4.1. The fact that the

Table 4.1: Coding gains for the (127,36) and (127,64) BCH codes.

	$G_c(B = 10^{-4})$	$G_c(B = 10^{-6})$	$G_{c,max}$
(127,36)	1.46	2.15	4.54
(127,64)	1.94	2.52	5.54

upper bound $G_{c,max}$ is not tight will have impact on the present energy comparison, see Figure 4.7.

The above example shows the optimistic nature of the bound in (4.44) for bit error rates $B \geq 10^{-6}$: it appears that the bound is at least a factor two too large. Therefore, in order to obtain a simple, yet reasonably realistic, model of the coding gain, we make the following assumption:

Assumption 4.2 *The coding gain of a rate R_c block code with error-correcting capability t is*

$$G_c = R_c(t + 1)/2. \tag{4.45}$$

Transmission-to-processing concentration. When employing an error correcting code, the purpose is here reduction of the transmit energy, but it also increases the processing energy. The result is a decrease in the transmission-to-processing ratio to

$$\rho'_c = \rho'_u \frac{R_c}{G_c^{1/g}} \tag{4.46}$$

where the second factor $R_c G_c^{-1/g} < 1$ for all codes of interest. We thus expect to observe a transmission-to-processing ratio concentration, as discussed previously when we considered Shannon-limit rate optimisation. Error correcting codes will counteract very large transmission-to-processing ratios. We return to this matter in the numerical results.

4.3.2   Energy saving through error correction

Employing a rate R_c code saves a fraction w of energy if

$$(1 - w)\breve{\mathcal{E}}_{\text{tot,u}} = \breve{\mathcal{E}}_{\text{tot,c}}. \tag{4.47}$$

By the use of (4.41) and (4.42) we obtain

$$(1 - w)\left(1 + \rho_u'\right) = \frac{1}{R_c} + \rho_u'\frac{1}{G_c^{1/g}}, \tag{4.48}$$

where we have invoked our simplifying assumption $\breve{\mathcal{E}}_{\text{CD}} = 0$. We can rearrange the above as

$$w = \frac{1 + \rho_u'\left(1 - G_c^{-1/g}\right) - R_c^{-1}}{1 + \rho_u'}. \tag{4.49}$$

Note that (4.49) is optimistic from a coding perspective as it neglects coding/decoding processing. We can only compute it for specific examples for given target bit error rates B; the coding gain G_c depends on the target bit error rate B and grows larger as B decreases. The use of the upper bound in (4.44) leads to the saving

$$w = \frac{1 + \rho_u'\left(1 - (R_c(t + 1))^{-1/g}\right) - R_c^{-1}}{1 + \rho_u'}, \tag{4.50}$$

which is indeed optimistic.

4.3.3   Adaptive coding within a class of BCH codes

Up to this point we have compared uncoded transmission with coded, but we have not compared different code rates. As we discussed previously, the globally optimal solution is difficult to obtain due to the lack of good energy models and the complexity of an exhaustive search over all codes. But by the use of our bound in (4.50) we can find the best sequence of codes within a specified class. The result will certainly not be the absolutely best choice but it will give good guidance and serve a useful purpose. Indeed, it is commonplace to optimise within a specified class of codes for practical reasons, see for example the work by Vucetic (1991) or Rice and Wicker (1994). There is a large body of literature on adaptive coding, but note that our topic here is different in that we are concerned with energy efficiency

under processing costs rather than throughput and spectral efficiency which are the common themes.

In analogy with (4.48) we can compare two codes (R_a, G_a) and (R_b, G_b) by the use of

$$\frac{1}{R_a} + \rho'_u \frac{1}{G_a^{1/g}} \geq \frac{1}{R_b} + \rho'_u \frac{1}{G_b^{1/g}}, \tag{4.51}$$

and hereby quantify the (uncoded) transmission-to-processing ratio ρ'_u required to change from code a to code b. The change-point is given by

$$\rho'_u \geq \frac{R_a - R_b}{R_a R_b} \frac{G_a^{1/g} G_b^{1/g}}{G_b^{1/g} - G_a^{1/g}}. \tag{4.52}$$

Let us now restrict ourselves, like Rice and Wicker (1994), to the class of BCH linear block codes and also make use of the upper bound $G_{c,max}$ on the coding gain. We performed the search over all BCH codes listed by Proakis (2001, Table 8.1-6) to find the ones with the smallest energy consumption

$$\breve{\mathcal{E}}_{tot} = \frac{1}{R_c} + \rho'_u \frac{1}{G_c^{1/g}} \tag{4.53}$$

for different transmission-to-total-processing ratios ρ'_u. It turned out that all the chosen codes were of block length 255 (the longest length within the class).[16] The use of a single block length simplifies the implementation of adaptive block coding, so this feature is welcome. Our results are given in Table 4.2 for two cases. First, we show the results for the ideal case when $G_c = G_{c,max}$ and the power amplifier consumption scales proportionally to the radiated power, that is $g = 1$. Second, we give results for a more realistic scenario when only half of the upper-bound gain is realised, $G_c = G_{c,max}/2$, and the power amplifier degradation exponent is $g = 2.8$. At first, the results appear counter-intuitive. As a larger degradation exponent g, and the reduced coding gain, reduces the transmission energy gains achievable by the coding, we would have expected higher code rates for $g = 2.8$ than for $g = 1$. This is also the case for transmission-to-total-processing ratios $\rho'_u < 0.5$, but for $\rho'_u > 0.5$ the result is the opposite. It seems that for large ρ'_u it is worth compensating the loss caused by the amplifier degradation by stronger coding; in the ideal case it suffices with higher code rates.

---

[16]Longer block lengths can, if included, of course turn out to be more favourable than the ones we have included here. Our goal is however not to find the overall optimal error correcting code, but to capture the first order energy-effects from coding. Moreover, if very short data packets are used, even the codes we consider here can be too long.

Table 4.2: *Adaptive BCH coding based on (4.52) and the upper bound on the coding gain given by (4.44). The two rightmost columns give the transmission-to-processing threshold at which the code should be changed for the next one in the leftmost column. All numbers correspond to a receiver-to-transmitter processing ratio $\alpha = 1$.*

Code (n,k)	Rate R_c	Transmission-to-total-processing threshold ρ'_u	
		$G_{c,max}, g = 1$	$G_{c,max}/2, g = 2.8$
(1,1)	1.0	0.00	0.00
(255,247)	0.969	0.07	(-)
(255,239)	0.937	0.22	(-)
(255,231)	0.906	0.47	0.55
(255,223)	0.875	0.84	0.76
(255,215)	0.843	1.37	1.11
(255,207)	0.812	2.12	1.57
(255,199)	0.780	3.14	2.18
(255,187)	0.733	3.46	2.21
(255,179)	0.702	8.89	5.32
(255,171)	0.671	12.7	7.47
(255,163)	0.639	18.7	10.6
(255,131)	0.514	21.4	11.4
(255,115)	0.451	163	82.3

4.3.4 Results for adaptive BCH coding

In Figure 4.7 the results from an optimisation according to (4.52) is shown for both the ideal case ($G_c = G_{c,max}, g = 1$) and the more realistic ($G_c = G_{c,max}/2, g = 2$). We observe the following from Figure 4.7:

Code rate. As expected, the code rate R_c decreases as the transmission-to-processing ratio increases: When the transmission costs become dominant it is motivated to use stronger codes. In the ideal case, coding is attractive already for $\rho'_u = 0.1$ because of the optimistic upper bound $G_{c,max}$ in conjunction with the constant power amplifier efficiency. In the more realistic case where we consider both efficiency degradation and smaller coding gains, the transmission-to-total-processing ratio must exceed 0.4 to motivate coding. Quite counter-intuitively, however, the code rates are lower in the more realistic case once they are introduced. This is probably due to the fact that even the higher code rates pay off very well in the ideal case. In any case, very low code
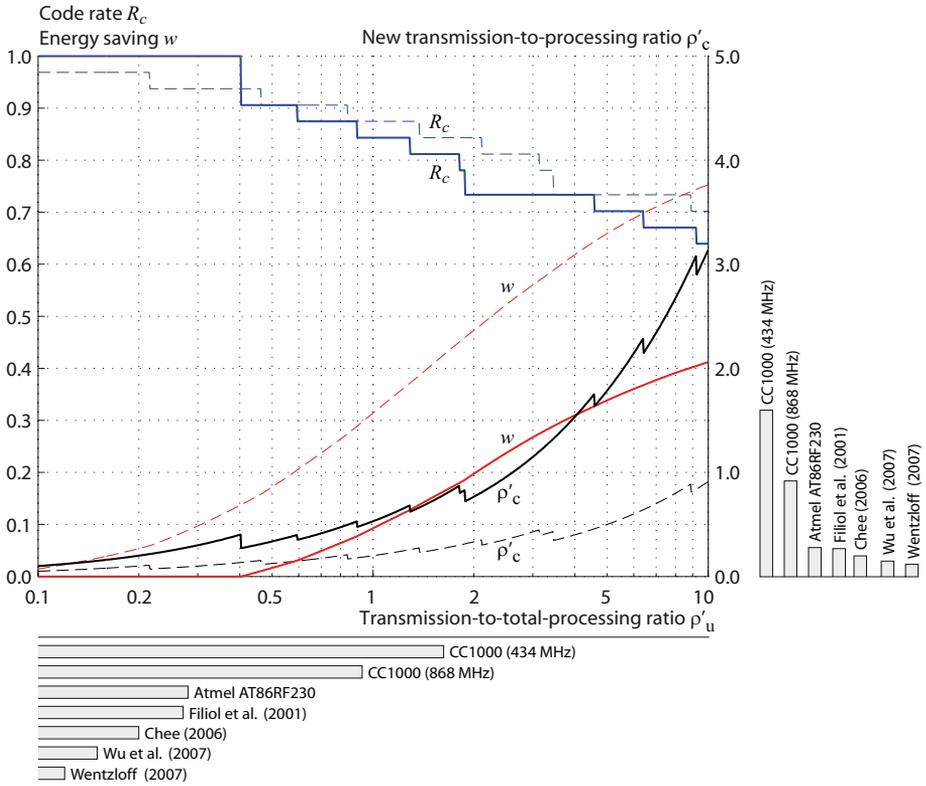
Figure 4.7: Code rate R_c (blue, left axis), energy saving factor w (red, left axis) and resulting transmission-to-total-processing ratio ρ'_c (black, right axis) resulting from an energy-optimal code rate choice within a class of BCH codes. Two cases are given; the ideal case $G_c = G_{c,max}$ and $g = 1$ (dashed lines), and the realistic case $G_c = G_{c,max}/2$ and $g = 2$ (solid lines). Note the drastically reduced efficiency of coding in the realistic setting.

rates $R_c < 1/2$ are never used over the range studied. The lowest rate used by the CC100 radio is $R_c = 0.73$.

Energy saving. There is a significant difference between the ideal and the more realistic case when it comes to energy savings. Note that, out of the radio examples shown, it is only the CC1000 radio that would benefit from error correcting codes with a maximum saving of about 17 percent according to the non-ideal characteristics.

New ρ'_c. Adoption of error correcting codes decreases the transmission costs and hence also the transmission-to-processing ratio. The more transmission energy we have to expend, the lower the optimum code rate

will be. This is the concentration phenomenon noted for the Shannon-limit optimisation in Section 4.1. It is clear from Figure 4.7 that there is such an effect, although it is less pronounced in the more realistic of the two cases shown.

Fading channels

Our results are based on the assumption of a static channel. In the unlikely case of a *fast* fading channel (changing on a symbol-time scale, see Definition 2.1), the errors can be sufficiently independent over symbols to render the coding efficient. On the other hand, fast power control would require immense feedback to counter the fast changes: The combination of slow power control and coding seems an appropriate choice.

For slow channels the power control feedback would suffice to determine, for each packet, the best code rate within a class of codes. A small additional overhead would however be required to reveal the code rate in use, but we here omit this overhead under the assumption that it is significantly smaller than the power control related overhead common to all code rates. It would affect the choice between codes or no codes, but not the choice between codes.

Range extension

Saving energy is not the only reason to use error correcting codes. Another, displayed in Figure 4.7, is that the codes can extend the transmission range. Consider the CC1000 radio operating at 434 MHz. It has a $\rho'_{\max} = 1.6$ and the use of adaptive BCH coding can at most save 17 percent of the total uncoded energy per bit at maximum output power (go from the horizontal axis at $\rho'_{\mathrm{u}} = 1.6$ in Figure 4.7 to the curve for w). However, the rate 0.81 code then in use has decreased the transmission-to-total-processing ratio to $\rho'_{\mathrm{c}} \approx 0.8$ and the radio is not operating at maximum output power; it can transmit over longer ranges. Considering the coded transmission-to-total-processing ratio ρ'_{R}, we can start from vertical, rightmost, axis at $\rho'_{\mathrm{R}} = 1.6$ to find that the lowest, energy-efficient code rate for the CC1000 radio is $R_{\mathrm{c}} = 0.73$ – going below that would always be inefficient although it could still be motivated by a need for range-extension.

4.3.5 Summary error correcting codes

Judging from the most realistic case of adaptive BCH codes that we have considered, error correcting codes require uncoded transmission-to-total-process-

ing ratios $\rho_{\mathrm{u}}' > 0.4$ to be energy-wise attractive. The power control scheme of Section 4.2 becomes attractive somewhat earlier and it is therefore reasonable to assume that the gains from error correcting codes can be reaped by this scheme. Code rates between $R_{\mathrm{c}} = 1$ and $R_{\mathrm{c}} = 2/3$ are the ones used, and lower rates seem to be of limited energy-saving value.

4.4 Adaptive QAM modulation

Energy-constrained adaptive modulation, with and without error correcting codes, was explored by Cui et al. (2005) who considered M-ary frequency shift keying (MFSK) and M-ary quadrature amplitude modulation (MQAM). The general conclusion was that optimised MQAM could save considerable amounts of energy, for short distances when high rates could be used, by reducing the transmission time. We here review this modulation optimisation briefly, with our focus on the transmission-to-processing ratio ρ, and then investigate whether the relatively complex and processing intensive hardware platform that adaptive MQAM requires can be motivated by the attained energy savings. Our conclusion is that it can not be motivated solely by energy consumption arguments.

4.4.1 Energy consumption models

Regarding coherent MQAM transmissions, we begin with the observation that binary phase shift keying (2QAM) and quaternary phase shift keying (4QAM) have the same signal-to-noise ratio requirement (Proakis, 2001) and it is therefore always better to use 4QAM in the present context since we can transmit the data in half the time.[17] We let

$$b \equiv \text{ number of bits per symbol}, \tag{4.54}$$

where $b \geq 2$ since we restrict ourselves to $M = 2^b \geq 4$. Let us define the optimisation with respect to 4QAM, $b = 2$, just like we defined the Shannon-limit optimisation with respect to a reference spectral efficiency C_0, see (4.6). Including the power amplifier efficiency degradation (2.14) and the peak-to-average power ratio of MQAM (more on this later on), our

---

[17]We are here assuming that the system always processes both the in-phase component and the quadrature component. Of course, a system designed solely for 2QAM could work with less processing costs than a system designed for 4QAM, but in our present problem the adaptivity requires both components and the possible gain for 2QAM is thus unattainable.

normalised total-energy model, which conforms to the model used by Cui et al. (2005), is

$$
\begin{aligned}
\breve{\mathcal{E}}_{\text{tot}}(b) &= \frac{\mathcal{E}_{\text{tot}}(b)}{\mathcal{E}_{\text{RP}}(b=2)} \\
&= \frac{\mathcal{E}_{\text{RP}}(b)}{\mathcal{E}_{\text{RP}}(b=2)} + \frac{\mathcal{E}_{\text{T}}(b)}{\mathcal{E}_{\text{RP}}(b=2)} \\
&= \frac{2}{b} + \rho_2' \frac{\epsilon(2)}{\epsilon(b)} \left(\frac{\gamma(b,B)}{\gamma(2,B)} \right)^{1/g},
\end{aligned}
\tag{4.55}
$$

where ρ_2' is the transmission-to-total-processing ratio for $b = 2$. In (4.55), the last line, the first term on the righthand side is the normalised processing energy which is inversely proportional to b, while the second term is the normalised transmission energy, both with respect to 4QAM, $b = 2$: if we insert $b = 2$ we arrive at $\breve{\mathcal{E}}_{\text{tot}} = 1 + \rho_2'$. Moreover, we have in (4.55) introduced the large-symbol penalty

$$
\frac{\epsilon(2)}{\epsilon(b)},
\tag{4.56}
$$

due to increased peak-to-average power ratio for larger b.

Assumption 4.3 *For MQAM symbols with b bits per symbol, consideration of the symbols peak-to-average power ratio leads to a large symbol energy penalty relative to $b = 2$ of*

$$
\frac{\epsilon(2)}{\epsilon(b)} = 3 \frac{\sqrt{2^b} - 1}{\sqrt{2^b} + 1}.
\tag{4.57}
$$

∎

See Appendix 4.G for a motivation of Assumption 4.3. The larger symbols also require better signal-to-noise ratio per bit to satisfy the target bit error rate B, and this enters (4.55) through the factor

$$
\frac{\gamma(b,B)}{\gamma(2,B)}.
\tag{4.58}
$$

Assumption 4.4 *The bit error rate B for Gray coded MQAM, $M = 2^b$, $b \geq 2$, is*

$$
B = \frac{2}{b} \cdot \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \cdot \text{erfc} \left(\sqrt{\frac{3b}{2^b - 1} \frac{\gamma}{2}} \right),
\tag{4.59}
$$

where γ is the received signal-to-noise ratio per bit.

∎

See Appendix 7.A for our motivation of assumption 4.4.

We make the following observations before we move on to numerical results:

- Similar to the Shannon-limit optimisation in Section 4.1, optimisation of the number of bits per symbol b rests on the trade-off between decreased processing time and increased transmission power. The differences are in the transmit energy term, where we now include: a peak-to-average power penalty for large b; the power amplifier efficiency degradation through g; a slightly different scaling behaviour of the radiated energy though (4.59).

- A degradation exponent $g > 1$ will tend to favour large b because larger b require more output power and hence the amplifier can operate more efficiently.

- Judging from (4.59), the received signal-to-noise ratio per bit γ must scale approximately like

$$\frac{2^b - 1}{b} \tag{4.60}$$

to maintain a given bit error rate B. This is the same behaviour as the Shannon limit in (4.1). In fact, as we outline in Appendix 4.I, this Shannon-type scaling holds approximately also for Nakagami-m fading channels. This means that the results obtained by the use of (4.59) will be representative for a larger range of conditions than expected.

4.4.2   Optimisation results

The energy-optimised number of bits b per symbol is shown in Figure 4.8 together with the continuous rate optimisation based on the Shannon limit, see (4.6). First we observe that there is a close correspondence between these two optimisations when the power amplifier degradation exponent $g = 1$, which was assumed in Section 4.1. They are separated by an approximately constant distance of 4 dB due to the peak-to-average factor in Assumption 4.3. These results are also in agreement with the ones provided by Cui et al. (2005). Second, we see that inclusion of the degradation exponent g significantly increases the preferable number of bits b per symbol. This effect signifies that power amplifiers should be operating close to their maximum efficiency, and that the peak-to-average penalty of large symbols thereby is outweighed. Third, we would like to stress that constellation sizes grow exponentially with the number of bits, $M = 2^b$. In practice it is not reasonable
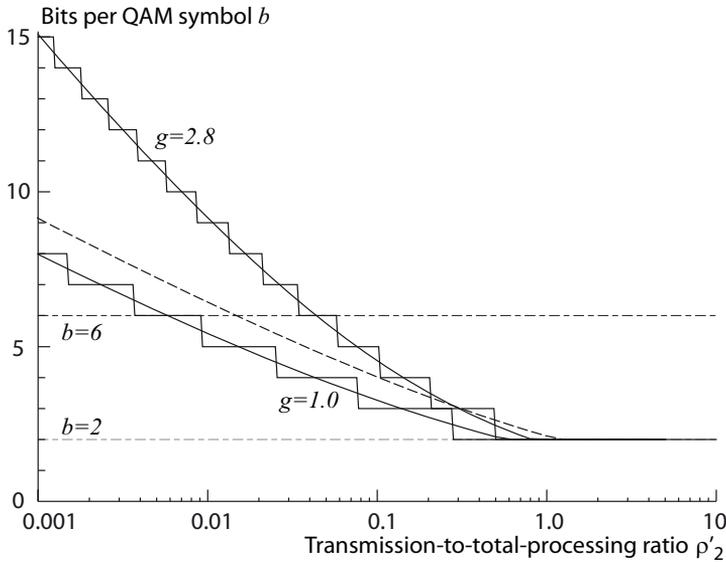
Figure 4.8: Number of bits b per optimised MQAM symbol (solid lines), and the optimised spectral efficiency based on the Shannon limit (dashed line), in the static channel scenario. All results are obtained with $B = 10^{-6}$.

to assume that $b = 16$, say, could be used as it corresponds to 65536-QAM. Therefore we indicate the limit $b = 6$, that is 64-QAM, as a reasonable practical constraint in the figure. We comment more on practical considerations in Section 4.4.4. Fourth, for transmission-to-total-processing ratios $\rho' > 1/2$ there is no preference for large MQAM symbols since the processing cost is not overly dominant.

Energy savings and transmission-to-processing ratio

As usual, we quantify the obtained energy reduction in relative terms, presently compared to the energy consumption for $b = 2$,

$$w = \frac{\breve{\mathcal{E}}_{\text{tot}}(2) - \breve{\mathcal{E}}_{\text{tot}}(b)}{\breve{\mathcal{E}}_{\text{tot}}(2)}, \tag{4.61}$$

with w being the fraction of saved energy, and $\breve{\mathcal{E}}_{\text{tot}}(b)$ is given by (4.55). By transmitting with b bits per MQAM symbol, the transmission-to-total

processing ratio is found from (4.55),

$$
\begin{aligned}
\rho_b' &= \frac{\rho_2' \frac{\epsilon(2)}{\epsilon(b)} \left(\frac{\gamma(b,B)}{\gamma(2,B)} \right)^{1/g}}{\frac{2}{b}} \\
&= \rho_2' \frac{b}{2} \frac{\epsilon(2)}{\epsilon(b)} \left(\frac{\gamma(b,B)}{\gamma(2,B)} \right)^{1/g} .
\end{aligned}
\tag{4.62}
$$

This is, naturally, steadily increasing with b for a fixed ρ_2': increasing b means increasing the per bit transmission costs and decreasing the per bit processing costs.

The fraction w of energy saved by MQAM optimisation is shown in Figure 4.9, which also displays the resulting transmission-to-total-processing ratio ρ_b' after applying the new symbol size. Clearly, for small transmission-to-total-processing ratios $\rho_2' < 0.1$ there are large processing costs to be saved by balancing them better against the transmission costs – note the significant increase in the transmission-to-total-processing ratio ρ_b' achieved. We observe that even if we constrain the optimisation to $b \leq 6$, there seems to be large attainable reductions in energy consumption. So, given that we have nodes with the flexibility exploited here we should definitely use it for short distances. However, there are several practical constraints not included in the model, and also the question if the increased energy consumption of a flexible hardware platform pays off in the end. So let us now turn to this question.

4.4.3   How much is adaptivity worth?

A radio supporting coherent detection of adaptive MQAM is naturally more complex than, say, a radio designed for envelope (square-law, non-coherent) detection of binary frequency shift keying (NCFSK). So even if the adaptive approach can save substantial amounts of processing energy at small transmission-to-processing ratios it will start from a higher level than a simple non-coherent radio. Let us see how much larger the initial processing cost is allowed to be for adaptivity to be worth its price.

We use the subscripts Q and N to denote the MQAM and binary NCFSK schemes respectively. To motivate the adaptive MQAM system the following inequality must be satisfied,

$$
\mathcal{E}_{\mathrm{RP,Q}}(b) + \mathcal{E}_{\mathrm{T,Q}}(b) < \mathcal{E}_{\mathrm{RP,N}} + \mathcal{E}_{\mathrm{T,N}},
\tag{4.63}
$$

where we have used the *total* processing energy per bit $\mathcal{E}_{\mathrm{RP}}$ and the transmission energy per bit $\mathcal{E}_{\mathrm{T}}$. In accordance with (4.55) and the reasoning
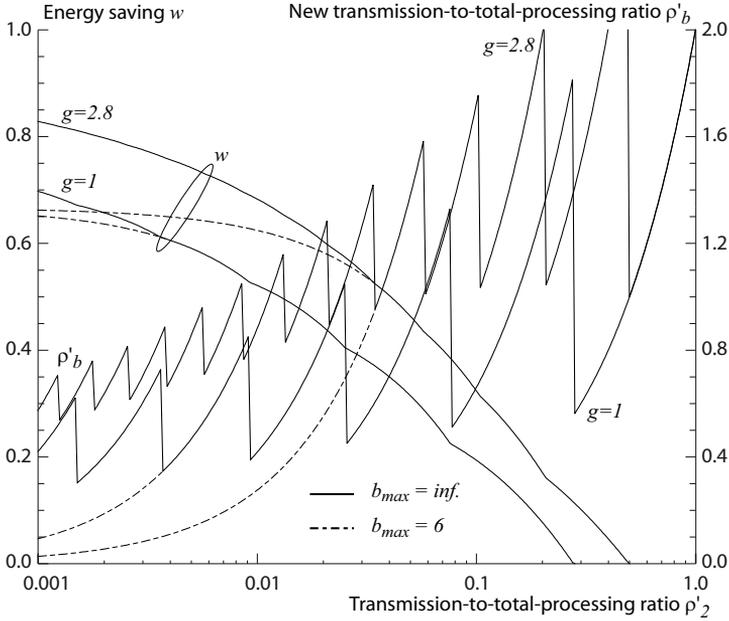
Figure 4.9: Achieved energy savings w and the corresponding transmission-to-processing ratio ρ'_b after optimisation.

regarding amplifier efficiency we can rewrite the inequality as

$$\frac{\mathcal{E}_{\mathrm{RP,Q}}(2)}{b/2}+\mathcal{E}_{\mathrm{T,Q}}(2)\frac{\epsilon_{\mathrm{Q}}(2)}{\epsilon_{\mathrm{Q}}(b)}\left(\frac{\gamma_{\mathrm{Q}}(b,B)}{\gamma_{\mathrm{Q}}(2,B)}\right)^{1/g}<\mathcal{E}_{\mathrm{RP,N}}+\mathcal{E}_{\mathrm{T,Q}}(2)\frac{\epsilon_{\mathrm{Q}}(2)}{\epsilon_{\mathrm{N}}}\left(\frac{\gamma_{\mathrm{N}}(B)}{\gamma_{\mathrm{Q}}(2,B)}\right)^{1/g},$$

$$(4.64)$$

where the left-hand side is optimised with respect to b. Presently, we are interested in the maximum allowed processing cost of the adaptive hardware relative to the simple non-coherent radio. For this reason we define the processing-to-processing ratio

$$\beta \equiv \frac{\mathcal{E}_{\mathrm{RP,Q}}(b=2)}{\mathcal{E}_{\mathrm{RP,N}}}. \qquad (4.65)$$

Solving (4.64) for β yields the upper limit

$$\beta < \left(\frac{2}{b}+\rho'_2\left[\frac{\epsilon_{\mathrm{Q}}(2)}{\epsilon_{\mathrm{Q}}(b)}\left(\frac{\gamma(b,B)}{\gamma(2,B)}\right)^{1/g}-\frac{\epsilon_{\mathrm{Q}}(2)}{\epsilon_{\mathrm{N}}}\left(\frac{\gamma_{\mathrm{N}}(B)}{\gamma(2,B)}\right)\right]\right)^{-1} \qquad (4.66)$$

with $\rho'_2 = \mathcal{E}_{\mathrm{T,Q}}(2)/\mathcal{E}_{\mathrm{RP,Q}}(2)$ the adaptive transmitter's transmission-to-total-processing ratio. Note that b is a function of ρ'_2 for an optimised MQAM transmission.

Two entities are still unspecified: The ratios $\epsilon_Q(2)/\epsilon_N$ and $\gamma_N(B)/\gamma(2,B)$. The former is determined by the amplifier efficiencies, and since NCFSK can operate with a non-linear amplifier its efficiency is superior to the MQAM amplifier. Without further due we assume that the ratio is 3/4. For a given bit error rate criterion B, the penalty in required per-bit signal-to-noise ratio γ is for NCFSK about 4 times, or 6 dB, relative to 4-QAM (Proakis, 2001). With these numbers we calculate the upper limit in (4.66); the result is displayed in Figure 4.10. When the non-coherent system operates under balanced conditions, that is with a transmission-to-processing ratio $\rho'_N \approx 1$, we see that the adaptive MQAM system can not afford more than roughly 50 percent more processing energy than its non-coherent competitor; $\beta \approx 1.5$. To be uniformly superior, the implementation of adaptivity can add at most 15 percent processing cost. As we deem such a design virtually impossible, our conclusion is that an adaptive hardware platform supporting MQAM will save energy for nodes communicating of good links but waste energy for nodes in more demanding conditions; it will take from the poor and give to the rich. Greater energy-unbalance than necessary can be caused between "rich" and "poor" nodes. [18]

4.4.4 Practical limitations

Unconstrained MQAM optimisation can as we have seen result in extreme symbol sizes. Here we give a brief list of practical considerations not included in our present analysis, considerations that disfavour an adaptive MQAM implementation.

Analog to digital converter (ADC). The use of larger MQAM symbols puts harder requirements on the receiver's ADC. Imprecise discretisation can increase the effective noise level and cause error floors. An example of the impact of the number of ADC levels is given by Shen and Zhang (2002) who study the caused loss in signal-to-noise ratio for MQAM modulation. A 6-bit ADC causes a 6 dB loss for 256-QAM and a 2 dB loss for 64-QAM. In order to get below 0.5 dB in loss, an 8-bit ADC is needed for 64-QAM and a 9-bit ADC is needed for 256-QAM.

Phase noise. In addition to other noise sources, the phase noise of the local oscillator used in the receiver can cause problems for large MQAM symbols. Costa and Pupolin (2002) study the impact of phase noise

---

[18]Observe that the increase in β for $\rho'_N > 1$ is due to the transmission-energy advantage of MQAM, but this advantage would decrease by the use of error correcting codes as previously studied.
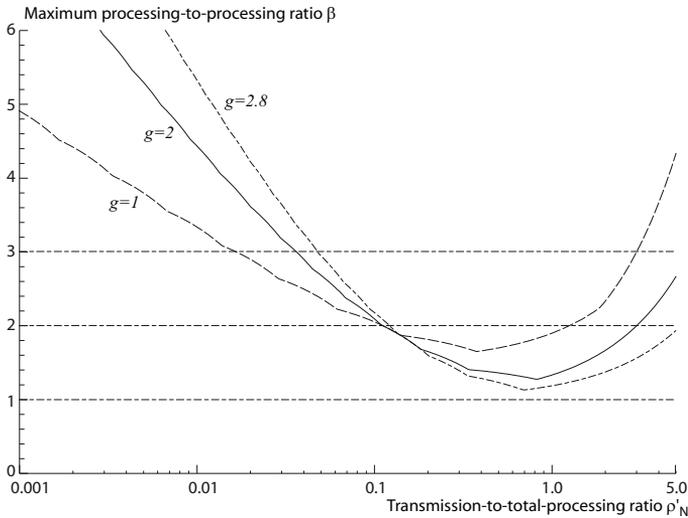
Figure 4.10: Maximum allowed processing-to-processing ratio β for superiority of the adaptive MQAM system over the binary non-coherent frequency shift keying alternative. We here use $\alpha = 1$.

and we can in there work see energy penalties exceeding 5 dB even for moderate levels of phase noise in a 16-QAM receiver. The penalty increases significantly with symbol size.

Channel uncertainties. With increasing number of bits per symbol, the sensitivity to channel estimation error becomes more pronounced and translates to a need for more overhead transmissions. Correct detection of large MQAM symbols requires significantly more precise channel state information than does non-coherent detection of orthogonal signals.

Non-linear decrease in processing costs. Assuming that all processing costs scale like $1/b$ is obviously optimistic as some overhead costs, such as the startup energy consumption, are unaffected by the symbol size. Interestingly, Shih et al. (2001) found that the start-up time has significant impact, and if it is not short enough the gains from MQAM are erased.

Monetary costs. Complexity costs money, and the sophisticated transceivers needed for MQAM can come at a cost that outweighs the saving in energy (although this is highly application dependent).

4.4.5   Summary adaptive QAM

Indeed, if nodes are equipped with flexible QAM transceiver hardware, it is difficult to argue against the use of adaptive MQAM – except possibly by considering the overhead required for adapting the rate. However, the other negatives in terms of complexity, monetary costs, increased processing energy and greater energy imbalance between nodes rule out the use of sophisticated MQAM radios for most energy-limited wireless sensor networks.

4.5   Conclusions

Shannon-limit rate optimisation.   The idealised energy optimisation highlights two matters:

- The fundamental trade-off between processing and transmission energy that can not be avoided in the present sensor-network context.

- Transmission optimisation will, typically, tend to balance transmission and processing energies; as soon as one of them dominates there is probably a better alternative.

Power control.   Regarding truncated channel inversion we come to the conclusion that if we aim at actually saving energy the transmission-to-total-processing ratio should be larger than 1/3. Below this threshold there is really no energy gain from power control, and we note that several existing radios therefore should use a fixed output power. Above the threshold the gain increases steadily even for substantial amounts of control information feedback.

- Use power control to counter slow fading, when limited feedback can be used, and leave fast fading to the channel codes.

- Coarse grained power control is probably enough; it will reap the major gains at a small feedback fraction. There is no need for adjustable power below $\rho' = 1/3$.

Error correcting codes.   Channel coding is a good complement to power control and should be used for transmission-to-total-processing ratios $\rho' > 0.4$. Simple, fixed block length, codes will suffice. Single length blocks facilitates simple, slow, adaption of the code rate.

Adaptive MQAM. Requires complex and processing intensive radio architecture; the processing energy is in fact too large to motivate the adaptivity from en energy point-of-view. Therefore, adaptive MQAM should be used *only* if the platform supports it for other reasons.

Appendix 4.A   Proof of Lemma 4.1

Setting the derivative of (4.6), with respect to C_e, to zero we obtain an equation to solve for the optimum spectral efficiency C_{opt},

$$\rho_0' \left(1 - 2^{C_{\text{opt}}} + C_{\text{opt}}2^{C_{\text{opt}}} \ln(2)\right) = 2^{C_0} - 1. \tag{4.67}$$

Rearranging terms and exploiting that $2^x = e^{\ln(2)x}$ we take the following steps

$$C_{\text{opt}}2^{C_{\text{opt}}} \ln(2) - 2^{C_{\text{opt}}} = \frac{2^{C_0} - 1 - \rho_0'}{\rho_0'}, \tag{4.68}$$

$$\ln(2)C_{\text{opt}}e^{\ln(2)C_{\text{opt}}} - e^{\ln(2)C_{\text{opt}}} = \frac{2^{C_0} - 1 - \rho_0'}{\rho_0'}, \tag{4.69}$$

$$\ln(2)C_{\text{opt}}e^{\ln(2)C_{\text{opt}}-1} - e^{\ln(2)C_{\text{opt}}-1} = \frac{2^{C_0} - 1 - \rho_0'}{e\rho_0'}, \tag{4.70}$$

$$\left(\ln(2)C_{\text{opt}} - 1\right)e^{\ln(2)C_{\text{opt}}-1} = \frac{2^{C_0} - 1 - \rho_0'}{e\rho_0'}. \tag{4.71}$$

We can now use the Lambert W function W_{L} since it is the solution to $x = W_{\text{L}}(x)\exp(W_{\text{L}}(x))$;

$$\ln(2)C_{\text{opt}} - 1 = W_{\text{L}}\left(\frac{2^{C_0} - 1 - \rho_0'}{e\rho_0'}\right), \tag{4.72}$$

$$C_{\text{opt}} = \frac{1}{\ln(2)}\left[1 + W_{\text{L}}\left(\frac{2^{C_0} - 1 - \rho_0'}{e\rho_0'}\right)\right]. \tag{4.73}$$

∎

Appendix 4.B   Proof of Theorem 4.1

First we rearrange (4.67) to

$$1 - 2^{C_{\text{opt}}} + C_{\text{opt}}2^{C_{\text{opt}}} \ln(2) = \frac{2^{C_0} - 1}{\rho_0'}. \tag{4.74}$$

Then we invoke (4.7), which states that the post-optimisation transmission-to-total-processing ratio

$$\rho_{\text{opt}}' = \rho_0'\frac{2^{C_{\text{opt}}} - 1}{2^{C_0} - 1}, \tag{4.75}$$

on the right hand side of (4.74). Isolating ρ_{opt}' we arrive at

$$\rho_{\text{opt}}' = \frac{2^{C_{\text{opt}}} - 1}{\ln(2)C_{\text{opt}}2^{C_{\text{opt}}} - (2^{C_{\text{opt}}} - 1)}. \tag{4.76}$$

With the aid of Mathematica 6 (Wolfram Research Inc., 2007) we can then solve (4.76) for the spectral efficiency C_{opt} to arrive at equations (4.11) and (4.12) in Theorem 4.1. ∎

Appendix 4.C   Proof of Corollary 4.1

Begin by multiplying both sides of (4.10) by C_{opt};

$$\rho'_{\text{opt}}C_{\text{opt}} = \frac{C_{\text{opt}}\left(2^{C_{\text{opt}}} - 1\right)}{\ln(2)C_{\text{opt}}2^{C_{\text{opt}}} - \left(2^{C_{\text{opt}}} - 1\right)}. \tag{4.77}$$

Then rewrite as

$$\rho'_{\text{opt}}C_{\text{opt}} = \frac{\left(1 - 2^{-C_{\text{opt}}}\right)}{\ln(2) - \frac{1 - 2^{-C_{\text{opt}}}}{C_{\text{opt}}}} \tag{4.78}$$

to see that

$$\lim_{C_{\text{opt}} \to \infty} \rho'_{\text{opt}}C_{\text{opt}} = \frac{1}{\ln(2)}. \tag{4.79}$$

To obtain the limit when $C_{\text{opt}} \to 0$ we make use of the first l'Hôpital rule – valid for limits of the form $0/0$ – which is based on the limit of the ratio of the derivatives. Differentiating the numerator and denominator in (4.77) with respect to C_{opt} we obtain

$$
\begin{aligned}
\lim_{C_{\text{opt}} \to 0^+} \rho'_{\text{opt}}C_{\text{opt}} &= \lim_{C_{\text{opt}} \to 0^+} \frac{\frac{\mathrm{d}}{\mathrm{d}C_{\text{opt}}}C_{\text{opt}}\left(2^{C_{\text{opt}}} - 1\right)}{\frac{\mathrm{d}}{\mathrm{d}C_{\text{opt}}}\left(\ln(2)C_{\text{opt}}2^{C_{\text{opt}}} - \left(2^{C_{\text{opt}}} - 1\right)\right)} \\
&= \lim_{C_{\text{opt}} \to 0^+} \frac{\left(2^{C_{\text{opt}}} - 1\right) + \ln(2)C_{\text{opt}}2^{C_{\text{opt}}}}{\ln^2(2)C_{\text{opt}}2^{C_{\text{opt}}}} \\
&= \lim_{C_{\text{opt}} \to 0^+} \frac{1}{\ln(2)}\left[1 + \frac{1}{\ln(2)}\frac{1 - 2^{-C_{\text{opt}}}}{C_{\text{opt}}}\right] \\
&= \frac{1}{\ln(2)} + \frac{1}{\ln^2(2)}\lim_{C_{\text{opt}} \to 0^+}\frac{1 - 2^{-C_{\text{opt}}}}{C_{\text{opt}}} \\
&= \frac{2}{\ln(2)}.
\end{aligned}
\tag{4.80}
$$

In the last step we used a second application of the first l'Hôpital rule to find that

$$
\begin{aligned}
\lim_{C_{\text{opt}} \to 0^+} \frac{1 - 2^{C_{\text{opt}}}}{C_{\text{opt}}} &= \lim_{C_{\text{opt}} \to 0^+} \ln(2)2^{-C_{\text{opt}}} \\
&= \ln(2).
\end{aligned}
\tag{4.81}
$$

It remains to show that the derivative of $\rho'_{\text{opt}}C_{\text{opt}}$ in (4.77) with respect to C_{opt} is non-positive. Differentiating the right hand side of (4.77) with respect to C_{opt} we obtain

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}C_{\text{opt}}}\rho'_{\text{opt}}C_{\text{opt}} &= \frac{\mathrm{d}}{\mathrm{d}C_{\text{opt}}}\frac{C_{\text{opt}}\left(2^{C_{\text{opt}}} - 1\right)}{\ln(2)C_{\text{opt}}2^{C_{\text{opt}}} - \left(2^{C_{\text{opt}}} - 1\right)} \\
&= \frac{2^{C_{\text{opt}}}C_{\text{opt}}^2\ln^2(2) - \left(2^{C_{\text{opt}}} - 1\right)^2}{\left(\ln(2)C_{\text{opt}}2^{C_{\text{opt}}} - \left(2^{C_{\text{opt}}} - 1\right)\right)^2},
\end{aligned}
\tag{4.82}
$$

where the denominator is positive for all $C_{\text{opt}} > 0$. Let us now perform a series expansion of the nominator in (4.82) around $C_{\text{opt}} = 0$,

$$2^{C_{\text{opt}}} C_{\text{opt}}^2 \ln^2(2) - \left(2^{C_{\text{opt}}} - 1\right)^2 = \sum_{n=4}^{\infty} \frac{\ln^n(2)}{n!} \left(2 + n(n-1) - 2^n\right) C_{\text{opt}}^n. \quad (4.83)$$

Since all coefficients $2 + n(n-1) - 2^n$ in the series expansion are negative for all $n > 1$, the derivative must be negative over the whole positive range of spectral efficiencies C_{opt}. Note that the end-point limits of the derivative in (4.82) are

$$\begin{aligned}
\lim_{C_{\text{opt}} \to 0^+} \frac{2^{C_{\text{opt}}} C_{\text{opt}}^2 \ln^2(2) - \left(2^{C_{\text{opt}}} - 1\right)^2}{\left(\ln(2) C_{\text{opt}} 2^{C_{\text{opt}}} - \left(2^{C_{\text{opt}}} - 1\right)\right)^2} &= -\frac{1}{3} \\
\lim_{C_{\text{opt}} \to \infty} \frac{2^{C_{\text{opt}}} C_{\text{opt}}^2 \ln^2(2) - \left(2^{C_{\text{opt}}} - 1\right)^2}{\left(\ln(2) C_{\text{opt}} 2^{C_{\text{opt}}} - \left(2^{C_{\text{opt}}} - 1\right)\right)^2} &= 0,
\end{aligned} \quad (4.84)$$

as obtained by the use of Mathematica 6 (Wolfram Research Inc., 2007). ■

Appendix 4.D   Proof of Theorem 4.2

By the use of (4.20) and the channel inversion rule (4.24) we express the received signal-to-noise ratio

$$\begin{aligned}
\gamma_{\text{ICI}} &= K x_{\text{f}} \frac{\mathcal{E}_{\text{ICI},0}}{x_{\text{f}}} \\
&= K \mathcal{E}_{\text{ICI},0}.
\end{aligned} \quad (4.85)$$

This constant signal-to-noise ratio per bit must be the same as for the static channel, γ_{sta}, and we can by the use of Definition 4.1 find that

$$\mathcal{E}_{\text{ICI},0} = \frac{\mathcal{E}_{\text{LM}}}{M_{\text{L}}}. \quad (4.86)$$

The inversion rule at the transmitter can hence be written

$$\mathcal{E}_{\text{ICI}} = \frac{\mathcal{E}_{\text{LM}}}{M_{\text{L}}} \frac{1}{x_{\text{f}}}. \quad (4.87)$$

Through averaging over our gamma probability distribution for x_{f} we obtain by the use of Gradshteyn and Ryzhik (2000) or Mathematica 6 (Wolfram Research Inc., 2007)

$$\begin{aligned}
\overline{\mathcal{E}}_{\text{ICI}} &= \frac{\mathcal{E}_{\text{LM}}}{M_{\text{L}}} \int_0^{\infty} \frac{1}{x_{\text{f}}} \frac{m^m}{\Gamma(m)} x_{\text{f}}^{m-1} e^{-m x_{\text{f}}} \mathrm{d}x_{\text{f}} \\
&= \frac{m}{m-1} \frac{\mathcal{E}_{\text{LM}}}{M_{\text{L}}},
\end{aligned} \quad (4.88)$$

where the last step is valid for $m > 1$. ■

Appendix 4.E   Proof of Lemma 4.3

At the receiver, the signal-to-noise ratio per bit is $\gamma_{\mathrm{rec}} = K x_{\mathrm{f}} \mathcal{E}_{\mathrm{rad}}$, see (4.20). The power-limited control scheme in (4.27) will achieve a signal-to-noise ratio

$$
\begin{aligned}
\gamma_{\mathrm{CI}} &= K x_{\mathrm{f}} \mathcal{E}_{\mathrm{CI}} \\
&= \begin{cases} \overline{\gamma}_{\mathrm{LM}} \dfrac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}} & , \quad x_{\mathrm{f}} \geq \dfrac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}} \\[2mm] \overline{\gamma}_{\mathrm{LM}} \dfrac{\mathcal{E}_{\mathrm{max}}}{\mathcal{E}_{\mathrm{LM}}} x_{\mathrm{f}} & , \quad x_{\mathrm{f}} < \dfrac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}} \end{cases} ,
\end{aligned} \tag{4.89}
$$

where the two cases correspond to a perfectly inverted channel and the residual fading when the transmitter uses its maximum output $\mathcal{E}_{\mathrm{max}}$, respectively. For a fading probability distribution $p(x_{\mathrm{f}}|I)$, we obtain through a change of variables from x_{f} to γ_{CI} according to (4.89) the following

$$
\begin{aligned}
p(\gamma_{\mathrm{CI}}|I) =& \delta\left(\gamma_{\mathrm{CI}} - \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}} \overline{\gamma}_{\mathrm{LM}}\right) \int\limits_{\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}}^{\infty} p(x_{\mathrm{f}}|I) \mathrm{d}x_{\mathrm{f}} \\
&+ \left[1 - \Theta\left(\gamma_{\mathrm{CI}} - \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}} \overline{\gamma}_{\mathrm{LM}}\right)\right] p\left(\gamma_{\mathrm{CI}}|\gamma_{\mathrm{CI}} = \overline{\gamma}_{\mathrm{LM}} \frac{\mathcal{E}_{\mathrm{max}}}{\mathcal{E}_{\mathrm{LM}}} x_{\mathrm{f}}, I\right),
\end{aligned} \tag{4.90}
$$

where $\delta(\cdot)$ is Dirac's delta function and $\Theta(\cdot)$ is Heaviside's theta function. The first term is the probability that the channel can be inverted, and the the received signal-to-noise ratio per bit thereby is kept constant. The second term is the probability density function for the remaining, non-inverted but reduced, channel variations. Now, assuming perfect channel estimates at the receiver, we can average the static-channel bit error rate $B_{\mathrm{sta}}(\gamma)$ over (4.90):

$$
\begin{aligned}
\overline{B}_{\mathrm{IC}} =& \int\limits_{0}^{\infty} B_{\mathrm{sta}}(\gamma_{\mathrm{CI}}) p(\gamma_{\mathrm{CI}}|I) \mathrm{d}\gamma_{\mathrm{CI}} \\
=& P\left(x_{\mathrm{f}} \geq \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}} \Big| I\right) B_{\mathrm{sta}}\left(\overline{\gamma}_{\mathrm{LM}} \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}}\right) \\
&+ \int\limits_{0}^{\overline{\gamma_{\mathrm{CI}}}_{\mathrm{LM}} \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}}} p\left(\gamma_{\mathrm{CI}}|\gamma_{\mathrm{CI}} = \overline{\gamma_{\mathrm{CI}}}_{\mathrm{LM}} \frac{\mathcal{E}_{\mathrm{max}}}{\mathcal{E}_{\mathrm{LM}}} x_{\mathrm{f}}, I\right) B_{\mathrm{sta}}(\gamma_{\mathrm{CI}}) \mathrm{d}\gamma_{\mathrm{CI}}
\end{aligned} \tag{4.91}
$$

We have through Assumption 2.6 on page 29 assigned a gamma distribution for x_{f}, $\mathrm{Gam}(x_{\mathrm{f}}|m,1)$. We obtain the following probability density function for the received signal-to-noise ratio per bit γ_{CI} by the use of (4.90),

$$
\begin{aligned}
p(\gamma_{\mathrm{CI}}|I) =& \delta\left(\gamma_{\mathrm{CI}} - \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}} \overline{\gamma}_{\mathrm{LM}}\right) \frac{\Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma(m)} \\
&+ \left[1 - \Theta\left(\gamma_{\mathrm{CI}} - \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}} \overline{\gamma}_{\mathrm{LM}}\right)\right] \frac{m^m \gamma_{\mathrm{CI}}^{m-1} e^{-\frac{m}{\left(\frac{\mathcal{E}_{\mathrm{max}}}{\mathcal{E}_{\mathrm{LM}}} \overline{\gamma}_{\mathrm{LM}}\right)} \gamma_{\mathrm{CI}}}}{\Gamma(m) \left(\frac{\mathcal{E}_{\mathrm{max}}}{\mathcal{E}_{\mathrm{LM}}} \overline{\gamma}_{\mathrm{LM}}\right)^m},
\end{aligned} \tag{4.92}
$$

where we have used Mathematica 6 (Wolfram Research Inc., 2007) to find the integrals in (4.90). The bit error rate for differentially detected binary phase shift keying is in the static channel case (Proakis, 2001)

$$B_{\mathrm{sta}} = \frac{1}{2} e^{-\gamma_{\mathrm{sta}}}. \tag{4.93}$$

Inserting (4.93) and (4.92) into (4.91) we arrive at

$$
\begin{aligned}
\overline{B}_{\mathrm{CI}} =& \frac{\Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)} \cdot \frac{1}{2} e^{-\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{LM}}}\gamma_{\mathrm{LM}}} \\
&+ \frac{\Gamma\left(m\right) - \Gamma\left(m, \mathcal{E}_{\mathrm{CI},0}\left(\frac{m}{\mathcal{E}_{\mathrm{max}}} + \frac{\gamma_{\mathrm{LM}}}{\mathcal{E}_{\mathrm{LM}}}\right)\right)}{\Gamma\left(m\right)} \cdot \frac{1}{2} \left(\frac{m\mathcal{E}_{\mathrm{LM}}}{m\mathcal{E}_{\mathrm{LM}} + \mathcal{E}_{\mathrm{max}}\gamma_{\mathrm{LM}}}\right)^{m}
\end{aligned}
\tag{4.94}
$$

by the use of Mathematica 6 (Wolfram Research Inc., 2007).   ∎

Appendix 4.F   Proof of Theorem 4.3

The radiated energy per bit $\mathcal{E}_{\mathrm{CI}}$ follow

$$
\mathcal{E}_{\mathrm{CI}} = \begin{cases} \mathcal{E}_{\mathrm{max}} & , \quad x_{\mathrm{f}} \leq \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}, \\ \frac{\mathcal{E}_{\mathrm{CI},0}}{x_{\mathrm{f}}} & , \quad x_{\mathrm{f}} > \frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}. \end{cases}
\tag{4.95}
$$

The probability density function $p(\mathcal{E}_{\mathrm{CI}}|\mathcal{E}_{\mathrm{CI},0}, \mathcal{E}_{\mathrm{max}}, m, I)$ can be found from our gamma distribution for x_{f} and the limits given by (4.95). We obtain

$$
\begin{aligned}
p(\mathcal{E}_{\mathrm{CI}}|\mathcal{E}_{\mathrm{CI},0}, \mathcal{E}_{\mathrm{max}}, m, I) =& \delta(\mathcal{E}_{\mathrm{CI}} - \mathcal{E}_{\mathrm{max}})P(x_{\mathrm{f}} \leq \mathcal{E}_{\mathrm{CI},0}/\mathcal{E}_{\mathrm{max}}|m, I) \\
&+ \left[1 - \Theta\left(\mathcal{E}_{\mathrm{CI}} - \mathcal{E}_{\mathrm{max}}\right)\right] p(\mathcal{E}_{\mathrm{CI}}/x|m, I), \\
=& \delta(\mathcal{E}_{\mathrm{CI}} - \mathcal{E}_{\mathrm{max}}) \left(1 - \frac{\Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma(m)}\right) \\
&+ \left[1 - \Theta\left(\mathcal{E}_{\mathrm{CI}} - \mathcal{E}_{\mathrm{max}}\right)\right] \frac{m^{m}}{\Gamma(m)\mathcal{E}_{\mathrm{CI},0}^{m}} \mathcal{E}_{\mathrm{CI}}^{-m-1} e^{-m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{CI}}}},
\end{aligned}
\tag{4.96}
$$

where $\delta(\cdot)$ is Dirac's delta function and $\Theta(\cdot)$ is Heaviside's theta function. The first term thus represents the occasions when the output power is not enough to invert all channel conditions, while the second term represents the invertible conditions for which $\mathcal{E}_{\mathrm{CI}} \propto 1/x_{\mathrm{f}}$. We can by the use of the definition of the incomplete gamma function $\Gamma(x, y)$ (Wolfram Research Inc., 2007), find the mean value of $p(\mathcal{E}_{\mathrm{CI}}|\mathcal{E}_{\mathrm{CI},0}, \mathcal{E}_{\mathrm{max}}, m, I)$ as

$$
\overline{\mathcal{E}}_{\mathrm{CI}} = \mathcal{E}_{\mathrm{max}} \frac{\Gamma\left(m\right) - \Gamma\left(m, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)} + m\mathcal{E}_{\mathrm{CI},0} \frac{\Gamma\left(m-1, m\frac{\mathcal{E}_{\mathrm{CI},0}}{\mathcal{E}_{\mathrm{max}}}\right)}{\Gamma\left(m\right)}.
\tag{4.97}
$$

∎

Appendix 4.G MQAM peak-to-average ratio

The peak-to-average power ratio of square MQAM is defined as the power of the outermost signal point divided by the average power over all signal constellation points. The possible constellation points in the IQ-plane[19] are $\pm 1, \pm 2, \ldots, \pm\sqrt{M}$, where $M = 2^b$ is the alphabet size and b is the number of bits per symbol. The ratio can be found by considering the north-east quadrant of the constellation with 2^{b-2} points. The sum $\mathcal{P}_{\text{sum}}$ of the squared Euclidian distances (that is the sum of the powers) in this quadrant is

$$\mathcal{P}_{\text{sum}} = \sum_{n=1}^{\sqrt{2^{b-2}}} \left[2(2n-1)^2 + 2\sum_{k=1}^{n-1} \left((2n-1)^2 + (2k-1)^2\right) \right], \tag{4.98}$$

where (n, k) is the coordinate of each point. The average power $\overline{\mathcal{P}} = \mathcal{P}_{\text{sum}}/2^{b-2}$ can, after some manipulations, or with the aid of Mathematica 6 (Wolfram Research Inc., 2007) be evaluated to

$$\overline{\mathcal{P}} = \frac{2}{3}\left(2^b - 1\right). \tag{4.99}$$

Now, the maximum power $\mathcal{P}_{\text{max}}$ is given by the outermost point and is

$$\mathcal{P}_{\text{max}} = 2\left(2\sqrt{2^{b-2}} - 1\right)^2. \tag{4.100}$$

Finally we get the peak-to-average ratio

$$Q_{\text{PAR}}(b) = \frac{\mathcal{P}_{\text{max}}}{\overline{\mathcal{P}}} = 3\frac{\sqrt{2^b} - 1}{\sqrt{2^b} + 1}. \tag{4.101}$$

Assuming that the efficiency of the amplifier is inversely proportional to $Q_{\text{PAR}}(b)$, we obtain

$$\frac{\epsilon(2)}{\epsilon(b)} = 3\frac{\sqrt{2^b} - 1}{\sqrt{2^b} + 1}. \tag{4.102}$$

For non-square constellations the expression is not valid, but we will use the above peak-to-average ratio for all rectangular constellations bearing in mind that we then underestimate the peak-to-average ratio for non-square constellations. We include the above effect mostly for the purpose of comparability with previous research results (for instance Cui et al. (2004) make use of (4.102)).

Appendix 4.H MQAM bit error rates

First, let us note that one factor influencing the bit error rate behaviour is the constellation design, that is the mapping of bits to signal points. Although it is possible to optimise the MQAM constellations, see for instance (Proakis, 2001,

---

[19]The In-phase and Quadrature plane.

Section 5.2.9), we here restrict the analysis to square Gray encoded constellations for simplicity.

The symbol error rate S for *square* MQAM, corresponding to even values of b, in the case of a time-invariant additive white Gaussian noise channel is (Goldsmith, 2005, p. 177)

$$S = 1 - \left(1 - \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \cdot \mathrm{erfc}\left(\sqrt{\frac{3}{2^b - 1} \frac{\gamma_s}{2}}\right)\right)^2 \qquad (4.103)$$

where γ_s is the signal-to-noise ratio per symbol. With Gray coded symbols and use of the approximation that one symbol error equals one bit error – this is a good approximation for large signal-to-noise ratios – we can write

$$B = \frac{1}{b}\left(1 - \left(1 - \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \cdot \mathrm{erfc}\left(\sqrt{\frac{3b}{2^b - 1} \frac{\gamma}{2}}\right)\right)^2\right) \qquad (4.104)$$

where we have used $\gamma_s = b\gamma$ with γ being the signal-to-noise ratio per bit. Expanding the square in (4.104) and neglecting second order terms $\mathrm{erfc}^2(\cdot)$ we obtain for $b \geq 2$

$$B = \frac{2}{b} \cdot \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \cdot \mathrm{erfc}\left(\sqrt{\frac{3b}{2^b - 1} \frac{\gamma}{2}}\right) \qquad (4.105)$$

which is exact for $b = 2$ and a good approximation for $b > 2$.

By the use of the first term

$$B_1 = \frac{2}{b} \cdot \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \cdot \frac{1}{\sqrt{\pi}} \sqrt{\frac{2^b - 1}{3b} \cdot \frac{2}{\gamma}} e^{-\frac{3b}{2^b - 1} \cdot \frac{\gamma}{2}}, \qquad (4.106)$$

in a Taylor series expansion of (4.106) we can solve for the signal-to-noise ratio per bit and obtain

$$\gamma(b, B) = \left[\frac{1}{3} W_{\mathrm{L}}\left(\frac{8}{\pi} \cdot \left(\frac{1 - 2^{-b/2}}{bB}\right)^2\right)\right] \frac{2^b - 1}{b}, \qquad (4.107)$$

where W_{L} denotes the Lambert W function. Observe that we here have, in the first factor, a quantification of the distance to the Shannon limit in (4.1); the rightmost factor has exactly the same scaling behaviour $(2^b - 1)/b$ as the Shannon limit (4.1) for the signal-to-noise ratio per bit. The distance to the Shannon limit is fairly constant with respect to the exponential behaviour of $(2^b - 1)/b$, and the optimisation results for MQAM will correspond closely with the Shannon-limit optimisation. For illustration, consider a bit error rate $B = 10^{-6}$. For $b = 2$ the signal-to-noise ratio in (4.107) is 8.8 dB from the Shannon-limit while it for $b = 20$ is 8.2 dB away: Compared to the 45 dB increase of the factor $(2^b - 1)/b$, this change of 0.6 dB is negligible.

Appendix 4.I   Average bit error for Nakagami channels

In the Nakagami fading case we calculate the bit error rate by the use of the gamma distribution $\mathrm{Gam}(m, \overline{\gamma})$ for the received signal-to-noise ratio, see Assumption 2.6 on page 29. Averaging the bit error rate in (4.106) we obtain (Wolfram Research Inc., 2007)

$$B(m, \overline{\gamma}) = (-1)^m \frac{2}{\sqrt{\pi}} \cdot \frac{1 - 2^{-\frac{b}{2}}}{b} \cdot \frac{\Gamma(1/2 + m)}{\Gamma(m)} \beta \left(-\frac{2m}{3\overline{\gamma}} \frac{2^b - 1}{b}, m, \frac{1}{2} - m \right), \quad (4.108)$$

where $\beta(z, x, y)$ denotes the incomplete beta function. If we for a moment ignore the fine details of our bit error rate expression, and concentrate on the interconnection between the average signal-to-noise ratio per bit $\overline{\gamma}$ and the number of bits b per symbol, we see that the former appears in the following company,

$$\overline{\gamma} \cdot \frac{b}{2^b - 1}. \quad (4.109)$$

Just like for the static channel and the outage criterion we thus find that the minimum energy per bit must scale approximately like the Shannon limit (4.1). There are of course differences, but a Shannon-type analysis actually captures the most prominent scaling effects and can serve as a good reference case.

Chapter 5

Polarisation Receiver Diversity

THERE has during the last two decades been a vital interest in spatial diversity/multiplexing techniques in the wireless communication community. Not that the idea is of spatial diversity is new, even Marconi used an antenna array in his 1901 transatlantic transmission (Bondyopadhyay, 2000), but the demand for robust high speed communication combined with important theoretical multiple-input multiple output (MIMO) results have boosted the interest. Spatial diversity techniques have also been proposed for the purpose of reducing the total energy consumption in energy limited wireless sensor networks. Because of the limited possibility to equip small nodes with antenna arrays the focus has been on cooperative techniques, see for instance the work by Laneman (2002), Laneman and Wornell (2003) on cooperative diversity, the work by Cui et al. (2004) on cooperative MIMO.

Curiously enough, the possibility to use polarisation diversity in place of spatially separated antennas has been largely overlooked in spite of a compact antenna configuration and the apparent suitability for small sensor nodes.[1] Perhaps there has been a suspicion that polarisations are not independent enough in wireless sensor network scenarios, but judging from our measurement results in Chapter 3 they are. Malik and Edwards (2007) consider the benefits of polarisation diversity with the motivation of compact design and reduced receiver complexity in an ultra wide-band system, but there is to the best of our knowledge no analysis of the overall energy-wise benefits of polarisation diversity. Of course, the transmission energy gains

---

[1]The use of polarisation diversity in mobile cellular networks was considered by Vaughan (1990) who found that considerable gains in signal-to-noise ratio were available. One can now see base stations using polarisation diversity in place of spatial diversity.

from dual branch receiver diversity will be similar for polarisation diversity and spatial diversity, but the important difference in the energy-limited context is the possibility to avoid cooperation and the increased processing costs by simple onboard receiver polarisation diversity. Cui et al. (2004) study the energy efficiency of array-antenna MIMO communications – nothing hinders the use of a MIMO scheme over polarised antennas – but consider only a complex receiver structure with large processing costs. We here assess simpler structures that can be used for polarisation receiver diversity (we study cooperative MIMO in the next chapter). Moreover, we include the degree of fading in our analysis since it has a considerable impact on the results.

In this chapter we treat *receiver* diversity only, but nothing prevents an extension to more than one transmit branch through a slight modification of our assay. We study three different reception diversity schemes, namely maximum ratio combining (MRC), selection diversity (SD) and switching (or threshold) diversity (SwD).[2] These techniques are illustrated in Figure 5.1 where we have applied them to a measurement at 868 MHz. Perfect channel state information at the receiver is assumed throughout our analysis. The works on dual branch receiver diversity by Simon and Alouini (1999) and Abu-Dayya and Beaulieu (1994) contain many of the formulas for bit error rates and outage probabilities that we use here but expressed in somewhat different terms. We also generalise their results in some cases.

5.1   Receiver diversity radiated-energy gains

The baseline comparison is between a single-branch (SB) system that uses one fixed receiver antenna polarisation, and a system where the receiver can use *two* differently polarised antennas.[3]

5.1.1   Nakagami-m channel assumptions

Our Nakagami-m probability assignment for the received envelope, see Assumption 2.6, amounts to a gamma assignment for the per-bit signal-to-noise ratios γ of each receiver branch.[4] Here m is the fading figure and $\overline{\gamma}$ is the

---

[2]Another technique commonly mentioned in this context is equal gain combining. We omit it here for brevity and note that it in terms of transmission performance reside between MRC and SD. The performance loss with respect to MRC is typically less than 1 dB (Goldsmith, 2005, p. 216–217)

[3]For simplicity we study only two branches although triple-polarised antennas can be used.

[4]We use, for notational convenience, the signal-to-noise power ratio $\gamma_{\mathcal{P}}$ and the signal-to-noise ratio per bit γ interchangeably in this chapter.
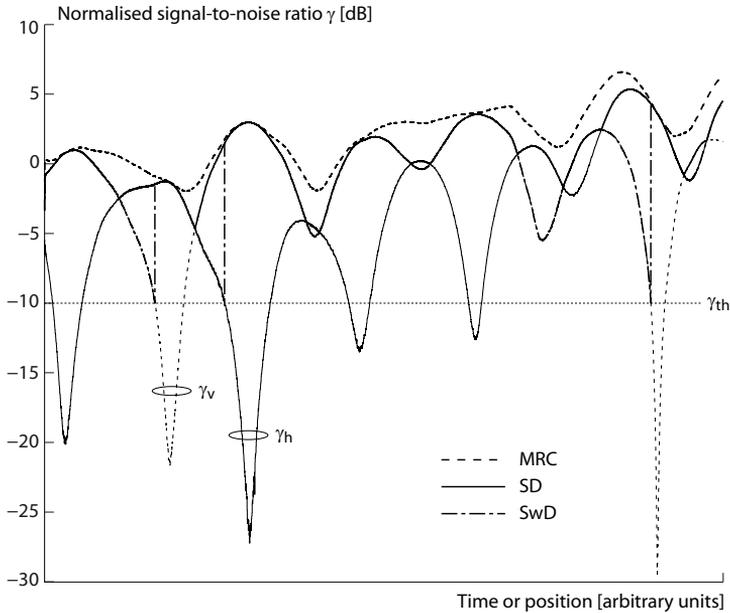
Figure 5.1: Polarisation diversity gains for non-line-of-sight indoor measurement at 868 MHz. The transmitter used horizontal polarisation, and the figure displays received signal-to-noise ratios γ_v and γ_h for vertical and horizontal receive polarisations. The maximum ratio combining (MRC) scheme makes an optimum combining of the two branches and therefore achieves the best signal-to-noise ratio. Selection diversity (SD) follows the best branch, while switched diversity (SwD) changes branch at a fixed threshold γ_{th}. Consequently, the switched scheme sometimes misses the largest signal-to-noise ratios. On the other hand, all three schemes avoid the bad conditions and thereby increase the robustness significantly.

average signal-to-noise ratio per bit. One concern in our comparison is the reference, single branch, performance when the two polarisation branches are of different quality.[5] Which "reference quality" should we use to assess the diversity gain? The best branch, the worst branch, or an average thereof? Either way, we might be considered unfair to one scheme or the other. Another concern regarding different branch qualities is that the analysis becomes much more involved and the resulting expressions unwieldy. We therefore introduce a simplifying assumption which we invoke for the major part of our analysis.

---

[5]The "quality" is not only determined by the average signal-to-noise ratio per bit $\overline{\gamma}$ but also by the fading degree as given by m.

Assumption 5.1 *Consider two Nakagami-m fading receiver branches (v) and (h) with respective fading figures m_v and m_h, and respective average signal-to-noise ratios per bit $\overline{\gamma}_v$ and $\overline{\gamma}_h$. Let I_I denote the following background information*

$$I_I \equiv \{m_v = m_h = m, \overline{\gamma}_v = \overline{\gamma}_h = \overline{\gamma}, c = 0\}, \tag{5.1}$$

c is the branch correlation coefficient defined in (3.5). ■

Our measurement results from Chapter 3 support the assumption that $c = 0$ quite strongly, but does not give support for the assumption on equal fading characteristics on the two polarisation branches. This condition will remove the reference-branch problem, and also simplifies the analysis. The use of Assumption 5.1, which is common in the literature, may lead to over estimation of the attainable energy savings, and this should be borne in mind. In Section 5.2.4 we analyse the effects of deviations from Assumptions 5.1 in terms of different per-bit branch signal-to-noise ratios, $\overline{\gamma}_v \neq \overline{\gamma}_h$.

5.1.2   Radiated-energy gains

The degree of fading and the performance requirements determine the attainable reduction G_D in the required energy per bit at the receiver. For outage and bit-error-rate criteria we will under I_I define the diversity gain

$$G_{D|I_I}(m, P_{\text{error}}) \equiv \frac{\overline{\gamma}_{\text{SB}}(m, P_{\text{error}})}{\overline{\gamma}_D(m, P_{\text{error}})} \tag{5.2}$$

where $\overline{\gamma}_{\text{SB}}$ and $\overline{\gamma}_D$ are the required average *branch-wise* signal-to-noise ratios per bit at the receiver for single branch (SB) and diversity (D) schemes respectively, m is the Nakagami fading parameter and P_{error} is the specified probability of error (typically outage P_{out} or bit error rate B). Recall from Chapter 3 and Definition 3.3 that the Nakagami-m channel has an inherent diversity order m with the corresponding error probability

$$P_{\text{error,SB}} \propto \overline{\gamma}^{-m}. \tag{5.3}$$

Based on our analysis in Appendix 5.A we find that the following assumption is well motivated for the analysis before us.

Assumption 5.2 *Under I_I in Assumption 5.1 the radiated energy gain from dual branch receiver polarisation diversity is*

$$G_{D|I_I}(m) = A_D(m) \frac{\overline{\gamma}_{\text{SB}}(m, P_{\text{error}})}{\overline{\gamma}_{\text{SB}}(\Delta m, P_{\text{error}})}, \tag{5.4}$$
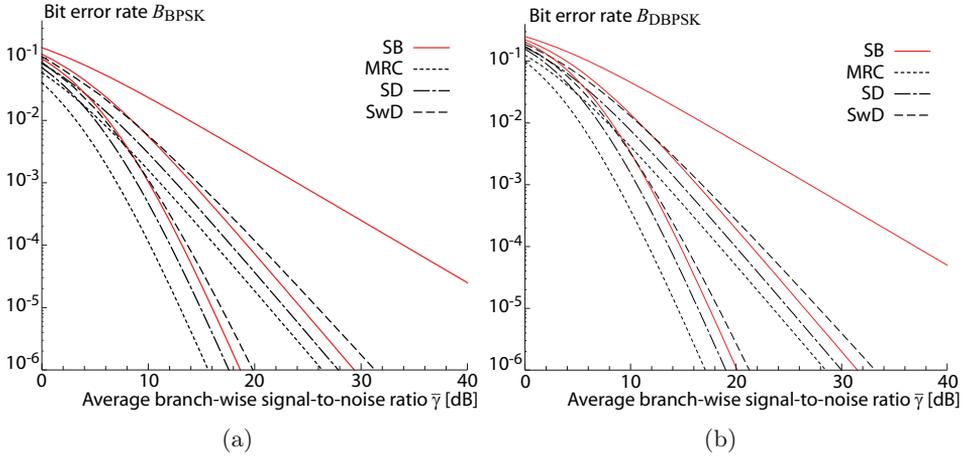
Figure 5.2: *Bit error rates for coherent and differential binary phase shift keying. The single branch performance (red solid lines) is shown for $m = 1, 2, 4$, while the performance of the diversity schemes are shown for $m = 1, 2$. It is evident that maximum ratio combining (black dotted lines), selection diversity (black dash-dotted lines) and switched diversity (black dashed lines) achieve almost identical diversity order gains $\Delta \approx 2$, but with different array gains that shift their curves horizontally.*

where $A_D(m) \leq 2$ is a diversity scheme dependent array gain and $\Delta = 2$ is the diversity order gain (see Definition 3.1 and Definition 3.4 respectively). Here, $\overline{\gamma}_{SB}(m)$ is the required signal-to-noise ratio per bit for a single branch receiver operating in Nakagami-m fading. ∎

Figure 5.2 illustrates that Assumption 5.2 holds with good accuracy for maximum ratio combining (MRC) and selection diversity (SD), but is slightly less accurate for switched diversity (SwD), for bit error rates $B \geq 10^{-6}$. The switched diversity scheme does not really attain diversity order gain $\Delta = 2$.

In the numerical results presented in this chapter, which for clarity is focused on maximum ratio combining(MRC) and switched diversity (SwD), we will use the following array gains:

Assumption 5.3 *The array gain A_D in (5.4) is for maximum ratio combining (MRC) and switched diversity (SwD), respectively,*

$$A_{D,MRC}(m) = 2$$
$$A_{D,SwD}(m) = \frac{3}{4}.$$

(5.5)

∎

5.2   Total energy consumption

A diversity receiver can reduce the required transmission energy per bit by making use of two received signals. The energy cost is that of increased overhead and/or receiver processing energy. The former can be required to allow switching between branches, and the latter could be the energy consumed by an additional receiver chain.

Maximum ratio combining and selection diversity require more receiver circuitry than a single branch scheme, and the additional receiver chain will consume processing energy. Let

$$\mathcal{E}_{\mathrm{Pr,D}} \equiv \text{ the diversity processing energy per bit} \tag{5.6}$$

required by the diversity scheme (D). This changes the receiver's processing energy to $\mathcal{E}_{\mathrm{Pr}} + \mathcal{E}_{\mathrm{Pr,D}}$. Switched diversity has the very attractive feature, at least for small energy-limited sensor nodes, of low hardware complexity. The switching consumes virtually no energy at all, $\mathcal{E}_{\mathrm{Pr,D}} = 0$, but in order to switch from a bad branch the receiver needs overhead bits. Abrupt switches in the middle of an ordinary packet reception would cause problems (Rustako et al., 1973). Abu-Dayya and Beaulieu (1994) analyse a scheme where the received signal-to-noise ratio is compared to the threshold γ_{th} only at certain discrete times, and under such a scheme "re-synchronisation bits" could be placed in the packets. Let

$$r_{\mathrm{D}} \equiv \text{ the fraction of diversity overhead,} \tag{5.7}$$

that is the amount of extra transmission time required.

Incorporation of receiver diversity (D) will save a fraction w of the total single branch (SB) per-bit energy consumption, where

$$\begin{aligned}
w &= \frac{\mathcal{E}_{\mathrm{tot,SB}} - \mathcal{E}_{\mathrm{tot,D}}}{\mathcal{E}_{\mathrm{tot,SB}}} \\
&= 1 - \frac{\mathcal{E}_{\mathrm{tot,D}}/\mathcal{E}_{\mathrm{RP}}}{\mathcal{E}_{\mathrm{tot,SB}}/\mathcal{E}_{\mathrm{RP}}} \\
&= 1 - \frac{(1 + r_{\mathrm{D}})\left[\frac{\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}} + \mathcal{E}_{\mathrm{Pr,D}} + \mathcal{E}_{\mathrm{T,D}}}{\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}}}\right]}{\frac{\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}} + \mathcal{E}_{\mathrm{T,SB}}}{\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}}}}.
\end{aligned} \tag{5.8}$$

We have in the last step included the increased transmission time through the diversity r_{D}, and the increased receive processing energy per bit through $\mathcal{E}_{\mathrm{Pr,D}}$. By the use of the transmission-to-total-processing ratio ρ'_{SB} in (2.8),

the receiver-to-transmitter processing ratio α in (2.9), the radiated-energy gain G_D in (5.2) and Mikami's degradation model in (2.14), we arrive at

$$
\begin{aligned}
w &= 1 - \frac{(1 + r_D)\left[1 + \frac{\alpha_D}{1+\alpha} + \frac{\rho'_{SB}}{G_D^{1/g}}\right]}{1 + \rho'_{SB}} \\
&= \frac{\rho'_{SB}}{1 + \rho'_{SB}}\left[1 - \frac{(1 + r_D)}{G_D^{1/g}} - \frac{r_D + \alpha'_D(1 + r_D)}{\rho'_{SB}}\right],
\end{aligned}
\tag{5.9}
$$

where g is the amplifier's efficiency degradation exponent and

$$
\begin{aligned}
\alpha_D &= \frac{\mathcal{E}_{Pr,D}}{\mathcal{E}_{Pt}}, \\
\alpha'_D &= \frac{\alpha_D}{1 + \alpha}.
\end{aligned}
\tag{5.10}
$$

For switched diversity we assume a fraction $r_{SwD} > 0$ of additional overhead is required to manage the switching, while the maximum ratio combining is assumed to perform well with $r_{MRC} = 0$. On the other hand, two receiver chains are required to do maximum ratio combining and we use $\alpha'_{MRC} > 0$, while $\alpha'_{SwD} = 0$ since the switching energy should be negligible.

Assumption 5.4 *Unless otherwise stated, the numerical results are calculated for*

$$
\begin{aligned}
r_{SwD} &= 0.03, \\
r_{MRC} &= 0, \\
\alpha'_{MRC} &= \frac{1}{6}, \\
\alpha'_{SwD} &= 0.
\end{aligned}
\tag{5.11}
$$

$\blacksquare$

The switched diversity overhead r_{SwD} corresponds to the use of every 32:nd bit, on the average, for switching purposes. Depending on the application and the packet lengths used, this may be reasonable or not. The maximum ratio processing energy α'_{MRC} amounts to 33 percent extra energy consumption in the receiver. Also here the relevance of this exact figure will depend on the application and its hardware. We discuss this further in Section 5.2.3.

5.2.1   Expected energy saving

Clearly, the benefits from receiver diversity are largest in severe fading and the radiated energy gain G_D depends on the Nakagami fading figure m, see Figure 5.2. Taking our uncertainty regarding m into account, the best estimate according to a quadratic loss function is the mean value of the probability distribution, see (2.69). This expected energy saving $\overline{w} \equiv E(w|I)$ is from (5.9)

$$\overline{w} = \frac{\rho'_{SB}}{1+\rho'_{SB}} \left[1 - (1+r_D)E\left(G_D^{-1/g}|I\right) - \frac{r_D + \alpha'_D(1+r_D)}{\rho'_{SB}} \right], \qquad (5.12)$$

where the expectation $E\left(G_D^{-1/g}|I\right)$ generally must be evaluated numerically. It is commonplace to use the variance of the distribution as a measure of accuracy, but in the present case the distribution for w is not symmetric and the variance can therefore be misleading. Let us instead complement the mean with the median w_{med}, see (2.70), the 90th and the 10th percentiles w_{90} and w_{10} to show the uncertainty in w. Thanks to the monotonicity of the radiated-energy gain $G_D(m)$, these values correspond to the median and the percentiles for our prior for m in (2.85),

$$\begin{aligned}
w_{10} &\leftrightarrow m_{90} = 8.83 \\
w_{med} &\leftrightarrow m_{med} = 1.88 \\
w_{90} &\leftrightarrow m_{10} = 1.10.
\end{aligned} \qquad (5.13)$$

The smaller the fading figure m, the larger the energy saving w.

5.2.2   Energy saving under bit error rate criterion

In Figure 5.3 and Figure 5.4 we display the energy savings for differential binary phase shift keying (DBPSK) under bit error rate criteria of $B = 10^{-3}$ and $B = 10^{-6}$ respectively. To begin with, we note that the behaviour is quite similar to the power control (channel inversion) scheme of Section 4.2; the bit error rate has large influence and there is a significant uncertainty in w stemming from the unknown fading figure m with its impact on the transmission energy per bit. Regarding the choice between maximum ratio combining (MRC) and switching diversity (SwD) there are two characteristics of importance.

- The processing penalty of switching diversity is considerably smaller than for maximum ratio combining. Hence, the penalty at small transmission-to-total-processing ratio ρ'_{SB} is smaller for the former of the two, leading to a smaller maximum energy loss.
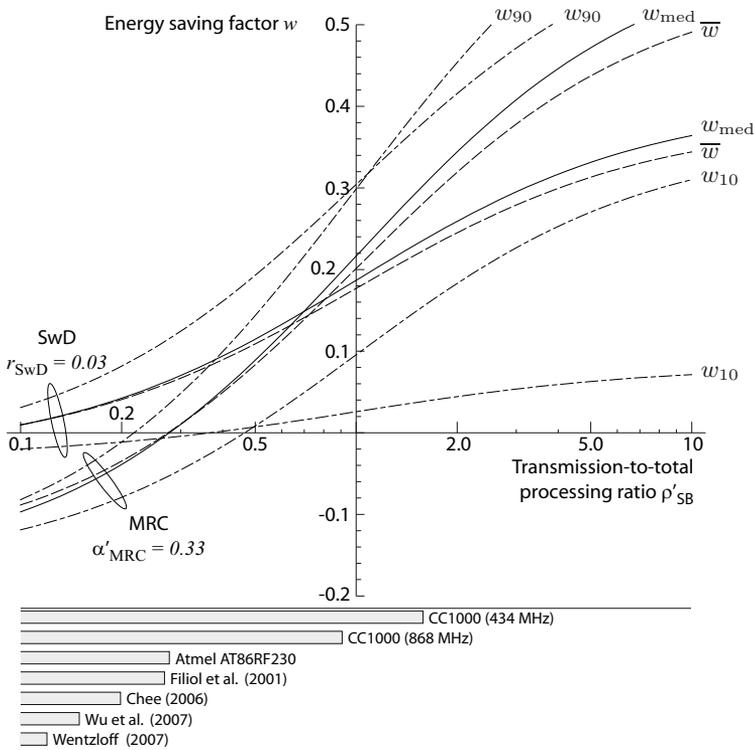
Figure 5.3: The fraction w of energy saved by exploitation of receiver po-larisation diversity for a target bit error rate $B = 10^{-3}$. The amplifier efficiency degradation exponent $g = 2$.

- The diversity performance of maximum ratio combining is superior to that of switching diversity and the reward comes at large transmission-to-total-processing ratios ρ'_{SB}. Then, the consistent performance out-weighs the larger "investment" in terms of processing energy. Switched diversity has significantly larger spread and even at large transmission-to-total-processing ratios ρ'_{SB} the degree of fading has significant impact.

For transmission-to-total-processing ratios $\rho'_{SB} > 1$ both schemes typically provide substantial expected energy savings. This is due to the poor performance of the single branch system in fading environments. Note that of the existing radios we consider, it is only the CC1000 that has an undisputable advantage of receiver diversity.

Remark 5.1 *Our results pertain to the use of a fixed link margin which is adjusted to the fading; it can be as much lower under a diversity scheme as*
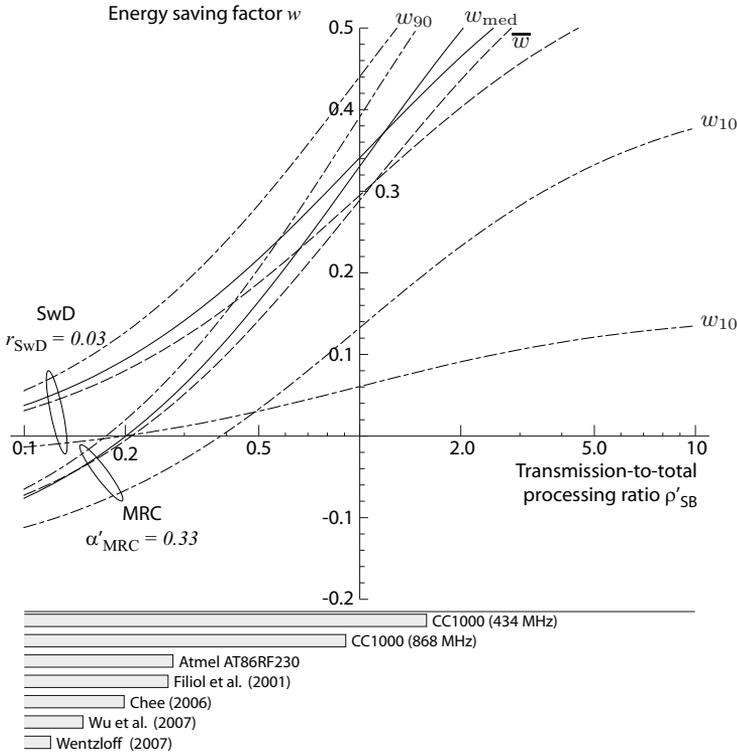
Figure 5.4: The fraction w of energy saved by exploitation of receiver po-larisation diversity for a target bit error rate $B = 10^{-6}$. The amplifier efficiency degradation exponent $g = 2$.

the radiated-energy gain G_D quantifies. Channel inversion through transmit power adaption is not considered in the present calculations. It is an alter-native, or a complement, to receiver diversity. We discuss this choice briefly in Section 5.3.

5.2.3 Maximum processing energy cost

This far, we have compared the diversity techniques with the single branch approach, but after consideration of the results presented in Figure 5.3 and Figure 5.4, a question presenting itself is "How much additional processing energy can we afford to pay for maximum ratio combining?" Let us compare the total *expected* energy consumption of the two diversity approaches,

$$(1 + r_\mathrm{SwD}) \left(1 + \rho'_\mathrm{SB} E\left(G_\mathrm{SwD}^{-1/g}|I\right)\right) = 1 + \alpha'_\mathrm{MRC} + \rho'_\mathrm{SB} E\left(G_\mathrm{MRC}^{-1/g}|I\right) \quad (5.14)$$
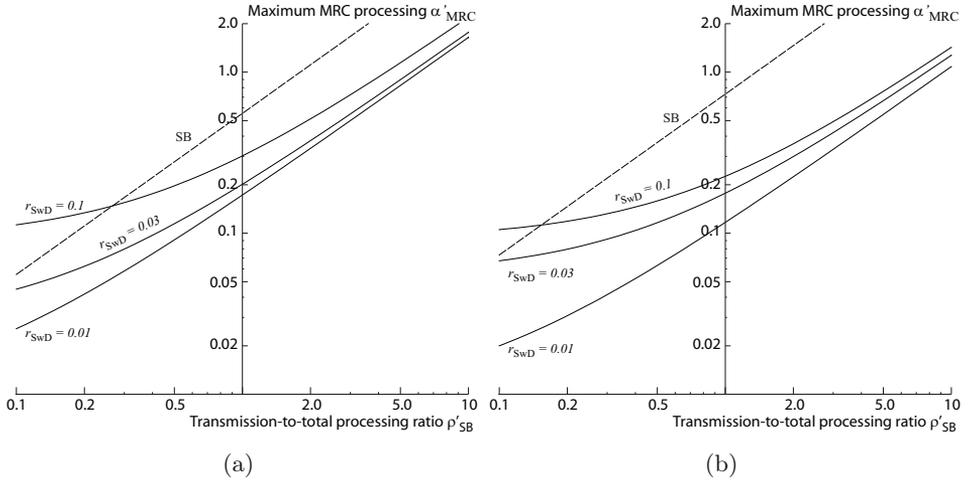
Figure 5.5: *Maximum additional normalised processing energy α'_{MRC} of maximum ratio combining. The left panel corresponds to $B = 10^{-3}$ and the right panel corresponds to $B = 10^{-6}$. The dashed curve corresponds to the single-branch receiver, while the solid curves correspond to the switched diversity receiver. The amplifier efficiency degradation exponent $g = 2$.*

where the left hand side is the (normalised) energy consumption of switched diversity while the right hand side is the counterpart for maximum ratio combining. Solving (5.14) for α'_{MRC}, see (5.10) and (5.11), we obtain

$$\alpha'_{\mathrm{MRC}} = r_{\mathrm{SwD}} + \left[(1 + r_{\mathrm{SwD}}) E\left(G_{\mathrm{SwD}}^{-1/g} | I \right) - E\left(G_{\mathrm{MRC}}^{-1/g} | I \right) \right] \rho'_{\mathrm{SB}}. \qquad (5.15)$$

By the use of $r_{\mathrm{SwD}} = 0$ and $G_{\mathrm{SwD}} = 1$, we obtain the single branch performance from and can thereby find the maximum α'_{MRC} with respect to the single branch receiver. In Figure 5.5 we show the boundaries between switched diversity and maximum ratio combining for different r_{SwD}, and also the boundary between maximum ratio combining and a single branch receiver. Remember, as a reference, that we previously assumed that $\alpha'_{\mathrm{MRC}} = 1/6 \approx 0.17$. From Figure 5.5 we draw the following conclusions.

- The allowed relative processing cost of maximum ratio combining, with respect to a single branch scheme, climbs steadily and is surprisingly large for transmission-to-total-processing ratios $\rho'_{\mathrm{SB}} > 2/3$. Then, the additional processing can approach half of the total, transmit and receive, single-branch processing, that is $\alpha'_{\mathrm{MRC}} = 1/2$.

- Maybe somewhat counter intuitively, the comparison with switched diversity shows that maximum ratio combining can afford more rel-

ative processing under the milder requirement $B = 10^{-3}$. This is because switched diversity provides a large radiated energy gain under the stricter requirement $B = 10^{-6}$, but this decreases significantly for $B = 10^{-3}$ and then the more consistent performance of maximum ratio combining pays off.

5.2.4   Deviations from the idealised assumptions

We have so far used Assumption 5.1 on equal branch fading characteristics to simplify the analysis and the choice of the reference single-branch parameters. Differences between the two polarisation branches' average signal-to-noise ratios $\overline{\gamma}_{\mathrm{v}}$ and $\overline{\gamma}_{\mathrm{h}}$, as well as positive correlations between the branches, have been ignored. If present, both of these effects will generally reduce the achievable diversity gains (Abu-Dayya and Beaulieu, 1994, Simon and Alouini, 1999). In the extreme cases of perfectly correlated branches, or infinitely different branch signal-to-noise ratios $\overline{\gamma}_{\mathrm{v}}/\overline{\gamma}_{\mathrm{h}} \to \infty$, the diversity schemes reduce to the single branch scheme since no diversity remains to exploit.

Average signal-to-noise ratios.   To begin with, we study the impact of different average branch signal-to-noise ratios, but still under independent probability distributions, and in the end we shortly discuss branch correlations. We use equal fading figures $m_{\mathrm{v}} = m_{\mathrm{h}} = m$ and restrict the analysis to differential binary phase shift keying. Nothing is changed in the maximum ratio scheme with respect to weighting and summing of the branch signals, it is only the probability distribution for the resulting signal-to-noise ratio γ_{MRC} that changes. The same holds also for the switched diversity scheme, except that the bit-error-rate optimal threshold γ_{th} will change and is no longer available. Let us assume for convenience that the single branch system uses the vertical polarisation and define the horizontal-to-vertical branch ratio

$$\nu \equiv \frac{\overline{\gamma}_{\mathrm{h}}}{\overline{\gamma}_{\mathrm{v}}}. \tag{5.16}$$

If $\nu < 1$, then the alternative branch h has a poorer average quality, and provides less average help than when $\nu = 1$. But it can still help significantly when the vertical-branch fading dips are deep, see for instance Figure 5.1.

Remark 5.2 *Even if the branches are static, $m \to \infty$, a single branch receiver can be stuck with the worst branch while a switched diversity receiver can choose the best one for a negligible energy cost. Hence, energy gains –*
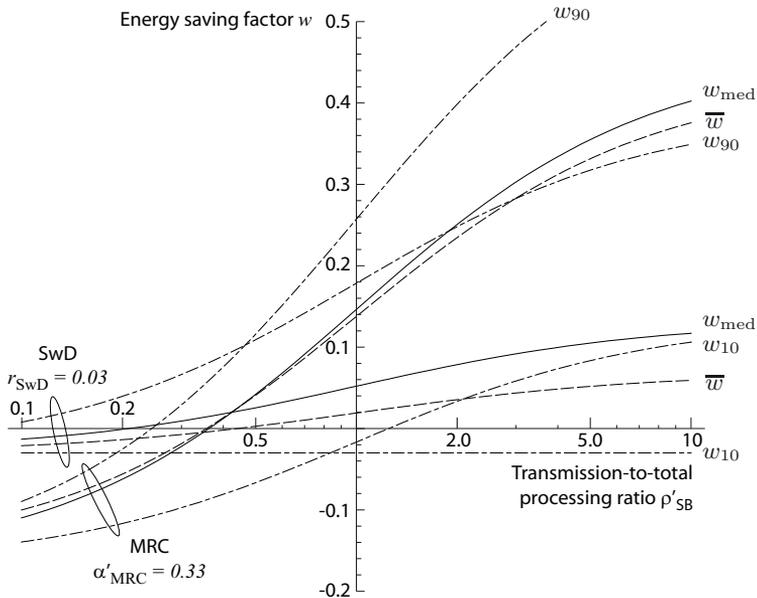
Figure 5.6: Expected energy savings for a horizontal-to-vertical branch ratio $\nu = 0.1$, that is a 10 dB difference in the branches' average signal-to-noise ratios. Target bit error rate $B = 10^{-3}$.

and robustness improvement – can actually be achieved even in non-fading environments.

In Figure 5.6 and Figure 5.7 we show the expected energy saving $\overline{w}$, accompanied by median and percentiles, for $B = 10^{-3}$ and $B = 10^{-6}$ respectively. All these results are found by numerical methods.

- As expected, the difference in branch signal-to-noise ratio decreases the benefits from receiver diversity. This is especially true for small degrees of fading, mild bit error rate requirements and switching receiver diversity, while the impact is less pronounced for maximum ratio combining in severe fading. In severe fading, the dips can be detrimental and any help is good help, but in less severe fading the risk is larger that the alternative branch is not above the dips.

- A distinctive effect is that the worst-case performance is significantly degraded for large transmission-to-processing ratios, because the chances to improve a less severely fading channel – larger fading figure m – diminishes when the alternative branch has a much smaller average gain. The uncertainty in w is increased substantially in this region for
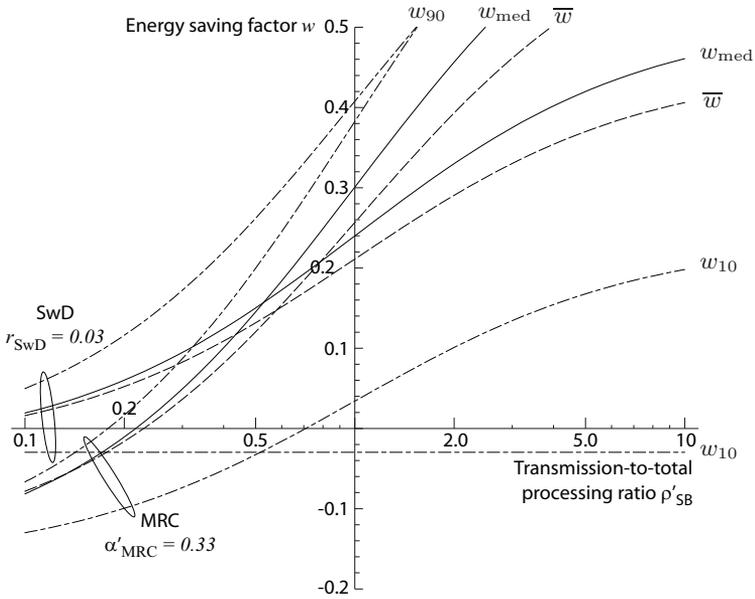
Figure 5.7: Expected energy savings for a horizontal-to-vertical branch ratio $\nu = 0.1$, that is a 10 dB difference in the branches' average signal-to-noise ratios. Target bit error rate $B = 10^{-6}$.

both maximum ratio combining and switched diversity; note that the 10th percentile w_{10} for switched diversity is virtually unaffected by the transmission-to-processing ratio.

Correlations. Regarding branch correlations we refer to the results by Simon and Alouini (1999, Fig. 4 and Fig. 5). Our conclusion based on their results, which include branch correlation coefficients between zero and 0.9, is that strong correlation has about the same impact as unequal branch qualities. The negative effects add up to larger overall penalty. The presence of an 0.9 branch correlation reduces, judging from the results by Simon and Alouini (1999), the gain by approximately 5 dB. Equivalently, horizontal-to-vertical ratio $\nu = 0.1$ yields a gain which is approximately 5 dB smaller than if $\nu = 1$. This gives us a coarse view of the impact of branch correlations on the energy efficiency of receiver diversity. Remember however the results in Chapter 3 which indicate that polarisations branches are practically uncorrelated; we need not worry too much about strong correlations.

5.3 Concluding remarks

Similarly to power control, receiver diversity will save no or little energy for transmission-to-total-processing ratios $\rho'_{\text{SB}} < 0.2$, but will become very attractive, from a total-energy perspective, for larger transmission-to-total-processing ratios $\rho'_{\text{SB}} > 1/2$. Some of our reference nodes shown in Figure 5.3 and Figure 5.4 would perhaps benefit from switched diversity but not maximum ratio combining; only the CC1000 radio is a candidate for the latter technique. Let us now consider transmission-to-total-processing ratios $\rho'_{\text{SB}} > 1/2$.

Channel inversion or polarisation diversity. Channel inversion and switched receiver diversity seem to provide similar energy savings in fading channels and the question then is which one to use. Our answer is: both. The use of receiver diversity reduces the dynamic range of the channel inversion scheme and could probably simplify the transmit power control, while the diversity schemes based on selection and switching relies on slow adjustment of the transmit power to reap the possible energy gains. Moreover, channel inversion can not replace receiver diversity in all respects. For instance, if the fading is very slow and one branch is in a dip, then the diversity receiver can likely switch out of the problem while the inversion scheme has to continue to struggle on high power. Diversity increases robustness.

You get what you pay for. Switched diversity carries the lightest overhead burden but does not achieve the consistent diversity performance that maximum ratio combining does at its larger processing cost. There is however a soft threshold around the transmission-to-total-processing ratio $\rho'_{\text{SB}} = 1$ at which maximum ratio combining becomes preferable to switched diversity.

Hardware complexity for MRC. Maximum ratio combining requires co-phasing of the branch signals, even if the subsequent detection is non-coherent. Therefore, as we did in Section 4.4.3 when considering adaptive MQAM, it could be motivated to study the increased processing cost of the coherent combining with respect to a fully non-coherent single branch (or switched) receiver. The maximum ratio combining processing energy α'_{MRC} will then increase, possibly by significant amounts, and might render the MRC approach much less attractive in spite of its consistent performance.

Who is paying the bill?   We have in this chapter considered the total energy consumption of a transmission, but the saving is actually achieved at the transmitter while largely paid for by the receiver. In a network where all nodes transmit and receive approximately equally often, this "load transfer" does not matter in the long run. But in a data gathering network – especially a single-hop network – most of the transmissions are from sensor nodes far away towards the central node. Clearly, nodes with no energy limitation should use maximum ratio combining to alleviate the burdens of the large mass of energy-limited sensor nodes sending data to them. In case some nodes are less energy-constrained than others – for instance in a heterogeneous network with some nodes capable of energy harvesting – the same approach could be used. The simpler sensor nodes can use switched diversity to increase robustness. These issues are further discussed in Chapter 8 when we study network sensing capacity.

Combine channel inversion, error correction and polarisation diversity.   Neither channel inversion or switched receiver diversity seem suited to counter fast fading, but this can be handled by the use of error correction. Our conclusion is that the combination of power control and receiver diversity to counter slow fading (see Definition 2.1 on page 30 for definitions of slow and fast fading), and error correcting codes to counter fast fading and noise is attractive and can be implemented successfully without great complexity (simple block codes, coarse grained power control and a suitable choice of receiver diversity).

Appendix 5.A   Diversity schemes details

All computations in this appendix have been carried out with the aid of Mathematica 6 (Wolfram Research Inc., 2007). For brevity, tedious derivations are not put in print here. Instead we focus on the most important results.

Appendix 5.A.1   Maximum ratio combining

Maximum ratio combining (MRC) amounts to summing the (co-phased) branch signals after weighting them so that the resulting signal-to-noise ratio γ_{MRCc} of the combined signals is maximised. The optimal weights are the branches' individual signal-to-noise amplitude-ratios $\sqrt{\gamma_{\mathrm{v}}}$ and $\sqrt{\gamma_{\mathrm{h}}}$, and that the resulting signal-to-noise ratio is simply the sum of the branches' signal-to-noise ratios (Goldsmith, 2005, p. 214). The average signal-to-noise after combining is therefore

$$\overline{\gamma}_{\mathrm{MRCc}} = \overline{\gamma}_{\mathrm{v}} + \overline{\gamma}_{\mathrm{h}}. \tag{5.17}$$

Under Assumption 5.1 this becomes $\overline{\gamma}_{\mathrm{MRCc}|I_{\mathrm{I}}} = 2\overline{\gamma}$; the 3 dB array gain of maximum ratio combining. This gain is independent of the nature of the fading.

There is no closed form solution for the distribution for γ_{MRCc} in the general case with different fading parameters for the two branches. However, if I_{I} is true, then we obtain a gamma probability distribution also for γ_{MRCc} (Wolfram Research Inc., 2007) and

$$m_{\mathrm{MRCc}|I_{\mathrm{I}}} = m_{\mathrm{v}} + m_{\mathrm{h}} \tag{5.18}$$

Under Assumption 5.1, $m_{\mathrm{v}} = m_{\mathrm{h}} = m$ and therefore $m_{\mathrm{MRCc}|I_{\mathrm{I}}} = 2m$. In addition to the array gain (5.17) we thus obtain a diversity order gain

$$\Delta_{\mathrm{MRC}} = 2, \tag{5.19}$$

which is exact.[6]

Radiated-energy gain.   Dual branch maximum ratio combining achieves according to (5.17) and (5.18) an array gain $A_{\mathrm{D}}(m) = 2$ and a diversity order gain $\Delta = 2$, see Definition 3.1 and Definition 3.4. It therefore has the exact radiated-energy gain

$$G_{\mathrm{MRC}|I_{\mathrm{I}}} = 2\frac{\overline{\gamma}_{\mathrm{SB}}(m, P_{\mathrm{error}})}{\overline{\gamma}_{\mathrm{SB}}(2m, P_{\mathrm{error}})}. \tag{5.20}$$

Appendix 5.A.2   Selection diversity

A selection diversity (SD) receiver monitors the branches' signal-to-noise ratios and selects the one with the best instantaneous signal-to-noise power ratio. The instantaneous "combined" signal-to-noise ratio is, ideally,

$$\gamma_{\mathrm{SDc}} = \max(\gamma_{\mathrm{v}}, \gamma_{\mathrm{h}}), \tag{5.21}$$

---

[6]In fact, the result in (5.18) holds also under the weaker condition that $m_{\mathrm{v}}/\overline{\gamma}_{\mathrm{v}} = m_{\mathrm{h}}/\overline{\gamma}_{\mathrm{h}}$ (Wolfram Research Inc., 2007).

see Figure 5.1. By the use of (2.109) on page 70 and the assumption of independent branch distributions we have

$$P_{\mathrm{SD}}(\gamma_{\mathrm{SDc}} < \gamma'|I) = \left[1 - \frac{\Gamma\left(m_{\mathrm{v}}, \frac{m_{\mathrm{v}}}{\overline{\gamma}_{\mathrm{v}}}\gamma'\right)}{\Gamma(m_{\mathrm{v}})}\right]\left[1 - \frac{\Gamma\left(m_{\mathrm{h}}, \frac{m_{\mathrm{h}}}{\overline{\gamma}_{\mathrm{h}}}\gamma'\right)}{\Gamma(m_{\mathrm{h}})}\right]. \tag{5.22}$$

This is the product of the independent probabilities that each branch is below γ'. Differentiation with respect to γ', and a subsequent substitution $\gamma' \to \gamma_{\mathrm{SD}}$, yield the probability density function

$$\begin{aligned}
p(\gamma_{\mathrm{SDc}}|I) &= \left[1 - \frac{\Gamma\left(m_{\mathrm{v}}, \frac{m_{\mathrm{v}}}{\overline{\gamma}_{\mathrm{v}}}\gamma_{\mathrm{SD}}\right)}{\Gamma(m_{\mathrm{v}})}\right]\frac{m_{\mathrm{h}}^{m_{\mathrm{h}}}}{\Gamma(m_{\mathrm{h}})\overline{\gamma}^{m_{\mathrm{h}}}}\gamma_{\mathrm{SD}}^{m_{\mathrm{h}}-1}e^{-\frac{m_{\mathrm{h}}}{\overline{\gamma}}\gamma_{\mathrm{SD}}} \\
&+ \left[1 - \frac{\Gamma\left(m_{\mathrm{h}}, \frac{m_{\mathrm{h}}}{\overline{\gamma}_{\mathrm{h}}}\gamma_{\mathrm{SD}}\right)}{\Gamma(m_{\mathrm{h}})}\right]\frac{m_{\mathrm{v}}^{m_{\mathrm{v}}}}{\Gamma(m_{\mathrm{v}})\overline{\gamma}^{m_{\mathrm{v}}}}\gamma_{\mathrm{SD}}^{m_{\mathrm{v}}-1}e^{-\frac{m_{\mathrm{v}}}{\overline{\gamma}}\gamma_{\mathrm{SD}}}.
\end{aligned} \tag{5.23}$$

Averaging (5.23) over γ_{SDc} results in the expected signal-to-noise ratio

$$\begin{aligned}
\overline{\gamma}_{\mathrm{SDc}} &= \overline{\gamma}_{\mathrm{v}} + \frac{(-1)^{-m_{\mathrm{v}}}\beta\left(-\frac{m_{\mathrm{v}}\overline{\gamma}_{\mathrm{h}}}{m_{\mathrm{h}}\overline{\gamma}_{\mathrm{v}}}, m_{\mathrm{v}}+1, -m_{\mathrm{v}}-m_{\mathrm{h}}\right)\Gamma(m_{\mathrm{v}}+m_{\mathrm{h}}+1)\overline{\gamma}_{\mathrm{v}}}{m_{\mathrm{v}}\Gamma(m_{\mathrm{v}})\Gamma(m_{\mathrm{h}})} \\
&+ \overline{\gamma}_{\mathrm{h}} + \frac{(-1)^{-m_{\mathrm{h}}}\beta\left(-\frac{m_{\mathrm{h}}\overline{\gamma}_{\mathrm{v}}}{m_{\mathrm{v}}\overline{\gamma}_{\mathrm{h}}}, m_{\mathrm{h}}+1, -m_{\mathrm{v}}-m_{\mathrm{h}}\right)\Gamma(m_{\mathrm{v}}+m_{\mathrm{h}}+1)\overline{\gamma}_{\mathrm{h}}}{m_{\mathrm{h}}\Gamma(m_{\mathrm{v}})\Gamma(m_{\mathrm{h}})}
\end{aligned} \tag{5.24}$$

This simplifies under Assumption 5.1,

$$\overline{\gamma}_{\mathrm{SDc}|\mathrm{I_I}} = \overline{\gamma}\left(2 + 4\frac{(-1)^{-m}\beta(-1, 1+m, -2m)\Gamma(2m)}{\Gamma^2(m)}\right). \tag{5.25}$$

The bracketed expression – the array gain $A_{\mathrm{D}}(m)$ corresponding to Definition 3.1 – is monotonically decreasing from $A_{\mathrm{SD}} = 3/2$ when $m = 1$ to $A_{\mathrm{SD}} = 1$ when $m \to \infty$. The array gain is smaller than for maximum ratio combining due to the fact that only one branch at a time is used.

Selection diversity error probability

Under Assumption 5.1 we obtain from (5.22) the outage probability

$$P_{\mathrm{SD}}(\gamma_{\mathrm{SDc}} < \gamma_{\mathrm{out}}|I) = P_{\mathrm{SB}}^2(\gamma < \gamma_{\mathrm{out}}|I), \tag{5.26}$$

where γ is the branch-wise signal-to-noise ratio. By the use of the series expansion in (2.110), we obtain

$$P_{\mathrm{out,SD}|I_{\mathrm{I}}} \propto \overline{\gamma}^{-2m}, \tag{5.27}$$

a diversity order gain $\Delta = 2$ as given by Definition 3.4.

Regarding bit error rates we first note that basically all modulations have at worst exponential error behaviour over a static channel for large signal-to-noise ratios per bit; $B \leq C_1 e^{-C_2 \gamma}$ for some constants $C_1 > 0$ and C_2. Specifically for differential binary phase shift keying we have

$$B_{\mathrm{sta,DBPSK}} = \frac{1}{2} e^{-\gamma}. \tag{5.28}$$

Assuming perfect receiver channel state information we can average the static channel bit error rate over the probability density function in (5.23) with the aid of Mathematica 6 (Wolfram Research Inc., 2007), giving the result

$$\overline{B}_{\mathrm{DBPSK}|I_{\mathrm{I}}} = \left(\frac{m}{m + \overline{\gamma}} \right)^m - \frac{\Gamma(2m) \, {}_2F_1 \left(m, 2m; m+1; -\frac{m+\overline{\gamma}}{m} \right)}{m \Gamma(m)^2}. \tag{5.29}$$

A series expansion for $\overline{\gamma} \gg m$ gives

$$\overline{B}_{\mathrm{DBPSK}|I_{\mathrm{I}}} \propto \overline{\gamma}^{-2m}. \tag{5.30}$$

For all modulations with exponential error characteristics over a static channel this relation will hold.

The selection scheme thus achieves, under Assumption 5.1, a diversity order gain $\Delta = 2$,

$$P_{\mathrm{error}} \propto \overline{\gamma}^{-2m}, \tag{5.31}$$

for both outage and bit error probabilities.

Radiated-energy gain. With a diversity order gain $\Delta = 2$ and an array gain $A_{\mathrm{D}}(m)$ corresponding to (5.25) the radiated-energy gain of selection diversity becomes

$$G_{\mathrm{SD}|I_{\mathrm{I}}} = A_{\mathrm{D}}(m) \frac{\overline{\gamma}_{\mathrm{SB}}(m, P_{\mathrm{error}})}{\overline{\gamma}_{\mathrm{SB}}(2m, P_{\mathrm{error}})}, \tag{5.32}$$

where $A_{\mathrm{D}}(m) \leq 3/2$.

Appendix 5.A.3 Switched diversity

The switched diversity scheme addresses the hardware complexity problem by switching a single receiver chain between, in the present case, two different antennas. Switches occur when the current signal-to-noise ratio drops below a predefined threshold value γ_{th}.[7] Deep fades are thereby avoided since the other branch is likely better, but the scheme will not always use the best branch. See Figure 5.1 for an illustration.

---

[7] It is typically assumed that the noise level is the same for both branches and the threshold value can be defined in terms of total received signal power. This simplifies the implementation.

Switching rule. Several switching rules can be implemented, but we here anal-yse the switching strategy outlined by Abu-Dayya and Beaulieu (1994, eq. (3a) and (3b)). Let

$$V_t \equiv \text{Vertical polarisation is used at time } t$$
$$H_t \equiv \text{Horizontal polarisation is used at time } t \tag{5.33}$$

and let γ_{v_t} and γ_{h_t} denote the signal-to-noise ratios at time t. The receiver's choice can now be expressed in logic form,[8] and it will at any time t choose

$$H_t \quad \text{iff} \quad H_{t-1}(\gamma_{h_t} \geq \gamma_{th}) + V_{t-1}(\gamma_{v_t} < \gamma_{th}) \tag{5.34}$$
$$V_t \quad \text{iff} \quad V_{t-1}(\gamma_{v_t} \geq \gamma_{th}) + H_{t-1}(\gamma_{h_t} < \gamma_{th}) \tag{5.35}$$

where "iff" is shorthand for "if and only if". The resulting signal-to-noise ratio γ_{SwDc} thus equals γ_{v_t} when V_t is true and γ_{h_t} when H_t is true.

From the switching rule in (5.33) and the rules of probability theory it follows that the probability for the resulting signal-to-noise ratio γ_{SwDc} to be below γ' is

$$
\begin{aligned}
P(\gamma_{\text{SwDc}} < \gamma'|I) &= P(V_t(\gamma_{v_t} < \gamma')|I) + P(H_t(\gamma_{h_t} < \gamma')|I) \\
&= P(V_t|(\gamma_{v_t} < \gamma')I)P(\gamma_{v_t} < \gamma'|I) + \\
&\quad P(H_t|(\gamma_{h_t} < \gamma')I)P(\gamma_{h_t} < \gamma'|I) \\
&= \{\text{Switching rule}\} \\
&= [P(V_{t-1}(\gamma_{v_t} \geq \gamma_{th})|(\gamma_{v_t} < \gamma')I) + \\
&\quad + P(H_{t-1}(\gamma_{h_t} < \gamma_{th})|(\gamma_{v_t} < \gamma')I)] \, P(\gamma_{v_t} < \gamma'|I) + \\
&\quad [P(H_{t-1}(\gamma_{h_t} \geq \gamma_{th})|(\gamma_{h_t} < \gamma')I) + \\
&\quad + P(V_{t-1}(\gamma_{v_t} < \gamma_{th})|(\gamma_{h_t} < \gamma')I)] \, P(\gamma_{h_t} < \gamma'|I).
\end{aligned} \tag{5.36}
$$

There will now be a special case, namely the case $\gamma_{th} \geq \gamma'$, when some probabilities are zero. Let us begin with this special case and use $I' = (\gamma_{th} \geq \gamma')I$.

$$
\begin{aligned}
P(\gamma_{\text{SwDc}} < \gamma'|I') &= P(H_{t-1}(\gamma_{h_t} < \gamma_{th})|(\gamma_{v_t} < \gamma')I)P(\gamma_{v_t} < \gamma'|I) + \\
&\quad P(V_{t-1}(\gamma_{v_t} < \gamma_{th})|(\gamma_{h_t} < \gamma')I)P(\gamma_{h_t} < \gamma'|I) \\
&= P(\gamma_{h_t} < \gamma_{th}|H_{t-1}(\gamma_{v_t} < \gamma')I)P(H_{t-1}|(\gamma_{v_t} < \gamma')I) \\
&\quad \times P(\gamma_{v_t} < \gamma'|I) + \\
&\quad P(\gamma_{v_t} < \gamma_{th}|V_{t-1}(\gamma_{h_t} < \gamma')I)P(V_{t-1}|(\gamma_{h_t} < \gamma')I) \\
&\quad \times P(\gamma_{h_t} < \gamma'|I) \\
&= \{\text{Independence assumptions}\} \\
&= P(\gamma_{h_t} < \gamma_{th}|I)P(H_{t-1}|I)P(\gamma_{v_t} < \gamma'|I) + \\
&\quad P(\gamma_{v_t} < \gamma_{th}|I)P(V_{t-1}|I)P(\gamma_{h_t} < \gamma'|I).
\end{aligned} \tag{5.37}
$$

---

[8] AND is represented by a logical product AB, while OR is represented by a logical sum $A + B$.

If $\gamma_{\mathrm{th}} < \gamma'$ we must add the terms that cancelled in the previous case. Let us use $I'' = (\gamma_{\mathrm{th}} < \gamma')I$.

$$
\begin{aligned}
P(\gamma_{\mathrm{SwDc}} < \gamma'|I'') &= P(\gamma < \gamma'|I') + \\
&\quad P(V_{t-1}(\gamma_{\mathrm{v}_t} \geq \gamma_{\mathrm{th}})|(\gamma_{\mathrm{v}_t} < \gamma')I)P(\gamma_{\mathrm{v}_t} < \gamma'|I) + \\
&\quad P(H_{t-1}(\gamma_{\mathrm{h}_t} \geq \gamma')|(\gamma_{\mathrm{h}_t} < \gamma')I)P(\gamma_{\mathrm{h}_t} < \gamma'|I) \\
&= P(\gamma < \gamma'|I') + \\
&\quad P(V_{t-1}|(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{v}_t} < \gamma')I)P(\gamma_{\mathrm{v}_t} \geq \gamma_{\mathrm{th}}|(\gamma_{\mathrm{v}_t} < \gamma')I) \\
&\quad \times P(\gamma_{\mathrm{v}_t} < \gamma'|I) + \\
&\quad P(H_{t-1}|(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{h}_t} < \gamma')I)P(\gamma_{\mathrm{h}_t} \geq \gamma_{\mathrm{th}}|(\gamma_{\mathrm{h}_t} < \gamma')I) \\
&\quad \times P(\gamma_{\mathrm{h}_t} < \gamma'|I) \\
&= \{\text{Independence assumptions}\} \\
&= P(\gamma < \gamma'|I') + \\
&\quad P(V_{t-1}|I)P(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{v}_t} < \gamma'|I) + \\
&\quad P(H_{t-1}|I)P(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{h}_t} < \gamma'|I).
\end{aligned}
\tag{5.38}
$$

To move forward we need to determine $P(V_t|I)$ and $P(H_t|I)$. They will follow the same type of recurrence equation, namely

$$
\begin{aligned}
P(V_t|I) &= P(\gamma_{\mathrm{v}_t} \geq \gamma_{\mathrm{th}}|V_{t-1}I)P(V_{t-1}|I) + P(\gamma_{\mathrm{h}_t} < \gamma_{\mathrm{th}}|H_{t-1}I)P(H_{t-1}|I) \\
&= \{\text{Independence assumptions}\} \\
&= P(\gamma_{\mathrm{v}_t} \geq \gamma_{\mathrm{th}}|I)P(V_{t-1}|I) + P(\gamma_{\mathrm{h}_t} < \gamma_{\mathrm{th}}|I)(1 - P(H_{t-1}|I)).
\end{aligned}
$$

Solving for the stationary solution we set $P(V_t|I) = P(V_{t-1}|I)$ and obtain

$$
P(V_t|I) = \frac{P(\gamma_{\mathrm{h}_t} < \gamma_{\mathrm{th}}|I)}{P(\gamma_{\mathrm{h}_t} < \gamma_{\mathrm{th}}|I) + P(\gamma_{\mathrm{v}_t} < \gamma_{\mathrm{th}}|I)}.
\tag{5.39}
$$

Identically, for H_t we have

$$
P(H_t|I) = \frac{P(\gamma_{\mathrm{v}_t} < \gamma_{\mathrm{th}}|I)}{P(\gamma_{\mathrm{v}_t} < \gamma_{\mathrm{th}}|I) + P(\gamma_{\mathrm{h}_t} < \gamma_{\mathrm{th}}|I)}.
\tag{5.40}
$$

Collecting (5.36)-(5.40) in one expression we finally have

$$
\begin{aligned}
P(\gamma_{\mathrm{SwDc}} < \gamma'|I) &= \frac{1}{P(\gamma_{\mathrm{v}_{t-1}} < \gamma_{\mathrm{th}}|I) + P(h_{t-1} < \gamma_{\mathrm{th}}|I)} \times \\
&\quad \Big[P(\gamma_{\mathrm{h}_{t-1}} < \gamma_{\mathrm{th}}|I)P(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{v}_t} < \gamma'|I) + \\
&\quad P(\gamma_{\mathrm{v}_{t-1}} < \gamma_{\mathrm{th}}|I)P(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{h}_t} < \gamma'|I) + \\
&\quad P(\gamma_{\mathrm{h}_{t-1}} < \gamma_{\mathrm{th}}|I)P(\gamma_{\mathrm{v}_t} < \gamma_{\mathrm{th}}|I)P(\gamma_{\mathrm{h}_t} < \gamma'|I) + \\
&\quad P(\gamma_{\mathrm{v}_{t-1}} < \gamma_{\mathrm{th}}|I)P(\gamma_{\mathrm{h}_t} < \gamma_{\mathrm{th}}|I)P(\gamma_{\mathrm{v}_t} < \gamma'|I) \Big]
\end{aligned}
\tag{5.41}
$$

where it is understood that $P(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{v}_t} < \gamma'|I) = 0$ and $P(\gamma_{\mathrm{th}} \leq \gamma_{\mathrm{h}_t} < \gamma'|I) = 0$ when $\gamma_{\mathrm{th}} \geq \gamma'$. By the use of (2.109) on page 70, each of the probabilities in (5.41) can be expressed as a Nakagami single branch outage probability.

The probability density function $p(\gamma_{\mathrm{SwDc}}|I)$ for γ_{SwDc} is found by differentiation of (5.41), but the expression becomes quite large and we choose not to display it here since no insights are gained from it. However, the average signal-to-noise ratio can under Assumption 5.1 be expressed as

$$\overline{\gamma}_{\mathrm{SwDc}|I_\mathrm{I}} = \overline{\gamma} \left(1 + \frac{\left(m \frac{\gamma_{\mathrm{th}}}{\overline{\gamma}} \right)^m e^{-m \frac{\gamma_{\mathrm{th}}}{\overline{\gamma}}}}{\Gamma\left(1 + m \right)} \right). \tag{5.42}$$

The bracketed factor has its maximum for $\gamma_{\mathrm{th}} = \overline{\gamma}$, and this maximum is $1 + m^m \exp(-m)/\Gamma(1+m) > 1$. However, this does not imply that we should use the threshold $\gamma_{\mathrm{th}} = \overline{\gamma}$. The performance metrics – bit error rate or outage probability – are typically highly non-linear functions of γ_{th} and in general the optimum threshold depends both on m, $\overline{\gamma}$ and the performance requirement. An example is given in (5.46) below.

Switched diversity error probability

Observe that if we insert $\gamma_{\mathrm{th}} = \gamma'$ in (5.41), the above expression reduces to

$$\begin{aligned} P(\gamma_{\mathrm{SwD}} < \gamma'|(\gamma_{\mathrm{th}} = \gamma')I) &= P(\gamma_{\mathrm{v}_t} < \gamma'|I)P(\gamma_{\mathrm{h}_t} < \gamma'|I) \\ &= P(\gamma_{\mathrm{SD}} < \gamma'|I), \end{aligned} \tag{5.43}$$

see (5.26). Consequently, from an outage point of view we should follow the intuition telling us that the threshold should be at the outage level since we never would like to follow one branch down into a fading dip without trying the other branch. Since the outage performance of switched diversity equals that of selection diversity we have from (5.27) that

$$P_{\mathrm{out,SwD}|I_\mathrm{I}} \propto \overline{\gamma}^{-2m}. \tag{5.44}$$

The average bit error rate achieved with the switching diversity is found from averaging the static channel bit error rate over $p(\gamma_{\mathrm{SwD}}|I)$, and under Assumption 5.1 the obtained expression simplifies to

$$\overline{B}_{\mathrm{SwD,DBPSK}|I_\mathrm{I}} = \frac{\left(\frac{m}{m+\overline{\gamma}} \right)^m \left(\Gamma(m) - \Gamma\left(m, \frac{m\gamma\mathrm{th}}{\overline{\gamma}} \right) + \Gamma\left(m, \frac{\gamma\mathrm{th}(m+\overline{\gamma})}{\overline{\gamma}} \right) \right)}{2\Gamma(m)}. \tag{5.45}$$

The optimum threshold γ_{th} is obtained by setting the derivative to zero and solving for γ_{th};

$$\gamma_{\mathrm{th,opt}|I_\mathrm{I}} = m \ln \left(\frac{m+\overline{\gamma}}{m} \right). \tag{5.46}$$

But even with this optimum threshold, a series expansion of (5.45 for $\overline{\gamma} \gg m$ reveals that switched diversity does not achieve a diversity order gain of two, $\Delta < 2$ (Wolfram Research Inc., 2007).

It turns out that switched diversity does not achieve the full diversity order gain of two. On the other hand, as illustrated in Figure 5.2(a) and Figure 5.2(b), the gain in diversity order is close to $\Delta = 2$ for bit error rates down to $B = 10^{-6}$. Within this range we will therefore assume that $\Delta = 2$ for switched diversity, bearing in mind that this is not true if harder error requirements are posed.

Received-energy gain. Although it is possible to attain an array gain $A_{\mathrm{D}}(m) > 1$ with a switched receiver, see (5.42), adapting the threshold to the performance requirement will sacrifice part of this array gain to improve the error behaviour, see Figure 5.2(a) and Figure 5.2(b). Our approximate radiated-energy gain is

$$G_{\mathrm{SwD}|I_{\mathrm{I}}} = A_{\mathrm{SwD}}(m)\frac{\overline{\gamma}_{\mathrm{SB}}(m, P_{\mathrm{error}})}{\overline{\gamma}_{\mathrm{SB}}(2m, P_{\mathrm{error}})}, \tag{5.47}$$

where $A_{\mathrm{D}}(m) < 1$.

Chapter 6

Multi-hop Communication

IN THE research literature on wireless sensor networks one assumes almost without exception that the networks will have a multi-hop structure in which data is forwarded hop-by-hop toward the final destination. This structure is not seldom motivated by its alleged energy efficiency stemming from the fact that the required (average) transmission energy per bit increases super-linearly with distance, giving us the chance to save energy by forwarding data over many short hops. More specifically, under the common power-law propagation loss model which we have adopted through Assumption 2.8 on page 30, the required radiated energy per bit scales like

$$\overline{\mathcal{E}}_{\mathrm{rad}} \propto \left(\frac{d}{d_0} \right)^{\kappa}. \tag{6.1}$$

The model states that the required energy per bit is proportional to the κth power of the relative distance d/d_0. In (6.1), κ is the environment-dependent propagation loss exponent which typically is in the range $[2, 5]$, see Section 2.4.2 or Section 2.B.2. It hence seems possible to reduce the transmission energy by a factor $N^{\kappa}/N = N^{\kappa-1}$ by doing N short hops of length d/N instead of one long hop. Judging from this the energy per bit could be reduced to almost negligible values. For instance, if $\kappa = 4$ and $N = 10$ the reduction in transmission energy is $10^{4-1} = 1000$; a promising thousand-fold reduction.

Our obvious objection to the above line of reasoning is that it ignores processing energy, and the inclusion of processing energy unfortunately erases most of the possible energy savings from multi-hopping. Inspired by Min and Chandrakasan (2003), who put the energy efficiency of multi-hopping in doubt by an insightful example, we have previously shown that sensor

networks will rarely save energy through the use of multiple hops (Björnemo et al., 2006). We found that the propagation losses had to exceed the transmission range of current sensor node radios in order to motivate multi-hopping energy-wise, and we thereby confirmed the suspicion raised by Min and Chandrakasan (2003) in greater generality. Zorzi and Rao (2004) and Tralli (2005) both studied the energy consumption for combined multi-hop and error correcting codes. Their analyses include transmission processing costs and show how the total energy consumption varies with end-to-end transmission distance.

In this section we further emphasise the impact of processing energy by focussing on the transmission-to-processing ratio ρ in (2.7). Thereby we remove the impact all system parameters that affect the transmission distance, but not the transmission-to-processing ratio, and gain better insights into the fundamental trade-off, see Example 2.1. We also extend previous analyses in various respects and include

"Inverse diversity". An effect in multi-hop communication which is sometimes overlooked is what we in lack of a better word will term "inverse diversity": several consecutive transmissions increases the end-to-end probability of error, see Definition 6.1 below. We include this effect in our analysis and make use of the interpretation of the Nakagami fading figure m as the channel's inherent diversity order as given by Definition 3.3 on page 75.

Diversity. In Chapter 5 we found that polarisation receiver diversity offers an possibly processing-cheap way to reduce channel variations. Considering the inverse diversity effect of multi-hopping, the use of diversity reception seems particularly attractive here.

Error correcting codes. As all other techniques that save transmission energy, the multi-hop alternative must overcome the increased processing costs before becoming attractive. Now, error correcting codes can be become attractive at smaller transmission-to-processing ratios than multi-hopping and we therefore include adaptive BCH codes (Section 4.3, page 114) in our analysis.

Amplifier efficiency degradation. The attainable energy gains are highly dependent on the efficiency characteristics of the power amplifier, and the multi-hop gain from shorter transmissions, indicated by (6.1), is diminished by amplifier back-off degradation. We here include this effect through the model by Mikami et al. (2007), see (2.14) on page 23.

Packet aggregation and data fusion. A multi-hop approach offers possibilities to aggregate packets to reduce overhead, and to compress redundant data to reduce packet size. These techniques can add to the transmission-wise benefits from shorter hops, and we study their impact on the overall energy efficiency.

Uncertain relay node positions. In the end of this chapter we include the important impact of the position of the relay node; it will most likely be situated in a non-optimal position with respect to multi-hop transmission energy reduction.

Remark 6.1 *In the previous chapters we have mainly used the transmission-to-total-processing ratio ρ' of (2.8), but here we will also use the transmission-to-processing ratio ρ of (2.7). This is because we sometimes need to distinguish between the processing energy of the destination node and the processing energy of the relaying nodes.*

Unless otherwise stated, to simplify comparisons across chapters the numerical results are presented for a receiver-to-transmitter processing ratio $\alpha = 1$ so that the transmission-to-processing ratios fulfill $\rho = 2\rho'$, see Assumption 2.2 on page 22. This has to be kept in mind when interpreting the results. However, all the equations are kept general.

Assumptions.   To simplify the analysis we initially assume the following:

1. Equidistant multi-hopping, that is N hops of equal length d/N in place of a single hop of length d. We study irregular distances in Section 6.5.

2. Homogeneous transmission environment in the sense that the same propagation loss exponent κ and Nakagami fading figure m can be used for all hop lengths.

3. Non-shadowed transmissions. Shadowing effects are treated in Chapter 8.

Remark 6.2 *Equidistant hopping is seldom an alternative in practice: nodes will rarely be placed carefully at regular distances, and even if they were it need not mean that a relay node is present at the midpoint between source and destination. Equidistant hopping is the most favourable scenario with respect to multi-hop transmission energy.*

Criterion. It is not self-evident what the "right" definition of energy efficiency is when considering a sensor network, and in the multi-hop scenario it is sometimes relevant to study the energy consumption of the most heavily loaded node in addition to the total energy consumption. The reason is that some nodes can suffer from a much larger communication load than others due to their position close to the central sink (or in any area that for some reason is a communication "hot spot"). So even if a scheme minimises the total energy per information bit it could be detrimental to the network's operation if it quickly drains the batteries in certain areas, and thereby causes blind spots or network partitioning. The issue of load imbalance in the multi-hop context has been studied by for instance Perillo et al. (2004) and Olariu and Stojmenovic (2006), and we also included that aspect in our previous work (Björnemo et al., 2006). However, we defer this discussion to Chapter 8 where we delve deeper into the definition of a network wide energy resource metric. Presently we focus on the saving in total energy consumption,

$$w \equiv \frac{\mathcal{E}_{\text{tot},1\text{H}} - \mathcal{E}_{\text{tot},\text{NH}}}{\mathcal{E}_{\text{tot},1\text{H}}}, \tag{6.2}$$

where 1H denotes single-hop and NH denotes N consecutive hops.

6.1   The transmission gain of shorter hops

Before we concentrate on the performance criterion in (6.2), we need to quantify the radiated-energy gain G_{NH} of the N-hop approach. Consider therefore the division of one hop (1H) into N equidistant hops (NH). Due to what we – in lack of a better word – have termed "inverse diversity", the per-hop target error probability must be smaller for the multi-hop scheme, $P_{\text{error},\text{NH}} < P_{\text{error},1\text{H}}$. Here P_{error} typically denotes bit errors or outage.

Lemma 6.1 *Consider N successive transmissions – N hops – each with an assigned error probability P independent of the other transmissions. The end-to-end error probability is then*

$$P(N) = NP + O((NP)^2). \tag{6.3}$$

   Proof: See Appendix 6.A.   ■

As a first order approximation, valid for $NP \ll 1$, we should consequently aim for a per-hop error probability of $P_{\text{error},\text{NH}} = P_{\text{error},1\text{H}}/N$ to attain an overall error probability $B_{1\text{H}}$ over N hops.

Definition 6.1 Inverse diversity *is the N fold increase in end-to-end error probability from P_{error} to NP_{error}, given in Lemma 6.1, caused by N consecutive transmissions of equal error probability.*

From an energy point of view, the increased per-hop requirement will have quite different impact under different channel conditions. We quantify this by the use of the Nakagami-m fading model; this model's m parameter serves the presently useful purpose of quantifying the diversity order of the channel.

Lemma 6.2 *Consider N equidistant hops, each of length d/N, replacing a single hop of length d, and assume that the first order approximation*

$$P_{\text{error}}(N) = NP_{\text{error}}$$

to (6.3) holds to with an error $|\varepsilon| < \delta$, where $\delta > 0$ is a small constant. To ensure equal end-to-end error probabilities for single-hop and multi-hop in a Nakagami-m fading environment, the average hop-wise received signal-to-noise ratio per bit must fulfill

$$\overline{\gamma}_{\text{NH}} \geq N^{\frac{1}{m}}\overline{\gamma}_{\text{1H}}, \tag{6.4}$$

where $\overline{\gamma}_{\text{1H}}$ and and $\overline{\gamma}_{\text{NH}}$ are the average received signal-to-noise ratios per bit for the single hop and each multi-hop respectively.

Proof: See Appendix 6.B. ∎

So, while the power-law propagation loss model in (6.1) assures that multiple hops can reduce the *radiated* energy per bit, according to Lemma 6.2 the *received* energy per bit must be larger in the multi-hop case; the power-law gain is counteracted by the inverse diversity effect.

Theorem 6.1 *Let $\overline{\mathcal{E}}_{\text{rad,NH}}$ be the radiated energy per bit and hop for an equidistant N-hop scenario. Consider the propagation loss model in (6.1) and a communication error performance given by (2.26) for a Nakagami-m channel. Assume that the N-hop error probability is given by*

$$P_{\text{error}}(N) = NP_{\text{error}}(1)$$

according to Lemma 6.1. The reduction in radiated energy per bit and hop achieved by using N equidistant short hops in place of a single long hop is then

$$G_{\text{NH}} \equiv \frac{\overline{\mathcal{E}}_{\text{rad,1H}}}{\overline{\mathcal{E}}_{\text{rad,NH}}} = N^{\kappa - \frac{1}{m}}. \tag{6.5}$$

   Proof: See Appendix 6.C.   ∎

Theorem 6.1 show the combined effect of shorter hops, reducing the per-hop energy by a factor N^κ, and inverse diversity, increasing the per-hop energy by a factor $N^{\frac{1}{m}}$. In effect, the impact of inverse diversity is a reduction in propagation loss exponent from κ to $\kappa - 1/m$. For Rayleigh fading, $m = 1$, this effect is significant.

   Yet another effect that counteracts the power-law energy gains of multi-hop is the degradation in power amplifier efficiency when output power is reduced.

Corollary 6.1 *Under the power amplifier efficiency degradation model (2.14), the use of N equidistant hops leads to the transmission-to-processing ratio*

$$\rho_{\mathrm{NH}} = N^{\frac{1/m-\kappa}{g}}\,\rho_{\mathrm{1H}}, \tag{6.6}$$

where g is the degradation exponent and

$$\rho = \frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{Pt}}}$$

as given by (2.7) on page 21.

   Proof: See Appendix 6.D.   ∎

We have in (6.6) an important relation between transmission-to-processing ratios that neatly includes the number of hops N, the degree of fading m and the effects from inverse diversity, the propagation loss exponent κ and the power amplifier back-off degradation parameter g. In terms of transmission energy, the expression in (6.6) shows quantitatively how much large N and κ speak in favour of multi-hop, and how much large g and m speak against its use. Observe that inclusion of inverse diversity and amplifier degradation has large impact and decrease the "effective propagation loss exponent" from κ to $(\kappa - 1/m)/g$. For instance, a large $\kappa = 5$ can be reduced to $(5-1/1)/2 = 2$ in Rayleigh fading $m = 1$ for $g = 2$. Indeed, because $\kappa > 1/m$, the transmission cost of multiple hops vanishes as $N \to \infty$, but at a rate that can be significantly slower than indicated by reasoning from (6.1) only. The effects of m and g are therefore important to include and we will base our assessment on (6.6).

   Earlier we have studied power control and diversity approaches to the reduction of transmission energy. In particular, we found that dual-branch receiver diversity (D) effectively can double the Nakagami fading figure; $m_{\mathrm{D}} = 2m$. This diversity order gain can be an important ingredient in

multi-hop schemes due to the inverse diversity effect evident in (6.6). With a sufficient degree of accuracy we can simply use $m_{\mathrm{D}} = 2m$ when we consider multi-hopping between nodes equipped with dual-branch diversity receivers (the additional processing costs are subsumed in the transmission-to-processing ratio as they affect multi-hop and single-hop equally).

It is straightforward to extend the analysis to the choice between N and $K < N$ hops by repeating the above derivation.

Corollary 6.2 *Comparing K equidistant hops with N equidistant hops under the power-law model (6.1) and Nakagami-m fading, the transmission-to-processing ratios relate as*

$$\rho_{\mathrm{NH}} = \left(\frac{N}{K}\right)^{\frac{1/m-\kappa}{g}} \rho_{\mathrm{KH}}. \tag{6.7}$$

Proof: Follows immediately from Corollary 6.1. ∎

This generalised expression will be used for instance when considering the optimum number of hops in conjunction with adaptive error correction coding.

6.2 Uncoded multi-hop

Consider a node, here called S, that has data to send to node R, which can be another sensor node or the central sink depending on the situation. We initially assume that all transmissions are uncoded fixed rate transmissions with a transmit power level that is adjusted to achieve a given end-to-end error rate (bit error or outage) over a Nakagami-m channel.[1] The single hop scheme is our reference scheme and we determine how large its transmission-to-processing ratio ρ_{1H} must be to motivate multiple hops.

---

[1]We assume a link margin approach for simplicity, and the adjustment of the power level is hence not aimed at dynamic channel inversion, only a static initial setting.

6.2.1 Energy consumption in a one-dimensional network

The use of N hops saves according to (6.2) an energy fraction

$$
\begin{aligned}
w &= \frac{\mathcal{E}_{\text{tot,1H}} - \mathcal{E}_{\text{tot,NH}}}{\mathcal{E}_{\text{tot,1H}}} \\
&= 1 - \frac{(N-1)(\mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr}} + \mathcal{E}_{\text{T,NH}}) + (\mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr,R}} + \mathcal{E}_{\text{T,NH}})}{\mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr,R}} + \mathcal{E}_{\text{T,1H}}} \\
&= 1 - \frac{(N-1)(1+\alpha) + (1+\alpha_{\text{R}}) + N\rho_{\text{NH}}}{1 + \alpha_{\text{R}} + \rho_{\text{1H}}},
\end{aligned}
\tag{6.8}
$$

where $\mathcal{E}_{\text{Pr,R}}$ and α_{R} refer to the absolute and relative receive processing of the destination node R, while $\mathcal{E}_{\text{Pr,R}}$ and α_{R} refer to the other nodes. Here we use the receiver-to-transmitter processing ratio $\alpha = \mathcal{E}_{\text{Pr}}/\mathcal{E}_{\text{Pt}}$ given in (2.9). Of course, nothing prevents these reception energies from being equal $\alpha = \alpha_{\text{R}}$, or, if the final destination is not energy-limited, the destination's reception energy from being zero, $\alpha_{\text{R}} = 0$.

Now we replace ρ_{NH} in (6.8) by the use of (6.6), and find that

$$
w = \frac{\rho_{\text{1H}}\left(1 - N^{1+\frac{1/m-\kappa}{g}}\right) - (N-1)(1+\alpha)}{1 + \alpha_{\text{R}} + \rho_{\text{1H}}}.
\tag{6.9}
$$

The first term in the nominator is the saved transmission energy while the second term is the additional processing energy.[2] A multi-hop optimistic upper bound on the energy saving can be found by letting $\kappa \to \infty$ so that the multi-hop transmission energy vanishes. We thereby obtain

$$
w_{\text{optimistic}} = \frac{\rho_{\text{1H}} - (N-1)(1+\alpha)}{1 + \alpha_{\text{R}} + \rho_{\text{1H}}}.
\tag{6.10}
$$

The maximum energy saving for a given κ is on the other hand found by letting $\rho_{\text{1H}} \to \infty$ so that transmission costs dominate totally,

$$
w_{\text{max}} = 1 - N^{1+\frac{1/m-\kappa}{g}}.
\tag{6.11}
$$

---

[2]For illustration, let $\rho_{\text{1H}} \to \infty$ to make processing costs negligible, let $m \to \infty$ to ignore inverse diversity, and finally set $g = 1$ to ignore the power amplifier degradation. We then arrive at the common quantitative argument for multi-hop, namely the saving in transmission energy

$$
w = 1 - N^{1-\kappa},
$$

in agreement with direct reasoning from the power-law propagation loss model (6.1).

Uncertainties in m and κ

Under the present power-law propagation loss model, a parameter of prime interest is of course the propagation loss exponent κ which determines the benefits of shorter hops. As seen in (6.9), the fading can have comparable impact on our conclusions: for small fading figures m, the terms $1/m$ and κ in the exponent in (6.9) are of similar magnitude (the fading figure m is not negligible if it is small). We therefore need to consider the uncertainties in both m and κ to judge the likely energy savings from multi-hopping. By the use of a conservative assignment of independent probability distributions,

$$p(\kappa, m | I) = p(\kappa | I) p(m | I), \tag{6.12}$$

we characterise the uncertainty in the energy saving w by plotting its expected value

$$\overline{w} = \iint w(\kappa, m) p(\kappa | I) p(m | I) \mathrm{d}m \mathrm{d}\kappa, \tag{6.13}$$

together with its median w_{med} and the 95th and 5th percentiles w_{95} and w_5 in Figure 6.1. The probability distributions for κ and m are given by (2.79) and (2.85) in Section 2.4.

Remark 6.3 *Intuitively, it is reasonable to assume that m and κ are negatively correlated in some way; it is hard to imagine fading free-space propagation or a non-fading environment where $\kappa \approx 4$. However, the introduction of a negative correlation is not problem-free since there are indeed situations where $\kappa < 2$ while fading effects are still present (Seidel and Rappaport, 1992, p. 210). For this reason we choose the conservative assignment in (6.12) with its larger uncertainty.*

In Figure 6.1 we display the saved fraction w of energy for an equidistant two-hop scheme, $N = 2$, under the assumption that each node radio is equipped with a single antenna. Figure 6.2 shows the corresponding results for sensors equipped with dual branch diversity receivers achieving a diversity order gain $\Delta = 2$, given in Definition 3.4 on page 76. This diversity order gain has the same effect as doubling the Nakagami fading figure, $m \rightarrow 2m$. From Figure 6.1 and Figure 6.2 we observe the following::

- While the absolute minimum transmission-to-total-processing ratio according to (6.10) is $\rho'_{1\mathrm{H}} = 1$, see the uppermost dashed curve in Figure 6.1, the more realistic parameter values reveal that even at $\rho'_{1\mathrm{H}} = 1.7$ we will with a 95 percent probability still *loose* energy by the use of non-diversity multi-hop communication.
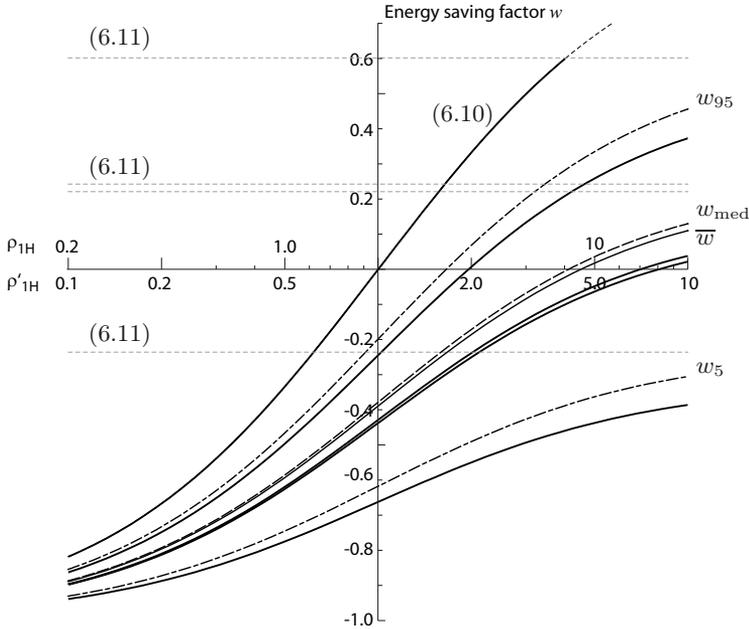
Figure 6.1: Fraction w of saved energy by an equidistant two-hop scheme, $N = 2$, in place of the single-hop scheme. Here under the assumption of non-diversity receivers. All curves correspond to $\alpha = \alpha_{\mathrm{R}} = 1$ and $g = 2$.

- We do not expect multi-hop to save energy until $\rho'_{1\mathrm{H}} > 4.5$ or $\rho'_{1\mathrm{H}} > 3.5$ for non-diversity and dual-branch diversity receivers respectively. Additionally, even at these transmission-to-total-processing ratios there is a fair risk of energy loss; from studying the median curve w_{med} we see that it is a fifty-fifty risk at $\rho'_{1\mathrm{H}} = 4.1$ and $\rho'_{1\mathrm{H}} = 3.3$ respectively for non-diversity and diversity receivers.

- The maximum *expected* energy savings $\overline{w}_{\mathrm{max}}$ in (6.11) are shown as horizontal curves in the figures. The maximum saving for non-diversity reception is 22 percent, while it increases to 29 percent with the use of dual-branch reception.

Remark 6.4 *The results in Figure 6.1 and Figure 6.2 were obtained with $\alpha_{\mathrm{R}} = 1$ and the implicit assumption that the destination node is energy-limited. If the destination node is instead a sink with unlimited power supply, modelled with $\alpha_{\mathrm{R}} = 0$, the energy saving/loss w in (6.9) would be more pronounced; larger energy losses would be induced by multi-hop for small $\rho_{1\mathrm{H}}$, and slightly larger savings would be achieved for large $\rho_{1\mathrm{H}}$. However, neither*
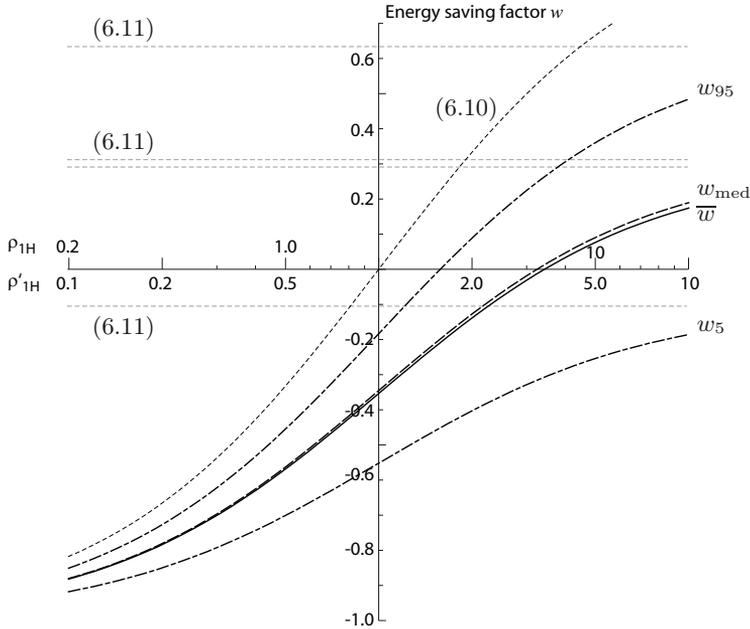
Figure 6.2: Fraction w of saved energy by an equidistant two-hop scheme, $N = 2$, in place of the single-hop scheme. Here under the assumption that dual-branch reception with a diversity order gain of two is used. All curves correspond to $\alpha = \alpha_R = 1$ and $g = 2$.

the maximum energy saving $w_{\max}$ in (6.11), nor the threshold transmission-to-processing ratio $\tilde{\rho}_{1H}$ which we give below in (6.15), would be affected.

Based on the discussion above and by taking full account of Figures 6.1 and 6.2 we reach the following conclusion.

The negative impact of processing energy and inverse diversity rules out uncoded multi-hop as a means to save energy for the currently available node radios since their maximum transmission-to-processing ratios are too small, see Section 2.2.2. Current nodes are forced to resort to multi-hop for range extension when a single hop would be more energy efficient had enough output power been available. Moreover, even at large transmission-to-processing ratios the energy-gains are uncertain because of incomplete knowledge of channel characteristics.

6.2.2   Transmission-to-processing concentration

Whatever the reason for using multi-hop, extended range and/or energy savings, the use of shorter hops will decrease the transmission-to-processing ratio ρ. For large ρ we will therefore see the concentrating effect mentioned in Chapter 2; the optimal approach tends to balance transmission and processing costs. We will investigate this next.

Rearranging (6.9) we obtain

$$\rho_{1\mathrm{H}} = \frac{(N-1)(1+\alpha) + w(1+\alpha_{\mathrm{R}})}{(1-w) - N^{1+\frac{1/m-\kappa}{g}}} \tag{6.14}$$

if N hops are to save a fraction w with respect to a single hop. With $w = 0$ we obtain the the threshold value

$$\tilde{\rho}_{1\mathrm{H}} = (1+\alpha)\frac{(N-1)}{1 - N^{1+\frac{1/m-\kappa}{g}}}. \tag{6.15}$$

Note that $\tilde{\rho}_{1\mathrm{H}}$ is independent of α_{R}. An optimistic lower bound is attained if we let $\kappa \to \infty$,

$$\tilde{\rho}_{1\mathrm{H},\mathrm{optimistic}} = (1+\alpha)(N-1), \tag{6.16}$$

or equivalently,

$$\tilde{\mathcal{E}}_{\mathrm{T},1\mathrm{H},\mathrm{optimistic}} = (\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}})(N-1). \tag{6.17}$$

This limit amounts to neglecting the multi-hop transmission energies and states the trivial fact that the transmission energy of a single hop must at least be as large as the additional processing cost of the $(N-1)$ additional hops. Observe that the optimistic limit in (6.16) increases with N, the number of hops, and $N = 2$ therefore is a case of special interest: if a multi-hop approach can not save energy for $N = 2$ we need not consider larger N.

An extension of (6.15) to the choice between N and $K < N$ hops can be achieved by the use of (6.7) and modification of the right hand side of (6.8). The subsequent derivation setting $w = 0$ yields the threshold

$$\tilde{\rho}_{K\mathrm{H}} = \frac{N-K}{K}\frac{1+\alpha}{1 - \left(\frac{N}{K}\right)^{1+\frac{1/m-\kappa}{g}}}. \tag{6.18}$$

A special case of interest is $N = K + 1$; the threshold when one more hop should be added to save energy. It is in this case instructive to let $K \to \infty$.

Theorem 6.2 *Consider the choice between K and $K+1$ equidistant hops of length d/K and $d/(K+1)$ respectively. When $K \to \infty$, the per-hop threshold transmission-to-total-processing ratio (6.18) becomes*

$$\tilde{\rho}_{\infty H} = (1 + \alpha) \frac{g}{\kappa - 1/m - g}. \tag{6.19}$$

Proof: See Appendix 6.E. ∎

First, we point out the fact that m and g increase the limit (6.19) while κ decreases it. Second, a more interesting fact is that we again see how the energy-optimal approach tends to equalise processing and transmission costs – in the limit $\rho_{KH} = \rho_{(K+1)H}$ – but the degree to which this can be done by multi-hopping depends on κ, m and g. For instance, assuming the use of dual-branch receiver diversity we obtain by inserting the median values $\kappa_{med} = 3.31$ and $m_D = 2m_{med} = 3.76$ together with $g = 2$ the limit $\tilde{\rho}_{\infty H} = 1.92(1 + \alpha)$. In the case of non-diversity receivers, we arrive at $\tilde{\rho}_{\infty H} = 2.57(1 + \alpha)$. Using $\alpha = 1$ we have $\tilde{\rho}_{\infty H} = 3.84$ and have $\tilde{\rho}_{\infty H} = 5.14$ for the two cases respectively. We will study the concentration in (6.19) in Section 6.3 coming next. Observe that the expected threshold is infinite since multi-hop can not save energy when $\kappa \leq g + 1/m$.

The concentration of the transmission-to-processing ratio shown in (6.19) have the following interesting implication.

> If multi-hopping can save energy, then the energy optimal chain of hops will in the equidistant case converge to a fixed transmission-to-processing ratio (6.19). This limit may serve as a design guideline for radios and their maximum transmission-to-processing ratio ρ_{max}: apart from considering the absolute energy consumption a designer should take the limit in (6.19) into account and aim for a $\rho_{max} > \tilde{\rho}_{\infty H}$ so that the full range of energy efficient hop lengths can be utilised.[3]

6.3 Multi-hop with error correcting codes

Error correcting codes start to pay off earlier than multi-hop for uncoded systems – compare Figure 4.7 on page 122 with Figure 6.1 on page 176 –

---

[3]Note that the choice of ρ_{max} is a typical design choice under uncertainty, and the range of propagation losses κ and fading figures m should be taken into account. We will not pursue this issue further here because the hardware design problem concerns many issues we are not considering in this thesis.

and this suggests that error correction can push the initial benefits of multi-hopping towards even larger single-hop distances.

Worth to mention in this context is the work by Zhong et al. (2005) which is focussed on the use of powerful codes to achieve single-hop sensor networks. They argue that error correcting codes with a simple encoding procedure should be used by the energy-limited sensors while the receiving central sink should do the energy-intensive decoding. In this manner one should be able to increase the transmission range to compete with multi-hop. In spite of the fact that they ignore the increased processing time, their line of reasoning fits well into our present analysis. It shows that when the transmission-to-processing ratio is large enough to motivate the extra processing energy incurred by multi-hopping, it is well beyond the minimum requirement for error correcting codes and the range over which single hop is preferable is further increased by error correcting codes.

Assumption 6.1 *We assume a decode and forward approach in which the message is decoded and re-encoded by every node in the multi-hop path.* ∎

Let us begin with a recapitulation from Chapter 4. The use of a rate R_c code with coding gain G_c over a single hop (1H) results in a normalised total energy

$$\breve{\mathcal{E}}_{\text{tot,1Hc}} = \frac{\mathcal{E}_{\text{tot,1Hc}}}{\breve{\mathcal{E}}_{\text{Pt}}} = \frac{1}{R_c}(1 + \alpha_R) + \frac{\rho_{1Hu}}{G_c^{1/g}}, \tag{6.20}$$

where ρ_{1Hu} is the transmission-to-processing ratio of uncoded (u) transmissions, and the subscript (c) denotes the coded scheme. Application of the error correcting code and N equidistant hops lower this ratio to

$$\rho_{NHc} = R_c \left(\frac{N^{1/m-\kappa}}{G_c} \right)^{1/g} \rho_{1Hu} \tag{6.21}$$

since the per-hop processing energy increases with a factor $1/R_c$ while the transmission energy is reduced with a factor $G_c^{1/g}$ due to coding and a factor $1/N^{(1/m-\kappa)/g}$ due to shorter hops. Joint use of multi-hopping and error correction provides more freedom in the trade-off because we can now adapt both the number of hops N and the code rate R_c in a way that makes the most out of the transmission energy gains from shorter hops and coding gain G_c. How these two techniques will combine depend on the propagation loss exponent κ and the fading figure m: we see this later in Figures 6.4 and 6.5.

The total (normalised) per-bit energy becomes

$$\breve{\mathcal{E}}_{\text{tot,NHc}} = \frac{\mathcal{E}_{\text{tot,NHc}}}{\mathcal{E}_{\text{Pt}}}$$

$$= \frac{1}{R_c} \underbrace{[1 + (N-1)(1+\alpha) + \alpha_R]}_{\text{Uncoded } N\text{-hop processing}} + N \underbrace{\left(\frac{N^{1/m-\kappa}}{G_c} \right)^{1/g}}_{N\text{-hop} - \text{coding gain}} \rho_{\text{1Hu}}, \quad (6.22)$$

By choosing the combination of hops N and code rate R_c that yields the minimum $\breve{\mathcal{E}}_{\text{NHc}}$ we achieve the optimum solution.[4] There is no closed form solution to the optimisation, so we perform a simple numerical search over the BCH codes also used in Chapter 4.

Let us start with $N = 2$ and study the joint choice of N and R_c next. In Figure 6.3 we see that the use of error correcting codes increases the required uncoded transmission-to-total-processing ratio ρ'_{1Hu} – compare with Figure 6.1 – and it seems that a transmission-to-total-processing ratio $\rho'_{\text{1Hu}} > 10$ is typically required to yield an energy saving $w > 0$.

However, since the codes themselves lower the transmission-to-processing ratios we need to consider the resulting, post-coding transmission-to-processing ratio (6.21) to get a more comprehensive picture. The concentrating effect on the transmission-to-processing ratio ρ'_{NHc} becomes clearer if we consider more than two hops and choose the jointly optimal code rate R_c and hop number N. In Figure 6.4 and Figure 6.5 we depict, for non-diversity and dual-branch diversity respectively, the optimum number of hops and the corresponding BCH code rate R_c and the resulting transmission-to-total-processing ratio ρ'_{NHc}. The results correspond to the median case, see w_{med} in Figure 6.3.

- There is a strong tendency to concentrate the transmission-to-processing ratio ρ'_{NHc}; it stabilises at the limiting value given by (6.19). Initially, however, there are significant fluctuations around this value.

- The code rate also stabilises at a certain value.[5]

- The equilibrium values of ρ'_{NHc} and R_c depend on the radio conditions. In situations more favourable to multi-hop, that is when the

---

[4]Observe that such a free choice of N requires a very dense network or the possibility to deploy sensors in a way that achieves the desired N.

[5]This might be the effect of the discrete set of codes used, and we are not sure this effect would remain in the limit of continuous code rates. If so, the coincidence between ρ'_{NHc} and the limit in (6.19) might disappear since the limit implicitly assumes a fixed code rate. Additionally, there might be properties particular to BCH codes that manifest themselves in the results.
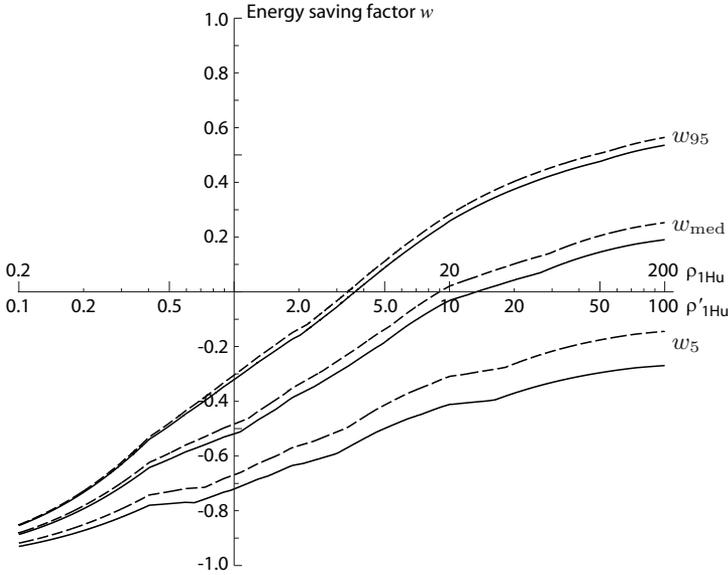
Figure 6.3: Energy saving through two hops versus the uncoded single-hop transmission-to-processing ratios ρ'_{1Hu} and ρ_{1Hu}. Both schemes use the energy-wise best BCH code. The dashed curves correspond to dual-branch diversity receivers, while the solid curves correspond to non-diversity receivers. Common to all results are $\alpha = \alpha_R = 1$, $g = 2$ and a coding gain $G_c = G_{c,max}/2$, where the upper bound $G_{c,max}$ from (4.44) on page 117 has been used.

propagation loss exponent κ is large and the fading figure m is small, the code rates and the number of hops will increase with the result that the equilibrium ρ'_{NHc} will decrease (and vice versa for less multi-hop favourable conditions under which coding will be preferred).[6]

- The existing radios which we have included in Figures 6.4 and 6.5 are all forced to use multi-hop earlier than energy-efficiency prescribes. Not even the CC1000's $\rho'_{max} = 1.6$ for 434 MHz is enough because the optimum scheme requires $\rho' > 3$ in both the cases considered.

We conclude the following from the presented results.

Error correcting codes increase the *distance* over which single-hop is the preferable choice. However, the use of error correct-ing codes decrease the *transmission-to-processing ratio* at which

---

[6]Contrary to the proposition of Zhong et al. (2005) to use very low codes rates R_c, we find that code rates typically stabilise at relatively large values like $R_c = 3/4$.
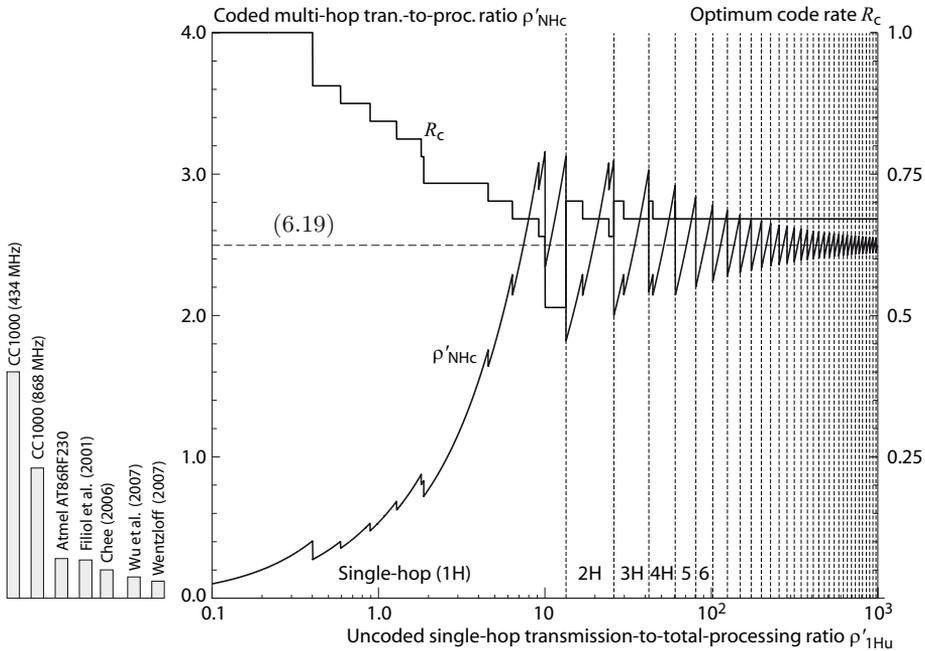
Figure 6.4: Transmission-to-total-processing ratio for multi-hopping with adaptive BCH coding; $\kappa - 1/m = 2.80$ corresponding to the median of (6.12) and non-diversity receivers. The ratio ρ'_{NHc} is kept small thanks to the reduction in transmission energy achieved by multi-hopping and error correction in conjunction. Vertical dotted lines denotes a change in the optimum number of hops.

multi-hop becomes the better than single-hop. Therefore, coding render more radio designs suitable for multi-hop but leads to fewer applications in need for it. The stabilising code rate R_c and coded transmission-to-processing ratio ρ_{NHc}, together with the initial fluctuations around their limiting values, provide valuable design guidelines. Radios should have a maximum transmission-to-total processing ratio ρ'_{max} that is larger than the stabilising value (6.19), about 50 percent judging from Figures 6.4 and 6.5. The minimum code rate must be small enough, around 2/3 judging from the figures, to facilitate a good trade-off between transmission and processing costs.
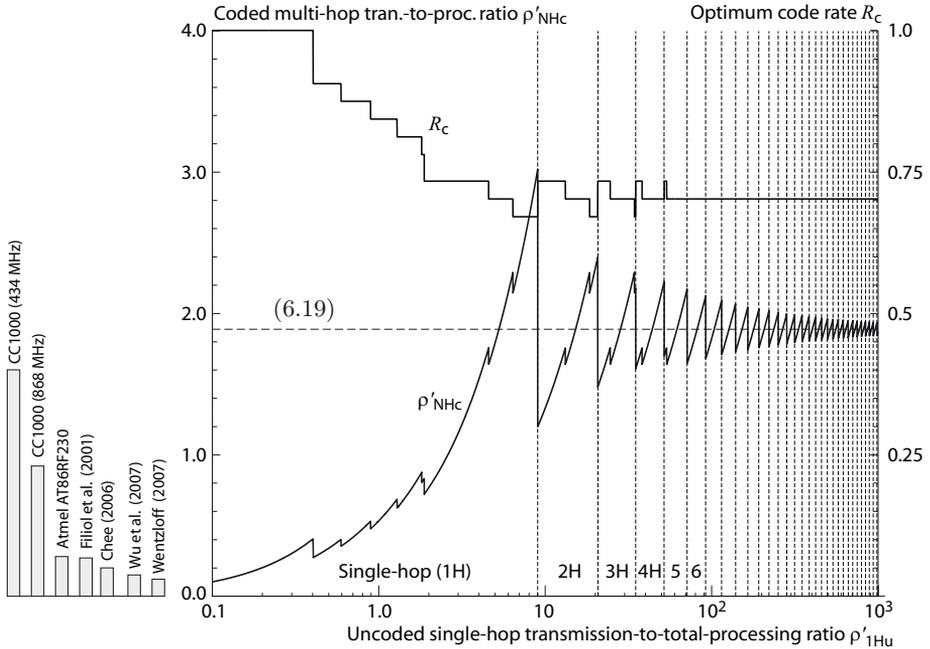
Figure 6.5: Transmission-to-total-processing ratio for multi-hopping with adaptive BCH coding; $\kappa - 1/m = 3.06$ corresponding to the median of (6.12) and dual-branch diversity receivers. The ratio ρ'_{NHc} is kept small thanks to the reduction in transmission energy achieved by multi-hopping and error correction in conjunction. Vertical dotted lines denotes a change in the optimum number of hops.

6.4 Packet aggregation and data fusion

In addition to the reduced transmission costs achieved by multi-hopping, a virtue of the technique that is often put forward is the possibility to perform packet aggregation and data fusion as the packets are being forwarded toward the final destination. Aggregation does not affect the information content of the packets; one achieves a reduction in packet overhead by combining many short packets into a long packet with the same content. Fusion does not affect the overhead but instead exploits data redundancy to reduce the packet size. One assumes that all forwarding nodes have data, partially redundant, to send along with the received data. Here we perform a brief study of data aggregation and fusion from an energy point of view. We limit the study to a one-dimensional network. No error correcting codes are included.

A complication is the combination of aggregation and fusion with inverse diversity effects; different bits in an aggregated packet should (perhaps) be

sent with different error protection as they experience a different number of hops and thereby different error probabilities. We circumvent this problematic issue by assuming that $m \gg 1$, that is to say that the impact of fading is negligible and inverse diversity can be ignored.

6.4.1 Overhead reduction through packet aggregation

Assume that the packet overhead is l bits which add to the L information bearing (sensor value) bits. If each node in the multi-hop path aggregates its own L bits with the incoming packet, the overhead-to-information ratio is reduced since l overhead bits now accommodate $2L$ information bits. Let

$$r_{\mathrm{o}} \equiv \frac{l}{L} \tag{6.23}$$

be the fraction of required overhead when sending an L-bit packet. We assume for simplicity that no extra tagging is needed when aggregating data, and that the aggregation processing consumes a negligible amount of energy.

Theorem 6.3 *Let N sensor nodes send L bits each to the final destination, node R. Number the nodes, from the farthest to the closest, by $n = 1, 2, \ldots, N$. The threshold transmission-to-processing ratio $\tilde{\rho}_{1\mathrm{H}}$, at which the energy consumption of the single-hop (1H) and the packet aggregating, equidistant multi-hop (NH) approaches are equal, is then for a non-fading environment given by*

$$\tilde{\rho}_{1\mathrm{H}} = N^{1+\kappa/g} \frac{\frac{N-1}{2} + \left(\frac{N-1}{2} + \frac{N-1}{N} r_{\mathrm{o}}\right) \alpha - \frac{N-1}{N} r_{\mathrm{o}} \alpha_{\mathrm{R}}}{\sum_{n=1}^{N} (1 + r_{\mathrm{o}})(N - n + 1)^{\kappa/g} - (r_{\mathrm{o}} + n)}, \tag{6.24}$$

where r_{o} is the fraction of packet overhead in (6.23).

    Proof: See Appendix 6.F. ∎

We illustrate the behaviour of the threshold for $N = 2$ in (6.24) in Figure 6.6. There, we show the threshold ratio $\tilde{\rho}_{1\mathrm{H}}$ corresponding to the median propagation loss exponent $\kappa_{\mathrm{med}} = 3.31$, and the percentiles $\kappa_{95} = 2.00$ and $\kappa_5 = 5.12$, of our prior for κ in (2.79). Maybe not unexpected, the impact of packet overhead becomes significant when it is comparable in size with the information content. Large overheads tend to lower the threshold above which multi-hop outperforms single-hop, but the effect is mostly weak for $r_0 < 0.25$. Above $r_{\mathrm{o}} = 0.25$ the threshold ratio decreases significantly and
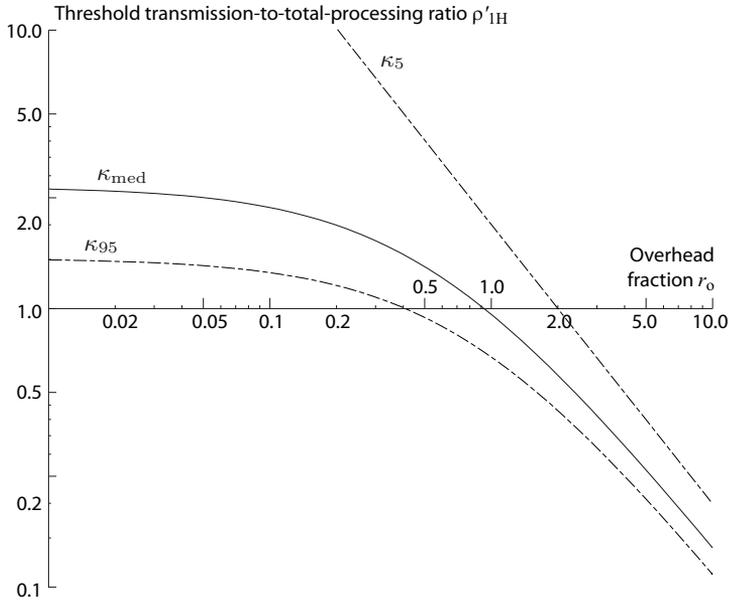
Figure 6.6: *The impact of overhead reduction through data aggregation in multi-hop communications. The overhead fraction r_o from (6.23) is given on the horizontal axis, and the corresponding threshold ratio $\tilde{\rho}'_{1H}$ from (6.24) is given on the vertical axis. Results are calculated for two hops $N = 2$, processing costs $\alpha = \alpha_R = 1$, and power amplifier efficiency degradation $g = 2$.*

above $r_o = 1$ the overhead reduction can even be the major motivating factor for multi-hop communication. To us, an overhead fraction $r_o > 1$ appears huge, but when single sensor readings are reported the overhead can indeed be quite large.[7]

6.4.2   Data fusion in multi-hop transmissions

If the data collected by the N nodes can be compressed without information loss, or without violation of a specified distortion criterion, there is a possibility to perform data fusion along the multi-hop route. Especially if the spatial phenomenon monitored by the network is spatially over-sampled there are compression gains to be attained. The research field of distributed

---

[7]In the recent initiative for using the Internet Protocol in sensor networks, the IPSO Alliance, the packet header is reduced to 6 bytes, see Dunkels and Vasseur (2008). This is however a significant overhead if one or a few bytes of data is going to be transmitted and then $r_o > 1$.

and incremental source coding is large and our aim here is not to fully account for the results achieved in that field. Rather, we try to capture the first-order effects of data fusion on multi-hop energy efficiency and for this purpose we will use the following simple model. We assume that the information in the NL data bits collected by N nodes can be compressed to $r_{\mathrm{f}}NL$ bits, where $1/N \leq r_{\mathrm{f}} \leq 1$. One can view $1/r_{\mathrm{f}}$ as the spatial over-sampling achieved by the present sensor node deployment.[8] Note that $r_{\mathrm{f}} = 1$ amounts to incompressible data while $r_{\mathrm{f}} = 1/N$ corresponds to the computation of a single sufficient statistic (such as the average of all measurements). We further assume that the compression can be carried out incrementally in each hop so that the first node sends L bits and the following node add bits in a linear manner so that node N sends $r_{\mathrm{f}}NL$ bits to the final destination.[9] We summarise this multi-hop data fusion model by letting node n transmit $L(n)$ bits according to

$$L(n) = L\left(1 + (n-1)\frac{r_{\mathrm{f}}N - 1}{N - 1}\right). \tag{6.25}$$

No compression is assumed possible in the single-hop case. We make two more simplifying assumptions before we go on. First, the processing energy of the compression is neglected. Second, we assume that the packet overhead $l = r_{\mathrm{o}}L$ is small enough to be neglected, $r_{\mathrm{o}} \ll 1$. Then we obtain the following result.

Theorem 6.4 *The threshold value $\tilde{r}_{\mathrm{f}}$ below which the multi-hop data fusion alternative is more energy-efficient than the single-hop scheme is given by*

$$\tilde{r}_{\mathrm{f}} = \frac{1 + 2\alpha_{\mathrm{R}} - \alpha + \rho_{\mathrm{1H}}\left(\frac{2}{N}\sum_{n=1}^{N}\left(\frac{N-n+1}{N}\right)^{\kappa/g} - N^{-\kappa/g}\right)}{N(1+\alpha) + 2(\alpha_{\mathrm{R}} - \alpha) + \rho_{\mathrm{1H}}N^{1-\kappa/g}}. \tag{6.26}$$

Proof: See Appendix 6.G. ■

We show $\tilde{r}_{\mathrm{f}}$ graphically in Figure 6.7 for $N = 2$; thereby $1/2 \leq r_{\mathrm{f}} \leq 1$. First, we note that if we have the best possible situation for data fusion, that is $r_{\mathrm{f}} = 1/N$, and $\alpha_{\mathrm{R}} = \alpha$, we would always prefer the multi-hop approach. This is because we can replace the N single-hop transmissions with N shorter transmissions without increased processing; the packet length does

---

[8]The number of signal dimensions, or target positions, reliably identified per sensor has been studied by Aeron et al. (2007) and Rachlin et al. (2005) recpectively. The interested reader is referred to these publications and references therein.

[9]Other incremental schemes are possible.
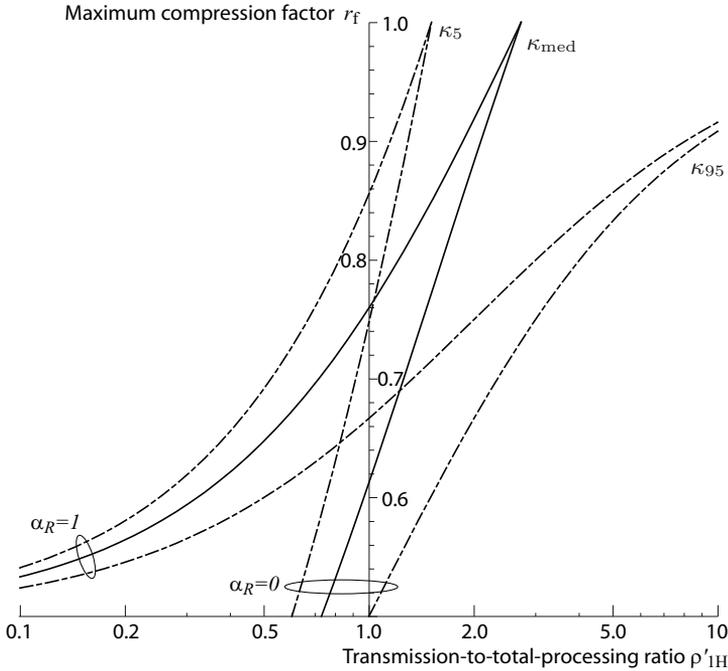
Figure 6.7: *The maximum allowable compression factor r_{f} for different transmission-to-processing ratios $\rho_{1\mathrm{H}}$. Below the curves the sensor network will benefit energy-wise from a combined multi-hop and data-fusion scheme. Here, $N = 2$, $\alpha = 1$ and $g = 2$.*

not increase and the fusion comes at negligible energy cost. On the other hand, if the data is reported to a central sink operating without an energy constraint, we can set $\alpha_{\mathrm{R}} = 0$. We then observe that multi-hopping is not always preferable even in the most optimistic compression case. The reason is that the receive processing costs is zero for single-hop while multi-hop adds receive processing costs at each hop, and these receive costs have to be outweighed by the shorter hops and the compression. At transmission-to-total-processing ratios $\rho'_{1\mathrm{H}} > 1$ the odds for this seem fairly good. In any case, if significant "spatial compression" is possible with a multi-hop approach, but not with the single-hop approach, it can be a strong reason to consider multi-hopping. It adds to the short-hop transmission benefits.

Our compression model in (6.25) is simplistic and many real applications of data fusion are conjectured to behave differently. We do however believe that our model captures the overall effect of data fusion and differences from the linear assumption will not have decisive impact as long as the overall compression factor r_{f} is the same. But, deviations will appear and our results

should be interpreted with this in mind.

From our first-order analysis of packet aggregation and data fusion we conclude the following.

> Large packet overheads and/or substantial data redundancy can offset the previous results that multi-hop is rarely efficient, making it attractive at least in equidistant cases. However, in addition to significant overhead reduction and data fusion, the destination node must be energy limited. In data gathering networks this may not be the case.

6.5 Multi-hop in irregularly deployed networks

Up to this point we have only considered nodes perfectly positioned on a grid. We have assumed that there are always relay-nodes present in the most favourable position[10], regardless of transmission distance and network density. This assumption is of course not very realistic. Therefore, we now include irregular node placement and study how the uncertainty in node position affects design choices. Additionally, we consider the uncertainty regarding the propagation loss exponent; what can be said about multi-hop energy efficiency if the propagation loss exponent κ is imprecisely known?

For convenience, let us consider the most interesting comparison, one hop versus two hops. To simplify the presentation, we focus on the impact of node position by assuming that no shadowing or small-scale fading occurs, and the reader can bear in mind that multi-hop is thus favoured by the absence of inverse-diversity fading effects.

6.5.1 The normalised analysis

To quantify the impact of the relay node's position on the transmission energy we consider the situation depicted in Figure 6.8. The node on the far right makes use of a node located in between itself and the destination node on the far left. We normalise the distance between source and destination to unity, and let the midpoint – which is also the best relay node position – be our Cartesian origin $(0,0)$ as marked by the star in Figure 6.8. The use of a relay node S_d situated on the straight line between source and destination, at a distance $\breve{d}$ from the midpoint, $0 \leq \breve{d} < 1/2$, will lead to a transmission

---

[10]Most favourable from a total transmission energy perspective, for instance a node positioned midway between the transmitter and the destination when comparing single-hop with two hops.
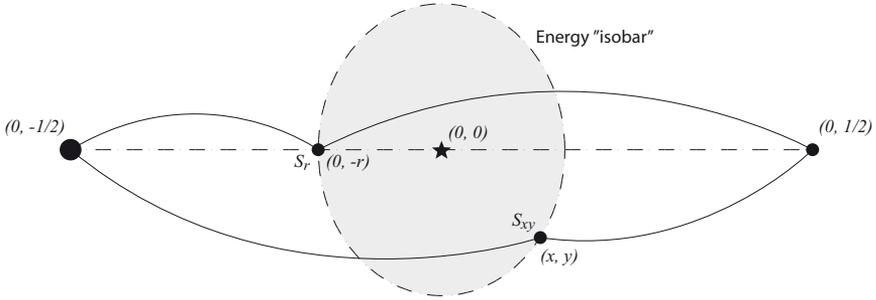
Figure 6.8: Multi-hop, here from right to left, via nodes S_d and S_{xy} are energy-wise equivalent if they are on the same energy "isobar" as defined by (6.28). An elliptic approximation is used to find the area within the isobar.

energy

$$\mathcal{E}_{\text{T,2H}} = \left[\underbrace{\left(\frac{1}{2} + \breve{d} \right)^{\frac{\kappa}{g}}}_{\breve{d}_1} + \underbrace{\left(\frac{1}{2} - \breve{d} \right)^{\frac{\kappa}{g}}}_{\breve{d}_2} \right] \mathcal{E}_{\text{T,1H}}. \tag{6.27}$$

Here we have made use of the power-law propagation loss model (6.1) for hop distances $\breve{d}_1$ and $\breve{d}_2$, together with the amplifier efficiency degradation model (2.14). For all $\kappa/g > 1$ we see that $\breve{d} = 0$ is the best location, and all offsets increase the energy consumption.[11] Consider now a relay node S_{xy} situated at position (x, y), see Figure 6.8. It will provide the same energy-efficiency as relay S_d if

$$\left(\frac{1}{2} + \breve{d} \right)^{\frac{\kappa}{g}} + \left(\frac{1}{2} - \breve{d} \right)^{\frac{\kappa}{g}} = \left(\left(\frac{1}{2} + x \right)^2 + y^2 \right)^{\frac{\kappa}{2g}} + \left(\left(\frac{1}{2} - x \right)^2 + y^2 \right)^{\frac{\kappa}{2g}}. \tag{6.28}$$

The solution defines an energy "isobar" depicted in Figure 6.8, and all relays on this "isobar" are equivalent from an energy point of view. When $\kappa/g = 2$ the solution to (6.28) is a circle of radius $\breve{d}$; $x^2 + y^2 = \breve{d}^2$. Otherwise, as we show in Appendix 6.H by the use of a series expansion of (6.28), the solution can be approximated by an ellipse given by

$$x^2 + \left(\frac{y}{\sqrt{\kappa/g - 1}} \right)^2 = \breve{d}^2. \tag{6.29}$$

---

[11]If $\kappa/g < 1$ the degradation in amplifier efficiency outweighs the transmission energy gain and longer hops are preferable since the higher output power levels correspond to better efficiency. Of course, if this is the case we should always use a single-hop approach (remember that we presently ignore shadowing effects and the possibility to circumvent shadowing objects by multi-hopping).

The ellipse has semi axes of length $\breve{d}$ and $\sqrt{\kappa/g - 1}\breve{d}$ respectively (it thereby includes the circle solution as a special case). We can use (6.29) to find an $\breve{d}$ which is energy-wise equivalent to any position (x, y), and thereby work with a single dimension instead of two.

Finding the best node.

The attainable transmission-energy saving is dependent on $\breve{d}$, that is the relay node's distance to the optimum relay position. Let $\breve{\lambda}$ be the normalised sensor node density, that is the (average) number of nodes within unit area. How close to the optimum $(0, 0)$ can we then hope to find a relay node? With reference to Corollary 2.1 on page 50, we assign a Rayleigh distribution for the semi-minor axis $\breve{d}$ of the smallest ellipse that includes one node, and the elliptic solution in (6.29) we obtain

$$p(\breve{d}|\breve{\lambda}, \kappa, I) = 2\pi\sqrt{\kappa/g - 1}\breve{\lambda}\breve{d}e^{-\pi\sqrt{\kappa/g-1}\breve{\lambda}\breve{d}^2}. \tag{6.30}$$

Large normalised node density and large propagation loss exponent increase our chances to find a good relay node; the Rayleigh distribution becomes narrower and contracts towards $\breve{d} = 0$.

Remark 6.5 *We know from before that the larger the propagation loss exponent κ is, the larger the attained energy saving can be, see (6.1). Additionally, for a given normalised node density $\breve{\lambda}$ we are according to (6.30) more likely to find a relay node in a good position when κ is large (due to the larger elliptic "isobar" in Figure 6.8). However, due to the normalisation of the distance, and hence also the node density, one should remember that $\breve{\lambda}$ actually depends on κ: the larger κ is, the smaller the attainable transmission range becomes, and the actual node density λ thus transforms into a smaller $\breve{\lambda}$. We return to this issue on page 195.*

Any relay node situated at $\breve{d} \geq 1/2$ would be farther away than the destination and would of course not be used since a single hop is then obviously preferable. The probability for finding a relay node closer is

$$\begin{aligned} P(\breve{d} < 1/2|\breve{\lambda}, I) &= \int_0^{1/2} p(\breve{d}|\breve{\lambda}, I)\mathrm{d}\breve{d} \\ &= \int_0^{1/2}\left(\int_1^{\infty} p(\breve{d}|\breve{\lambda}, \kappa, I)p(\kappa|I)\mathrm{d}\kappa\right)\mathrm{d}\breve{d}, \end{aligned} \tag{6.31}$$
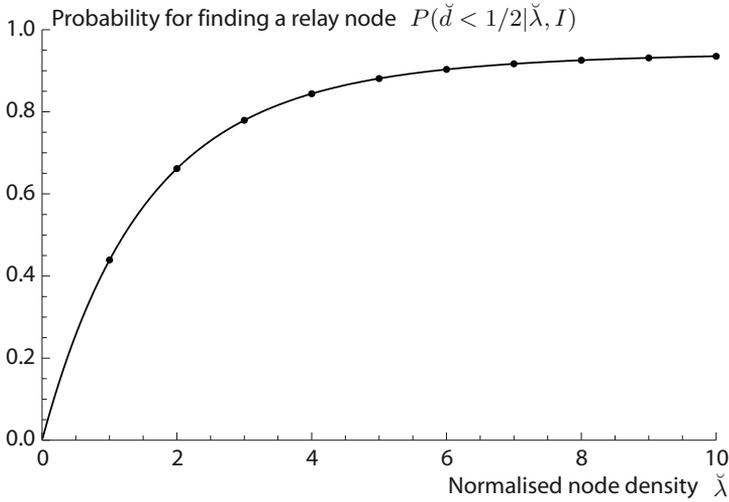
Figure 6.9: The probability for finding a relay node which would offer re-
duced transmission energy. Note that the probability saturates at 0.95
since our prior includes an 0.05 probability that $\kappa \leq 2 = g$; under these
circumstances single-hop is always preferable.

where our present prior $p(\kappa|I)$ is given by (2.79) on page 52. In Figure 6.9 we
show how the probability of finding a relay node depends on the normalised
node density.

Processing energy and total energy saving.

Let us now include processing energies and compare the total energy con-
sumptions of single-hop and two-hop. To save a fraction w of the single-hop
energy we must have

$$
\begin{aligned}
(1 - w)\mathcal{E}_{\text{tot,1Hu}} =& \mathcal{E}_{\text{tot,2Hu}}, \\
(1 - w)(\mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr,R}} + \mathcal{E}_{\text{T,1Hu}}) =& \mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr}} + \mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr,R}} \\
&+ \left[\left(\frac{1}{2} + \check{d}\right)^{\frac{\kappa}{g}} + \left(\frac{1}{2} - \check{d}\right)^{\frac{\kappa}{g}} \right] \mathcal{E}_{\text{T,1Hu}}, \\
(1 - w)(1 + \alpha_{\text{R}} + \rho_{\text{1Hu}}) =& 1 + \alpha + 1 + \alpha_{\text{R}} \\
&+ \left[\left(\frac{1}{2} + \check{d}\right)^{\frac{\kappa}{g}} + \left(\frac{1}{2} - \check{d}\right)^{\frac{\kappa}{g}} \right] \rho_{\text{1Hu}},
\end{aligned}
$$

$$(6.32)$$

where

$$\alpha = \frac{\mathcal{E}_{\mathrm{Pr}}}{\mathcal{E}_{\mathrm{Pt}}},$$
$$\alpha_{\mathrm{R}} = \frac{\mathcal{E}_{\mathrm{Pr,R}}}{\mathcal{E}_{\mathrm{Pt}}} \tag{6.33}$$

denote the normalised reception processing costs of the relay and destination nodes respectively, while

$$\rho_{1\mathrm{Hu}} = \frac{\mathcal{E}_{\mathrm{T,1Hu}}}{\mathcal{E}_{\mathrm{Pt}}} \tag{6.34}$$

is the single-hop uncoded transmission-to-processing ratio. Solving (6.32) for w we obtain

$$w = \frac{\rho_{1\mathrm{Hu}}\left[1 - \left(\left(\frac{1}{2} + \breve{d}\right)^{\frac{\kappa}{g}} + \left(\frac{1}{2} - \breve{d}\right)^{\frac{\kappa}{g}}\right)\right] - (1 + \alpha)}{1 + \alpha_{\mathrm{R}} + \rho_{1\mathrm{Hu}}}. \tag{6.35}$$

By the use of (6.35) and (6.30) we can now find $p(w|(\breve{d} < 1/2), \breve{\lambda}, I)$ – the probability distribution which manifests the uncertainties in κ and $\breve{d}$ – but only through numerical computations since the change of variables from κ and $\breve{d}$ to w is not analytically tractable for all cases. In Figure 6.10 we display the uncertainty in w *given that a relay node is found for $\breve{d} < 1/2$.* This qualification is important when interpreting Figure 6.10; the results could otherwise give a false indication that the sensitivity to the normalised node density is smaller than it actually is. The low-density network requires roughly double the transmission energy of the ideal case, while the denser network comes close to it.

Figure 6.9 and Figure 6.10 together give an idea of the impact of the normalised node density. A compound view is given in Figure 6.11 where we show the probability for $w > 0$, $P(w > 0|\breve{\lambda}, I)$, which is found by numerical integration over the probability distribution for κ and $\breve{d}$. We note that in the best case, $\breve{\lambda} = 50$, a transmission-to-processing ratio $\rho_{1\mathrm{Hu}} > 5.5$ is required to achieve a fifty-fifty chance of saving energy. For sparser networks, the limit increases steadily and the impact of low densities is more clearly visualised than in Figure 6.10.

Based our results and the discussion above we make the following observations.

> Suboptimal relay node placement reduces the multi-hop energy savings significantly. The uncertainty is therefore large when judging the energy efficiency of multi-hop over sparse networks,
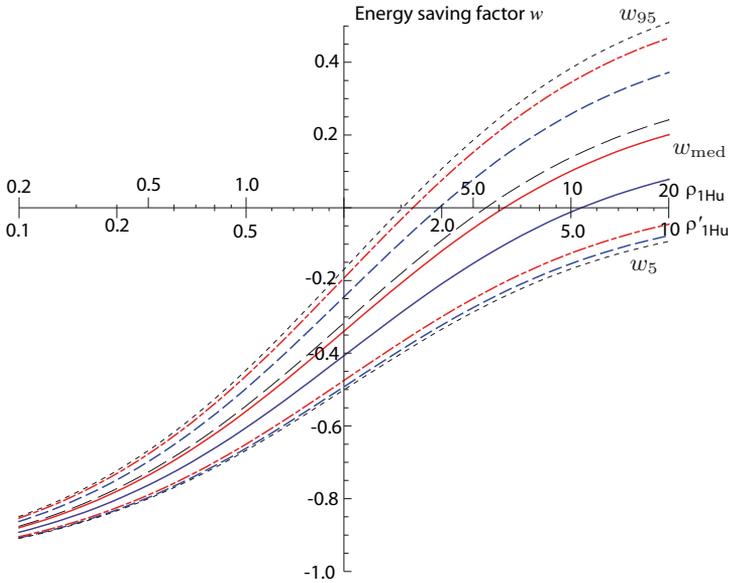
Figure 6.10: Energy saving w given that a relay node is found for $\breve{d} \leq 1/2$. Black lines correspond to the ideal, equidistant, case of (6.9) when $m \to \infty$, while the blue and red lines correspond to $\breve{\lambda} = 1$ and $\breve{\lambda} = 10$ respectively.
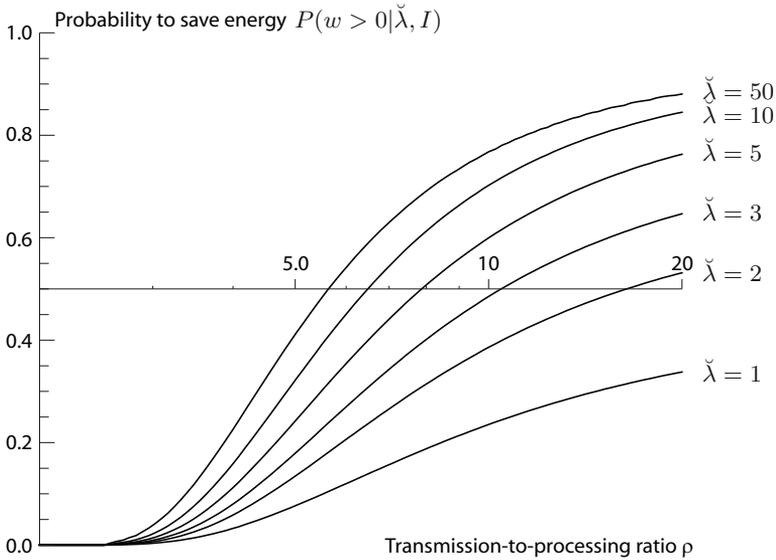
Figure 6.11: The probability of saving energy through a two-hop approach under different normalised node densities.

and this uncertainty probably translates into a larger design margin for the nodes' maximum transmission-to-processing ratio ρ_{max}: to avoid being forced to inefficient multi-hop due to badly positioned relays, the node needs additional output power (relative to the processing power) when relay nodes are not abundant.

6.5.2 Non-normalised analysis

By the preceding analysis we have found that the normalised node density, essentially the average number of nodes present in the unit area between sender and receiver, should be around five to sustain good multi-hop possibilities. But how does this relate to real node densities, given in nodes per square metre? Let us assume constant transmit power and use the propagation loss model (6.1) written in the following form,

$$\mathcal{E}_{\mathrm{rad}} = K_{\mathrm{T}} \left(\frac{d}{d_0} \right)^{\kappa}, \tag{6.36}$$

where d is the inter node distance, d_0 is the model's reference distance, and K_{T} is a transmission constant which depends on target error rate, fading figure, link margin, noise level, receiver noise figure, etc. All the latter contributions are assumed constant in the following calculations. The maximum distance, the maximum range of the radio, is for a free-space environment ($\kappa = 2$),

$$d_{\mathrm{max},2} = d_0 \left(\frac{\mathcal{E}_{\mathrm{rad,max}}}{K_{\mathrm{T}}} \right)^{1/2}. \tag{6.37}$$

where $\mathcal{E}_{\mathrm{rad,max}}$ is the radio's maximum radiated energy per bit. Now, from the Mikami power amplifier model (2.14) we have that

$$\begin{aligned} \mathcal{E}_{\mathrm{rad}} &= \mathcal{E}_{\mathrm{rad,max}} \left(\frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{T,max}}} \right)^{g} \\ &= \mathcal{E}_{\mathrm{rad,max}} \left(\frac{\rho}{\rho_{\mathrm{max}}} \right)^{g}, \end{aligned} \tag{6.38}$$

where g is the amplifier degradation exponent. By the use of (6.36), (6.37) and (6.38) we obtain the normalised density

$$
\begin{aligned}
\breve{\lambda} &= \lambda d^2 \\
&= \lambda d_0^2 \left(\frac{\mathcal{E}_{\mathrm{rad}}}{K_{\mathrm{T}}} \right)^{2/\kappa} \\
&= \lambda d_0^2 \left(\frac{\mathcal{E}_{\mathrm{rad,max}}}{K_{\mathrm{T}}} \right)^{2/\kappa} \left(\frac{\rho}{\rho_{\mathrm{max}}} \right)^{2g/\kappa} \\
&= \lambda d_0^2 \left(\frac{d_{\mathrm{max},2}}{d_0} \right)^{4/\kappa} \left(\frac{\rho}{\rho_{\mathrm{max}}} \right)^{2g/\kappa} ,
\end{aligned}
\tag{6.39}
$$

where λ is the average number of nodes per square metre. Observe that

- For a given transmission-to-processing ratio ρ, the normalised node density $\breve{\lambda}$ *decreases* with an increasing propagation loss exponent κ due to the shortened transmission distance achieved.

Therefore, while the use of a normalised node density $\breve{\lambda}$ gives a good idea of how many relay nodes that are needed (on the average) in the area between sender and receiver, it conceals an important "diluting" effect: smaller transmission-to-processing ratios correspond to shorter distances and hence fewer possible relay nodes. We illustrate this effect in Figure 6.12 where we show data transmissions corresponding to $\rho = 1$, but under different propagation loss exponents κ. Large κ means shorter transmission and smaller $\breve{\lambda}$.

Radio parameters

One of the merits of the transmission-to-processing metric is that we can carry out a more general analysis than if every radio and transmission parameter has to be specified; we have been able to focus on the transmission-processing trade-off regardless of absolute energy levels and specific propagation losses. But now, when the real distances start to interact with the trade-off model we are forced to specify additional characteristics of the propagation and the radio. From (6.39) we conclude that we need to specify d_0, $d_{\mathrm{max},2}$ and ρ_{max} respectively. We make the following selection.

- We will here use $\rho_{\mathrm{max}} = 2\rho'_{\mathrm{max}} = 20$, assuming that $\alpha = \alpha_{\mathrm{R}} = 1$, to facilitate the single-hop vs. multi-hop assessment over the previously considered range of ρ's. To the best of our knowledge there are no sensor node radios available with this large transmission-to-processing ratios – our choice is for illustrative rather than realistic purposes.
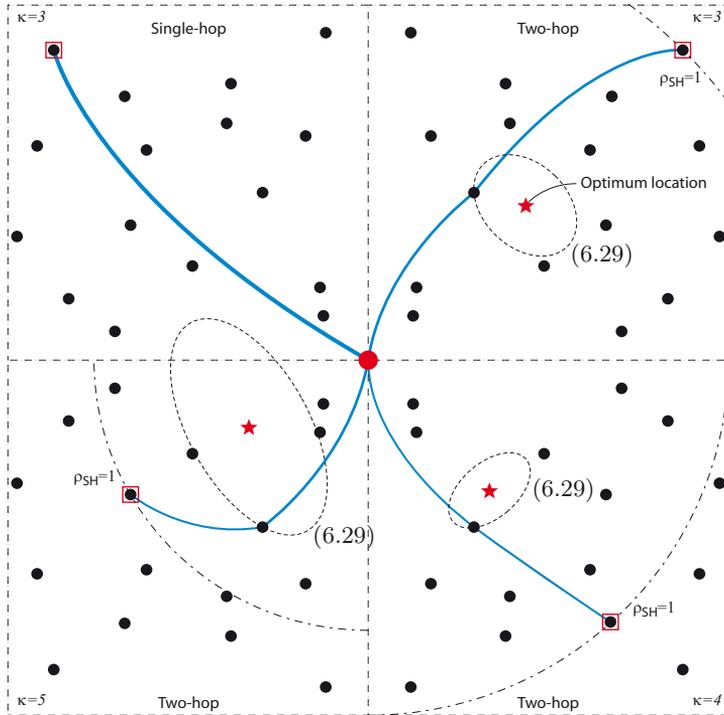
Figure 6.12: The impact of the propagation loss exponent κ on the normalised node density, that is the (average) number of nodes present between source and destination nodes. The stars indicate the optimum node positions and the ellipses show the energy "isobars" of the relay nodes actually present.

- For simplicity we set $d_0 = 1$ m even if this propagation loss parameter can be larger; Goldsmith (2005, page 47) holds that d_0 is typically between one and ten metres in indoor environments.

- Due to the considerable spread in output power and bit rates found for proposed radios, there is a sizeable spread in the free-space transmission range $d_{\mathrm{max},2}$. We have found radiated energies $\mathcal{E}_{\mathrm{rad}}$ per bit ranging from 0.1 nJ (Wu et al., 2007) to 0.9 μJ (Aerocomm, 2008), resulting in transmission ranges spanning from tens of metres up to several kilometres. We choose to use $d_{\mathrm{max},2} = 300$ m and $d_{\mathrm{max},2} = 2500$ m to illustrate the impact of transmission range.

In Figure 6.13 and Figure 6.14 we plot the probability $P(w > 0|\lambda, I)$ for saving energy by the use of two hops. The probability is found by integration over our probability assignment (2.79) for κ and the Rayleigh distribution for
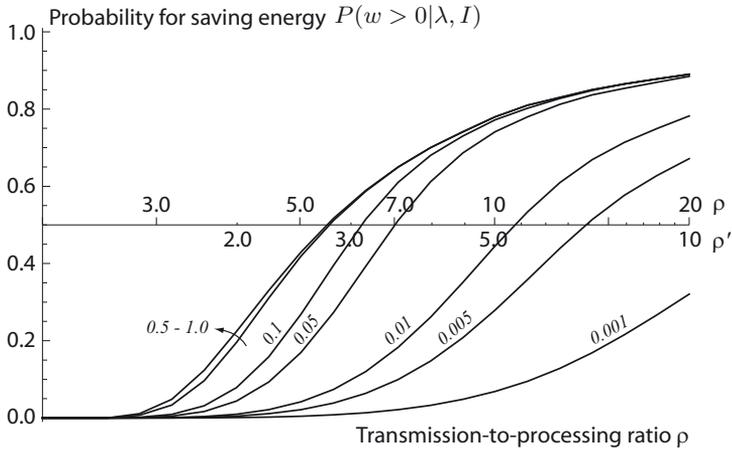
Figure 6.13: *Probability for saving energy by the use of two hops when $d_0 = 1$ m and $d_{\mathrm{max},2} = 300$ m. The node density in nodes per square metre is shown for each curve. Results pertain to $\rho_{\mathrm{max}} = 20$, $\alpha = \alpha_{\mathrm{R}} = 1$ and $g = 2$.*

r as given by Corollary 2.1 on page 50. The most prominent impact of the interdependence between the normalised node density $\breve{\lambda}$ and the transmission-to-processing ratio ρ is that it for low node densities λ takes a long single-hop distance d to attain a sufficient number of possible relay nodes. Hence, the threshold transmission-to-processing ratio can be increased as compared to Figure 6.11. However, once enough relay nodes are present, the number increases with ρ and the climb in probability $P(w > 0|\lambda, I)$ is steepened. The normalised node densities quickly become very large; with $\lambda = 0.01$ m^{-2} and $d_{\mathrm{max},2} = 300$, the normalised node density $\breve{\lambda} = 900$ when $\rho = 20$.

We can now make the following observation.

> Today's ultra-low power radios can cause problems through their limited transmission range, because even when the transmission-to-processing ratio ρ is large enough to motivate multi-hopping, the normalised node density $\breve{\lambda}$ may be too small to facilitate its use energy efficiently. Nodes may have to be more densely deployed than the sensing application requires in order to facilitate communication. In such a case, larger ρ_{max}, than previously suggested is motivated for connectivity reasons rather than energy reasons.
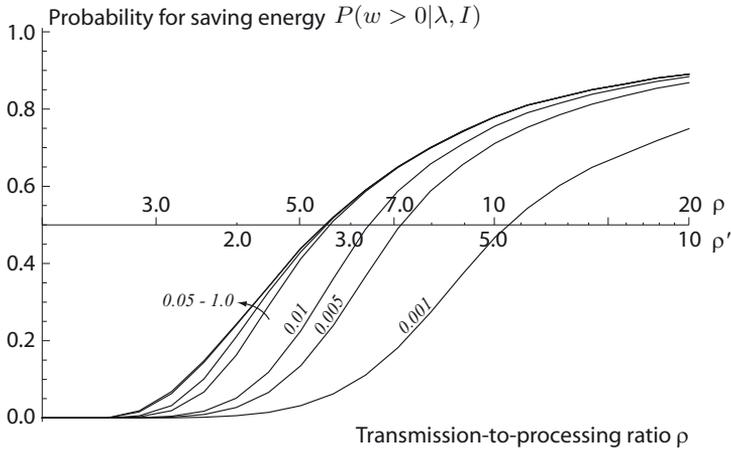
Figure 6.14: Probability for saving energy by the use of two hops when $d_0 = 1$ m and $d_{\mathrm{max},2} = 2500$ m. The node density in nodes per square metre is shown for each curve. Results pertain to $\rho_{\mathrm{max}} = 20$, $\alpha = \alpha_{\mathrm{R}} = 1$ and $g = 2$.

6.6 Concluding remarks

Observe that except for Section 6.5.2, where the actual transmission range comes into play, all our results should hold regardless of the absolute power level; the results pertain to high power radios as well as ultra-low power radios.

Multi-hop presently for range extension. Considering the range of transmission-to-processing ratios that presently available node radios exhibit, it stands clear that the present virtue of multi-hopping is range extension rather than energy-efficiency. This is especially true in the uncoded case, but holds also under an adaptive coding approach. Having said this, we would like to stress that according to our present models, multi-hopping *will* be motivated energy-wise at some point, and single-hop is not the uniformly best alternative for all network sizes. Otherwise, the most notable exceptions to the rule are networks with heavy packet overheads and/or considerable data redundancy where the destination node is energy-limited.

Transmission-to-processing ratio and radio design. As present nodes with their inadequately small transmission-to-processing ratios ρ_{max} can not fully utilise the single-hop benefits we propose as a general guideline that, if the processing costs of the architecture can not be decreased, the out-

put power should be increased to enable the optimal choice between single-hop and multi-hop. Maximum transmission-to-processing ratios approaching $\rho_{\max} = 10$ seems to be required to have some margin against the uncertainty stemming from unknown channel characteristics. Note that a very large maximum output power comes with a penalty in efficiency when the maximum is not utilised. Hence, the margin should not be exaggerated.

Some unconsidered factors. A possibly important factor that we have not considered here is the cost of wake-up, synchronisation and routing, all of which are probably larger for multi-hop than single-hop networks.

Appendix 6.A   Proof of Lemma 6.1

Consider first the end-to-end probability for bit error $P(N)$ after N hops. For each individual hop we assign independent bit error probabilities P. The recurrence equation for the probability of error after $N + 1$ hops is

$$P(N + 1) = P(N)(1 - P) + (1 - P(N))P, \qquad (6.40)$$

where the first term is the probability that the bit is erroneous after N hops and will remain erroneous, and the second term is the probability that the bit is correct but changes due to an incorrect decision at the receiver. The recurrence equation in (6.40) has the solution

$$P(N) = \frac{1}{2}\left[1 - (1 - 2P)^N\right], \qquad (6.41)$$

which can be found by application of the z-transform to (6.40). By expanding the solution $P(N)$ in a power series around $P = 0$ we obtain

$$\begin{aligned} P(N) &= NP - N(N - 1)P^2 + \frac{2}{3}N(N - 1)(N - 2)P^3 + \dots \\ &= NP + O((NP)^2). \end{aligned} \qquad (6.42)$$

Consider next an outage probability over N independent hops. We have then the probability

$$1 - P(N) = (1 - P)^N \qquad (6.43)$$

that no link is in outage, and the corresponding series expansion around $P = 0$ is

$$\begin{aligned} 1 - P(N) &= 1 - NP + \frac{1}{2}N(N - 1)P^2 - N(N - 1)(N - 2)P^3 + \dots \\ &= 1 - NP + O((NP)^2). \end{aligned} \qquad (6.44)$$

Hence, the error probability is $P(N) = NP + O((NP)^2)$. ∎

Appendix 6.B   Proof of Lemma 6.2

Consider a modulation with asymptotically exponential error behaviour in a static Gaussian channels $P_{\text{error,sta}} \propto \exp(-C_1\gamma)$, when $\gamma \gg 1$, for some positive constant C_1. With an average received signal-to-noise ratio $\overline{\gamma}$ over a Nakagami-m fading channel, the error probability P_{error} behaves like

$$P_{\text{error}} \propto \overline{\gamma}^{-m} \qquad (6.45)$$

for sufficiently large average signal-to-noise ratios $\overline{\gamma}$, see (2.108) on page 68. By the use of (6.45) and an overall target error rate $P_{\text{error,1H}} = NP_{\text{error,NH}}$ – it does not matter if we consider outage or bit errors – we find that

$$\begin{aligned} \overline{\gamma}_{\text{1H}}^{-m} &= N\overline{\gamma}_{\text{NH}}^{-m} \\ \overline{\gamma}_{\text{NH}} &= N^{\frac{1}{m}}\overline{\gamma}_{\text{1H}}. \end{aligned} \qquad (6.46)$$

∎

Appendix 6.C   Proof of Theorem 6.1

The average required radiated energy is proportional to the required average signal-to-noise ratio, $\overline{\mathcal{E}}_{\mathrm{rad}} \propto \overline{\gamma}$. Under the assumption that the overall channel conditions are equal for all hop lengths, we then invoke the power law propagation loss model (6.1) for both single-hop and multi-hop. In combination with Lemma 6.2 we obtain

$$\overline{\mathcal{E}}_{\mathrm{rad,NH}} \;=\; K_{\mathrm{T}}\overline{\gamma}_{\mathrm{NH}} \left(\frac{d}{Nd_0}\right)^{\kappa} \tag{6.47}$$

$$=\; K_{\mathrm{T}}N^{\frac{1}{m}-\kappa}\overline{\gamma}_{\mathrm{1H}} \left(\frac{d}{d_0}\right)^{\kappa} \tag{6.48}$$

$$\overline{\mathcal{E}}_{\mathrm{rad,1H}} \;=\; K_{\mathrm{T}}\overline{\gamma}_{\mathrm{1H}} \left(\frac{d}{d_0}\right)^{\kappa}, \tag{6.49}$$

where K_{T} is a transmission constant of no present interest. The resulting gain in radiated energy becomes

$$G_{\mathrm{NH}} \equiv \frac{\overline{\mathcal{E}}_{\mathrm{rad,1H}}}{\overline{\mathcal{E}}_{\mathrm{rad,NH}}} = N^{\kappa-\frac{1}{m}}. \tag{6.50}$$

∎

Appendix 6.D   Proof of Corollary 6.1

By including the amplifier back-off degradation factor g, see (2.14), we can relate the required transmission energies,

$$\overline{\mathcal{E}}_{\mathrm{T,NH}} \;=\; \frac{\mathcal{E}_{\mathrm{T,max}}}{\mathcal{E}_{\mathrm{rad,max}}^{1/g}}\overline{\mathcal{E}}_{\mathrm{rad,NH}}^{1/g}$$

$$=\; \frac{\mathcal{E}_{\mathrm{T,max}}}{\mathcal{E}_{\mathrm{rad,max}}^{1/g}} \left(N^{1/m-\kappa}\overline{\mathcal{E}}_{\mathrm{rad,1H}}\right)^{1/g}$$

$$=\; N^{\frac{1/m-\kappa}{g}}\overline{\mathcal{E}}_{\mathrm{T,1H}}. \tag{6.51}$$

Dividing by the transmitters per-bit processing cost $\mathcal{E}_{\mathrm{Pt}}$ we arrive at

$$\rho_{\mathrm{NH}} = N^{\frac{1/m-\kappa}{g}}\rho_{\mathrm{1H}} \tag{6.52}$$

by the use of the definition of the transmission-to-processing ratio ρ in (2.7). ∎

Appendix 6.E   Proof of Theorem 6.2

Replace N with $K+1$ in (6.18) and substitute $1/K \to x$ and $1+(1/m-\kappa)/g \to y$. We can then write

$$\tilde{\rho}_{KH} = (1+\alpha)\frac{x}{1-(1+x)^{y}}. \tag{6.53}$$

Treating x as a continuous variable, which is a good approximation when K approaches infinity, we make a Taylor series expansion around $x = 0$ and find that

$$\tilde{\rho}_{KH} = (1 + \alpha) \left(-\frac{1}{y} + \frac{y-1}{2y} x + O(x^2) \right). \tag{6.54}$$

Setting $x = 0$ then results in the desired limit (we achieve the same result by treating x as a continuous variable and applying the first l'Hopital rule to the limit of (6.53), which is of the indeterminate form $0/0$).

Appendix 6.F Proof of Theorem 6.3

Let $\rho_{1H}(n)$ be the transmission-to-processing ratio of node n transmitting over a fraction $(N - n + 1)/N$ of the unit distance between the farthest node 1 and the destination node R. If all N sensor nodes communicate directly with the destination node R in a single-hop fashion – each nod transmitting an $L + l = L + r_\text{o}L$ bit packet – the total normalised energy cost per information bit is

$$
\begin{aligned}
\breve{\mathcal{E}}_{1H} &= \frac{\mathcal{E}_{1H}}{\mathcal{E}_{Pt}} \\
&= \frac{1}{NL} \Bigg(N(L + r_\text{o}L) + N(L + r_\text{o}L)\alpha_\text{R} \\
&\quad + (L + r_\text{o}L) \sum_{n=1}^{N} \rho_{1H}(n) \Bigg) \\
&= \underbrace{(1 + r_\text{o})(1 + \alpha_\text{R})}_{\breve{\mathcal{E}}_{Pt} + \breve{\mathcal{E}}_{Pr}} + \underbrace{\frac{1 + r_\text{o}}{N} \sum_{n=1}^{N} \left(\frac{N - n + 1}{N} \right)^{\kappa/g} \rho_{1H}}_{\breve{\mathcal{E}}_\text{T}}, \tag{6.55}
\end{aligned}
$$

where we have used (6.7) to replace

$$\rho_{1H}(n) = \left(\frac{N - n + 1}{N} \right)^{\kappa/g} \rho_{1H}. \tag{6.56}$$

For N hops with aggregation, the packet size grows with the hop number n as $nL + l = nL + r_\text{o}L$. By adding processing energy and transmission energy for the

growing packet we obtain

$$
\begin{aligned}
\breve{\mathcal{E}}_{\mathrm{NH}} &= \frac{1}{NL}\left(\sum_{n=1}^{N}(nL + r_{\mathrm{o}}L) + (NL + r_{\mathrm{o}}L)\alpha_{\mathrm{R}} + \alpha\sum_{n=1}^{N-1}(nL + r_{\mathrm{o}}L)\right. \\
&\quad \left. + \rho_{1\mathrm{H}}N^{-\kappa/g}\sum_{n=1}^{N}(nL + r_{\mathrm{o}}L)\right) \\
&= \underbrace{\frac{1}{N}\sum_{n=1}^{N}(n + r_{\mathrm{o}})}_{\breve{\mathcal{E}}_{\mathrm{Pt}}} + \underbrace{(1 + r_{\mathrm{o}}/N)\alpha_{\mathrm{R}} + \alpha\frac{1}{N}\sum_{n=1}^{N-1}(r_{\mathrm{o}} + n)}_{\breve{\mathcal{E}}_{\mathrm{Pr}}} \\
&\quad + \underbrace{\rho_{1\mathrm{H}}N^{-\kappa/g}\frac{1}{N}\sum_{n=1}^{N}(r_{\mathrm{o}} + n)}_{\breve{\mathcal{E}}_{\mathrm{T}}}.
\end{aligned}
\tag{6.57}
$$

Here we use the receiver-to-transmitter processing ratios $\alpha = \mathcal{E}_{\mathrm{Pr}}/\mathcal{E}_{\mathrm{Pt}}$ and $\alpha = \mathcal{E}_{\mathrm{Pr,R}}/\mathcal{E}_{\mathrm{Pt}}$ for the sensor nodes and the destination node respectively. The threshold transmission-to-processing ratio $\tilde{\rho}_{1\mathrm{H}}$ is found by solving, preferably using Mathematica 6 (Wolfram Research Inc., 2007), $\breve{\mathcal{E}}_{1\mathrm{H}} = \breve{\mathcal{E}}_{\mathrm{NH}}$ for $\rho_{1\mathrm{H}}$ by the use of (6.55) and (6.57);

$$
\tilde{\rho}_{1\mathrm{H}} = N^{1+\kappa/g}\frac{\frac{N-1}{2} + \left(\frac{N-1}{2} + \frac{N-1}{N}r_{\mathrm{o}}\right)\alpha - \frac{N-1}{N}r\alpha_{\mathrm{R}}}{\sum_{n=1}^{N}(1 + r_{\mathrm{o}})(N - n + 1)^{\kappa/g} - (r_{\mathrm{o}} + n)}.
\tag{6.58}
$$

Appendix 6.G   Proof of Theorem 6.4

A network employing single-hop communication directly to the destination node R bears the total cost of N point-to-point transmissions of L bits over normalised hop length $(N - n + 1)/N$. In normalised cost per information bit we have that

$$
\begin{aligned}
\breve{\mathcal{E}}_{1\mathrm{H}} &= \frac{1}{r_{\mathrm{f}}NL}\left(NL(1 + \alpha_{\mathrm{R}}) + L\sum_{n=1}^{N}\rho_{1\mathrm{H}}(n)\right) \\
&= \underbrace{\frac{1 + \alpha_{\mathrm{R}}}{r_{\mathrm{f}}}}_{\breve{\mathcal{E}}_{\mathrm{Pt}}+\breve{\mathcal{E}}_{\mathrm{Pr}}} + \underbrace{\frac{1}{r_{\mathrm{f}}N}\sum_{n=1}^{N}\left(\frac{N - n + 1}{N}\right)^{\kappa/g}\rho_{1\mathrm{H}}}_{\breve{\mathcal{E}}_{\mathrm{T}}}.
\end{aligned}
\tag{6.59}
$$

Here $\rho_{1\mathrm{H}}(n)$ has been replaced according to Corollary 6.2. A multi-hop network obeying the data fusion model in (6.25) expends energy for transmitting $L(n)$ bits per hop over N equidistant hops of normalised length $1/N$. The normalised energy

per information bit is

$$
\begin{aligned}
\breve{\mathcal{E}}_{\mathrm{NH}} &= \frac{1}{r_{\mathrm{f}} NL} \left(\sum_{n=1}^{N} L(n) + \alpha \sum_{n=1}^{N-1} L(n) + \alpha_{\mathrm{R}} r_{\mathrm{f}} NL + \rho_{\mathrm{1H}} N^{-\kappa/g} \sum_{n=1}^{N} L(n) \right) \\
&= \frac{1}{2}\left(N + \frac{1}{r_{\mathrm{f}}} \right)(1+\alpha) + (\alpha_{\mathrm{R}} - \alpha) + \frac{1}{2}\rho_{\mathrm{1H}} N^{-\kappa/g} \left(N + \frac{1}{r_{\mathrm{f}}} \right),
\end{aligned}
\tag{6.60}
$$

where we have made use of the fact that

$$
\sum_{n=1}^{N} L(n) = L \sum_{n=1}^{N} \left(1 + (n-1)\frac{r_{\mathrm{f}} N - 1}{N-1} \right) = \frac{1}{2} LN(r_{\mathrm{f}} N + 1).
\tag{6.61}
$$

Equating the energy consumptions in (6.59) and (6.60) and solving for r_{f} we obtain

$$
\tilde{r}_{\mathrm{f}} = \frac{1 + 2\alpha_{\mathrm{R}} - \alpha + \rho_{\mathrm{1H}} \left(\frac{2}{N} \sum_{n=1}^{N} \left(\frac{N-n+1}{N} \right)^{\kappa/g} - N^{-\kappa/g} \right)}{N(1+\alpha) + 2(\alpha_{\mathrm{R}} - \alpha) + \rho_{\mathrm{1H}} N^{1-\kappa/g}}.
\tag{6.62}
$$

Appendix 6.H An elliptic approximation

First, expand the left hand side of (6.28) in a power series in $\breve{d}$ to obtain

$$
2^{1-\frac{\kappa}{g}} - \frac{2^{2-\frac{\kappa}{g}}(g-\kappa)\kappa\breve{d}^2}{g^2} + O\left(\breve{d}^3 \right).
\tag{6.63}
$$

Second, expand the right hand in x and y to arrive at

$$
2^{1-\frac{\kappa}{g}} + \frac{2^{2-\frac{\kappa}{g}}\kappa}{g}y^2 - \frac{2^{2-\frac{\kappa}{g}}(g-\kappa)\kappa}{g^2}x^2 + O\left((xy)^2 \right).
\tag{6.64}
$$

Keeping only second order terms in x, y and $\breve{d}$ we find the elliptic approximation from equation (6.63) and (6.64),

$$
x^2 + \left(\frac{y}{\sqrt{\kappa/g-1}} \right)^2 = \breve{d}^2.
\tag{6.65}
$$

Chapter 7

Cooperative MIMO

THE STRONG interest in wireless sensor networks has stimulated the development of a range of cooperative techniques that are aimed at efficient use of the network's limited energy resources. In this chapter we study the possible energy-wise benefits from cooperative multiple-input multiple-output (MIMO) communication; clusters of nodes cooperating in the transmission and reception to improve the communication performance. Cui et al. (2004) presented the first thorough investigation of the energy-efficiency of cooperative *including the processing energy*. Their conclusion was that in spite of the processing costs, cooperative MIMO yielded large energy savings above a modest threshold distance. By analysing the transmission-processing trade-off directly through the transmission-to-processing ratio ρ we come to the different conclusion that currently available nodes will not, or only rarely, save energy by the use of cooperative MIMO communication in place of direct node-to-node (SISO) communication. Our analyses favours the cooperative scheme by ignoring issues such as local transmission costs and synchronisation.

We also consider the comparison between multi-hop and cooperative MIMO in the case that single-hop SISO actually is less energy-efficient (for large transmission-to-processing ratios ρ). This comparison is not new, it is rather well studied as a special case of cooperative diversity, or cooperative relaying. However, in contrast to most of the research which is focussed on communication capacity and transmission energy, we concentrate on the first-order effects of processing energy. Unfortunately, we find no straightforward answer to the multi-hop or MIMO design question, mainly due to the uncertainty regarding the channel characteristics.

7.1   Cooperative MIMO-STBC

Let us consider a node S which has data intended for node R, see Figure 7.1.
With respect to the transmitter and receiver clusters in Figure 7.1, let

$$
\begin{aligned}
n_{\mathrm{t}} &\equiv \text{ the number transmit nodes,}\\
n_{\mathrm{r}} &\equiv \text{ the number receive nodes.}
\end{aligned}
\tag{7.1}
$$

A cooperative MIMO transmission from S to R can be divided into three
stages of operation: 1) Sharing the data from the source node S among the
n_{t} nodes participating on the transmit side. 2) Performing the synchronised
MIMO transmission from the n_{t} transmit nodes to the n_{r} receive nodes.
3) Gathering the received symbols at the destination node R. By contrast,
the SISO scheme involves only one point-to-point transmission from node
S to node R. The main energy trade-off is obviously that between pro-
cessing energy and transmission energy, but the cooperative scheme must
also expend some transmission energy in the stages of data sharing and
data gathering. The latter could of course be viewed as processing energy
if one likes, but the label has no bearing on the results. We now model the
two parts, processing and transmission energies, in a straightforward man-
ner under the assumption of fixed-rate transmissions through a flat fading
Nakagami-m channel, with channel state information at the receiver only.
The MIMO transmissions use orthogonal space-time block codes (STBC),
known for low-complexity decoding at the receiver. The SISO transmissions
are not protected by error correcting codes. Unless otherwise stated, all
nodes are identical and each node is supposed to have a single antenna.

7.1.1   Total processing energy per bit

We will here express the processing energies consumed during transmit coop-
eration, the long-range cooperative transmission and the receive cooperation.
In conformity with previous chapters we express the energies normalised by
the single node transmit processing energy per bit, $\mathcal{E}_{\mathrm{Pt}}$.

Transmit cooperation. Starting with the transmit side, we note that
the data from node S can be shared by the use of a single broadcast. This
broadcast is only required when $n_{\mathrm{t}} \geq 2$ and by the use of the unit step
function

$$
u(x) = \begin{cases} 0 &,\quad x < 0 \\ 1 &,\quad x \geq 0 \end{cases}
\tag{7.2}
$$
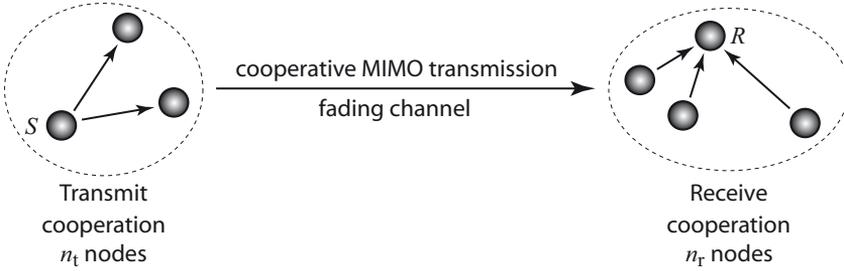
Figure 7.1: *A cooperative MIMO transmission of data from source node S to destination node R. The key idea is that exploitation of cooperative diversity will reduce the required transmission power and thereby save precious energy resources. The principal question is whether this transmission gain is larger than the cooperation cost.*

we can express the normalised processing energy per bit as

$$
\begin{aligned}
\breve{\mathcal{E}}_{\text{RPt}} &= \frac{\mathcal{E}_{\text{RPt}}}{\mathcal{E}_{\text{Pt}}} \\
&= \frac{u(n_{\text{t}} - 2)\mathcal{E}_{\text{Pt}} + (n_{\text{t}} - 1)\mathcal{E}_{\text{Pr}}}{\mathcal{E}_{\text{Pt}}} \\
&= u(n_{\text{t}} - 2) + (n_{\text{t}} - 1)\alpha,
\end{aligned}
\tag{7.3}
$$

where $\mathcal{E}_{\text{RPt}}$ is the total processing energy per bit, see (2.5) on page 20, and $\alpha = \mathcal{E}_{\text{Pr}}/\mathcal{E}_{\text{Pt}}$ is the receiver-to-transmitter processing ratio in (2.9) on page 21.

Long-range cooperative MIMO transmission. We include n_{t} transmitters and n_{r} receivers to arrive at

$$
\begin{aligned}
\breve{\mathcal{E}}_{\text{RP,long}} &= \frac{1}{R_{\text{c}}} \frac{n_{\text{t}}\mathcal{E}_{\text{Pt}} + n_{\text{r}}\mathcal{E}_{\text{Pr}}}{\mathcal{E}_{\text{Pt}}} \\
&= \frac{1}{R_{\text{c}}} \left(n_{\text{t}} + n_{\text{r}}\alpha \right),
\end{aligned}
\tag{7.4}
$$

where R_{c} is the code rate of the orthogonal space-time block code.[1] For SISO transmissions $R_{\text{c}} = 1$ while it depends on n_{t} and n_{r} for MIMO transmissions (Larsson and Stoica, 2003, Sec. 7.4).

---

[1]We will assume that an orthogonal design can be found, and this holds true for the small MIMO systems we consider.

Receive cooperation.   Finally we suppose that the data sharing on the receiver side requires $n_{\mathrm{r}} - 1$ point-to-point transmissions from the assisting nodes to node R, the final destination. Consequently, the associated processing energy per bit is

$$
\begin{aligned}
\breve{\mathcal{E}}_{\mathrm{RPr}} &= \frac{(n_{\mathrm{r}} - 1)(\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}})}{\mathcal{E}_{\mathrm{Pt}}} \\
&= (n_{\mathrm{r}} - 1)(1 + \alpha).
\end{aligned}
\tag{7.5}
$$

Total processing energy.   Summing all processing energies in (7.3), (7.4) and (7.5) we find that the total processing cost for transmit cooperation, long range transmission and receive cooperation is given by

$$
\breve{\mathcal{E}}_{\mathrm{RP}} = \underbrace{\left[u(n_{\mathrm{t}} - 2) + \frac{1}{R_{\mathrm{c}}} n_{\mathrm{t}} + n_{\mathrm{r}} - 1 \right]}_{\text{transmit proc.}} + \underbrace{\left[n_{\mathrm{t}} + \left(1 + \frac{1}{R_{\mathrm{c}}} \right) n_{\mathrm{r}} - 2 \right] \alpha}_{\text{receive proc.}}, \quad (7.6)
$$

where the braces collect the processing energies for all transmissions and receptions respectively (transmission and reception takes place in both the clusters of cooperating nodes). Observe that (7.6) is valid for both SISO and cooperative MIMO, that is for all positive integers n_{t} and n_{r}.

7.1.2   Total transmission energy per bit

Reasonably, the energy consumptions of the SISO and cooperative MIMO schemes should be compared under the requirement of equal end-to-end performance, that is from node S to node R. For convenience we will disregard from the fact that the local transmissions before and after the long-range transmission may be corrupted by errors.[2]

Transmit cooperation.   The transmit energy per bit expended by the source node S in the initial, local, broadcast will be denoted $\mathcal{E}_{\mathrm{Tt}}$. We include it in the model, but we will for simplicity mostly assume that it is negligible and that the corresponding transmission-to-processing ratio, given in (2.7),

$$
\rho_{\mathrm{t}} = \mathcal{E}_{\mathrm{Tt}}/\mathcal{E}_{\mathrm{Pt}}
\tag{7.7}
$$

---

[2]Remember that the inverse diversity effect, given by Definition 6.1 on page 171, was quite important in the multi-hop scheme applied in a fading environment (see Section 6.1, Lemma 6.2 on page 171). In a cooperative scheme parts of the information actually has to undergo multiple hops. We are here however considering short hops for which a very small probability of error can be achieved without excessive transmit power levels.

is zero $\rho_t = 0$. The motivation is that if cooperative MIMO is to be sensible at all, the distances within the transmit cluster should be significantly shorter than the long range transmission. We will thus underestimate the energy cost of the cooperative scheme somewhat, but not severely so.

Long-range transmission. Most important of the transmission energies is the long range transmission energy. Naturally, to save energy by the use of cooperative MIMO the transmit nodes must adjust their transmit power level; we here assume that they use a fixed link margin, or that any long-term adjustments of the transmit power comes at a negligible energy cost. We show in Appendix 7.A that the average bit error rate[3] for coherently detected quadrature amplitude modulated (QAM) transmissions over a Nakagami-m channel is accurately approximated by

$$
\begin{aligned}
\overline{B} = (-1)^{-mn_r n_t} & \frac{2}{b\sqrt{\pi}} \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \frac{\Gamma\left(mn_r n_t + \frac{1}{2}\right)}{\Gamma(mn_r n_t)} \\
& \times \beta\left(-\frac{2\left(2^b - 1\right) mn_t R_c}{3b\overline{\gamma}}, mn_r n_t, \frac{1}{2} - mn_r n_t\right)
\end{aligned}
\tag{7.8}
$$

where $b \geq 2$ is the number of bits per symbol, $\beta(x, p, q)$ is the incomplete beta function (Gradshteyn and Ryzhik, 2000, p. 900), and $\overline{\gamma}$ is the average received signal-to-noise ratio per bit. The result in (7.8) was derived under the assumption of identical and independent probability distributions across the $n_t n_r$ channels of which the receivers have perfect information. It is exact for $b = 2$, and we will henceforth assume that two bits per symbol are used, that is quaternary phase shift keying (QPSK). The radiated-energy gain can be found from the required average signal-to-noise ratios per bit for SISO (S) and cooperative MIMO (cM) respectively,

$$
G_{cM} \equiv \frac{\overline{\gamma}_S}{\overline{\gamma}_{cM}}.
\tag{7.9}
$$

Here, $\overline{\gamma}_S$ and $\overline{\gamma}_{cM}$ are found numerically from the bit error rate (7.8). Invoking the Mikami et al. (2007) amplifier efficiency degradation model (2.14) we can by use of (7.9) relate the transmit energies per bit as

$$
\mathcal{E}_{T,cM} = \frac{\mathcal{E}_{T,S}}{G_c^{1/g}},
\tag{7.10}
$$

---

[3]Sometimes, especially in the cases when the fading is slow, an appropriate performance criterion could be the maximum probability of outage, but we will for brevity focus on a bit error rate criterion. In Björnemo et al. (2007) we used the outage capacity and arrived at similar numerical results as we do here.

or equivalently, using the transmission-to-processing ratio ρ from (2.7), as

$$\rho_{\mathrm{cM}} = \frac{\rho_{\mathrm{S}}}{G_{\mathrm{c}}^{1/g}}. \tag{7.11}$$

We have here assumed that all system and transmission parameters not mentioned, such as noise figures and ambient noise levels, are identical for the two schemes.

Receive cooperation.   The last transmission energy term we need to consider is the sum of the $n_{\mathrm{r}} - 1$ point-to-point transmission energies consumed during the gathering of the received symbols at node R. For each transmission $n = 1, 2, \ldots, n_{\mathrm{r}} - 1$, let

$$\mathcal{E}_{\mathrm{Tr}}(n) \equiv \text{ the transmit energy of transmission } n. \tag{7.12}$$

Summing the $n_{\mathrm{r}} - 1$ transmission energies and normalising with respect to the receiver processing energy $\mathcal{E}_{\mathrm{Pt}}$ we obtain

$$
\begin{aligned}
\breve{\mathcal{E}}_{\mathrm{Tr}} &= \sum_{n=1}^{n_{\mathrm{r}}-1} \frac{\mathcal{E}_{\mathrm{Tr}}(n)}{\mathcal{E}_{\mathrm{Pt}}} \\
&= \sum_{n=1}^{n_{\mathrm{r}}-1} \rho_{\mathrm{r}}(n),
\end{aligned} \tag{7.13}
$$

where $\mathcal{E}_{\mathrm{Tr}}$ is the total transmission energy per bit.

Total transmission energy.   In summary, we obtain the total normalised transmission energy by summing (7.7), (7.11) and (7.13). The result is given by

$$
\begin{aligned}
\breve{\mathcal{E}}_{\mathrm{T}} &= u(n_{\mathrm{t}} - 2)\rho_{\mathrm{t}} + \rho_{\mathrm{cM}} + \sum_{n=1}^{n_{\mathrm{r}}-1} \rho_{\mathrm{r}}(n) \\
&= u(n_{\mathrm{t}} - 2)\rho_{\mathrm{t}} + \frac{\rho_{\mathrm{S}}}{G_{\mathrm{cM}}^{1/g}} + \sum_{n=1}^{n_{\mathrm{r}}-1} \rho_{\mathrm{r}}(n).
\end{aligned} \tag{7.14}
$$

If we use $n_{\mathrm{t}} = 1$ and $n_{\mathrm{r}} = 1$ in (7.14) it reduces to the SISO long-range transmission-to-processing ratio ρ_{S} since G_{cM} then is one.

7.2   Total energy comparison

As usual, we study the energy saving factor w through the relation

$$(1 - w)\mathcal{E}_{\mathrm{tot,S}} = \mathcal{E}_{\mathrm{tot,cM}}, \tag{7.15}$$

where $\mathcal{E}_{\mathrm{tot,S}}$ and $\mathcal{E}_{\mathrm{tot,cM}}$ are the total energy consumptions per bit for SISO (S) and cooperative MIMO (cM) respectively (recall from (2.3) that $\mathcal{E}_{\mathrm{tot}} = \mathcal{E}_{\mathrm{RP}} + \mathcal{E}_{\mathrm{T}}$). By the use of (7.6) and (7.14) and their normalised energies we obtain

$$\begin{aligned}
(1 - w)\left(1 + \alpha + \rho_{\mathrm{S}}\right) = {}& \left(u(n_{\mathrm{t}} - 2) + (n_{\mathrm{t}} - 1)\alpha + u(n_{\mathrm{t}} - 2)\rho_{\mathrm{t}}\right) \\
& + \left(\frac{1}{R_{\mathrm{c}}}(n_{\mathrm{t}} + n_{\mathrm{r}}\alpha) + \frac{\rho_{\mathrm{S}}}{G_{\mathrm{cM}}^{1/g}}\right) \\
& + \left((n_{\mathrm{r}} - 1)(1 + \alpha) + \sum_{n=1}^{n_{\mathrm{r}}-1} \rho_{\mathrm{r}}(n)\right),
\end{aligned} \tag{7.16}$$

where the left hand side is the total normalised energy per bit required for a SISO transmission while the right hand side is its cooperative MIMO counterpart. In the calculations below we will omit the impact of the transmission energies expended before and after the long range transmission; we henceforth set

$$\begin{aligned}
\rho_{\mathrm{t}} &= 0 \\
\rho_{\mathrm{r}}(n) &= 0
\end{aligned} \tag{7.17}$$

for all n. This will favour cooperative MIMO somewhat but for short intra cluster distances between the cooperating nodes the error will be negligible. We proceed by solving (7.16) for w, and we find that

$$\begin{aligned}
w = {}& \left(1 - G_{\mathrm{cM}}^{-1/g}\right) \frac{\rho_{\mathrm{S}}}{1 + \alpha + \rho_{\mathrm{S}}} \\
& - \frac{\frac{1}{R_{\mathrm{c}}}(n_{\mathrm{t}} + n_{\mathrm{r}}\alpha) + (u(n_{\mathrm{t}} - 2) + n_{\mathrm{r}} - 2) + (n_{\mathrm{t}} + n_{\mathrm{r}} - 3)\alpha}{1 + \alpha + \rho_{\mathrm{S}}}.
\end{aligned} \tag{7.18}$$

In (7.18) we observe that the first term on the right hand side is the fraction of energy saved by the cooperative MIMO gain G_{cM}, but the second term gives the penalty in terms of increased processing energy caused by the cooperation. By the use of (7.19) we can establish a fundamental limit regarding the energy-efficiency of cooperative MIMO relative to SISO.

7.2.1   Lower bound on the transmission-to-processing ratio

By solving (7.16) for the transmission-to-processing ratio, and inserting $w = 0$, we obtain the threshold

$$\tilde{\rho}_{\mathrm{S}} = \frac{\frac{1}{R_{\mathrm{c}}}\left(n_{\mathrm{t}} + n_{\mathrm{r}}\alpha\right) + \left(u(n_{\mathrm{t}} - 2) + n_{\mathrm{r}} - 2\right) + \left(n_{\mathrm{t}} + n_{\mathrm{r}} - 3\right)\alpha}{1 - G_{\mathrm{cM}}^{-1/g}} \tag{7.19}$$

First, we note the following regarding (7.19). The threshold transmission-to-processing ratio $\tilde{\rho}_{\mathrm{S}}$ is increasing with the number of cooperating nodes n_{t} and n_{r}. It also increases with decreasing code rate R_{c}. Of course, $\tilde{\rho}_{\mathrm{S}}$ decreases with the gain G_{cM}.

Next, let us be maximally optimistic regarding cooperative MIMO and assume that $G_{\mathrm{cM}} \to \infty$ at code rate $R_{\mathrm{c}} = 1$. The resulting bound on $\tilde{\rho}_{\mathrm{S}}$ is then

$$\tilde{\rho}_{\mathrm{S,optimistic}} = \left(n_{\mathrm{t}} + n_{\mathrm{r}} + u(n_{\mathrm{t}} - 2) - 2\right) + \left(n_{\mathrm{t}} + 2n_{\mathrm{r}} - 3\right)\alpha. \tag{7.20}$$

The above expression simply states that the transmission energy in the SISO approach must be at least as large as the additional processing energy in the cooperative approach. The smallest MIMO configurations are 2×1 MIMO and 1×2 MIMO. For these cases

$$\rho_{\mathrm{S21,optimistic}} = 2 + \alpha, \tag{7.21}$$
$$\rho_{\mathrm{S12,optimistic}} = 1 + 2\alpha, \tag{7.22}$$

represent the absolutely smallest transmission-to-processing ratios required in the SISO approach to render cooperative MIMO attractive from an energy point of view. For 2×2 MIMO we obtain

$$\rho_{\mathrm{S12,optimistic}} = 3 + 3\alpha, \tag{7.23}$$

which reveals a sizeable increase with respect to the smaller schemes. Comparing with the examples of existing nodes which we have given in Section 2.2.2, we see that very few of them can operate at the lowest limit, which for $\alpha = 1$ is $\rho_{\mathrm{S,optimistic}} = 3$. We are again, like in Chapter 6, seeing that current nodes seem to be slightly too processing dominated to reap the full energy gains from non-cooperative, that is single-hop SISO, communication. They will be forced to cooperation by their limited transmission range.

7.2.2 Achieved energy savings

Wennström (2002, p. 111) has shown that the diversity order gain of orthogonal space-time block codes over $n_t n_r$ independent Nakagami-m channels is precisely $n_t n_r$ (a series expansion of (7.8) confirms this). In this respect there is no difference between the use of transmit cooperation or receive cooperation, but, since we in our use of orthogonal space-time block coding (STBC) assume that channel state information is available on the receiver side only, receive cooperation will benefit from an array gain n_r which is absent in the case of transmit cooperation.[4] Considering further the diminishing returns from increased diversity and the increasing cooperative processing costs, we find that 1×2 MIMO will be the primary alternative to a direct SISO transmission. Larger MIMO constellations will spend significantly more processing energy – due an increased number of cooperating nodes and possibly also a lower code rate R_c – and can only be motivated when the 1×2 MIMO constellation becomes too transmission-dominated in its energy consumption (just like two hops always comes before three hops on the transmission-to-processing scale).

Uncertainty regarding the degree of fading. Basically, the reason that cooperative MIMO has the potential to drastically reduce the transmission energy is the poor performance of fixed-power SISO, rather than very good MIMO performance, in severely fading environments, for instance Rayleigh fading $m = 1$. Hence, the channel parameter of primary interest is here the Nakagami-m fading figure m. We would like to stress its importance again because of the prevalence of the pure-Rayleigh assumption in the literature, in spite of several studies showing the inadequacy of this assumption.[5] Our uncertainty regarding the Nakagami-m fading figure is quantified in (2.85)

---

[4]Our analysis is limited in this respect; we do not study the potential benefits from beam-forming or spatial multiplexing. However, the former requires tight transmit synchronisation and we comment on the problems of synchronisation in Section 7.5. The latter approach has no diversity advantage, which is the topic here.

[5]To head off a possible misunderstanding: we do not oppose to the assignment of a Rayleigh distribution *given that our background channel knowledge consists of average received power*; it is then the maximum entropy assignment and thereby well motivated. What we *do* object to is the assignment of a Rayleigh fading distribution *when it is known that other constraints on the fading are generally active*, see Section 3.3.2 for a discussion. Our use of one additional constraint, leading to the Nakagami-m fading distribution as the maximum entropy solution, is by no means the "final solution", but it does at least take the first order effects of other constraints into account (even if it is a weakness that the particular constraint in (2.64) on page 46 is difficult to motivate strongly on theoretical grounds).
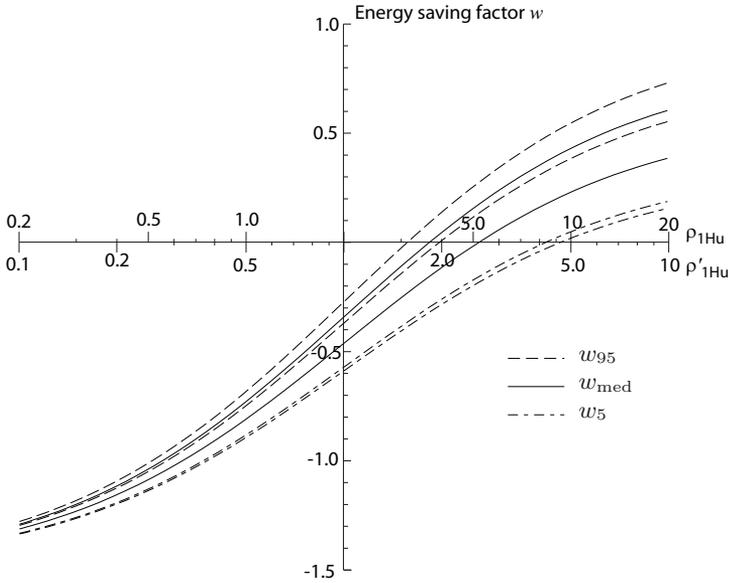
Figure 7.2: *Energy saving w, as a function of the SISO transmission-to-processing ratio ρ_S, achieved by a 1×2 cooperative MIMO scheme for target bit error rates $\overline{B} = 10^{-3}$ (lower curves) and $\overline{B} = 10^{-6}$ (upper curves) respectively. The percentiles come from our probability distribution for m in (2.85), and correspond directly to the percentiles for w. All results pertain to normalised receive energy $\alpha = 1$ and power amplifier degradation exponent $g = 2$.*

on page 54). We include in our results the median, the 5th percentile and the 95th percentile to present one likely result with error bars corresponding to 90 percent of the probability. Due to the fact that the transmission energy gain G_{cM} increases monotonically with decreasing fading figure m we can in our calculations use

$$w_5 \leftrightarrow m_{95} = 17.5$$
$$w_{med} \leftrightarrow m_{med} = 1.88 \quad (7.24)$$
$$w_{95} \leftrightarrow m_5 = 1.05.$$

Numerical results

In Figure 7.2 we show the energy saving w, under our uncertainty in the m parameter, as achieved by a 1×2 MIMO transmission. We include two different target bit error rates, namely $\overline{B} = 10^{-3}$ and $\overline{B} = 10^{-6}$. From Figure 7.2 we observe the following.

- In severe fading, $m_5 = 1.05$, and under a tough target bit error rate $\overline{B} = 10^{-6}$, the threshold transmission-to-processing ratio $\tilde{\rho}_S$ comes close to the optimistic threshold in (7.22) thanks to a very large gain $G_{cM} \approx 424$ (26 dB).

- There are good chances to save energy if $\rho_S > 5$, and the energy savings become consistent for $\rho_S > 10$. This is however more than currently available nodes regularly manage.

We reach, under the present circumstances, the following conclusion.

> Currently available node radios will generally not benefit energy-wise from cooperative MIMO-STBC even if they apply a link margin approach against the fading.

Differing conclusions

Why do we come to a quite different conclusion than other published works? For instance, in the well cited paper by Cui et al. (2004) it is stated that "tremendous energy saving is possible for transmission distances larger than a given threshold, even when we take into account the local energy cost necessary for joint information transmission and reception". While this is of course in a sense true[6] the problem is that the threshold distances given by Cui et al. (2004) correspond to a larger transmission-to-processing ratio than currently available nodes manage. We saw in Example 2.1 on page 20 how a threshold distance can change several orders of magnitude with relatively minor changes in the model parameters, yielding a potentially misleading conclusion. In the present case it is the link margin and the fading figure that make the difference: Cui et al. (2004) make use of a 40 dB link margin *and* a pure-Rayleigh assumption, thereby boosting the radiated energy per bit by several orders of magnitude. Unfortunately these assumptions are carried over to most citing papers, and nobody has, as far as we know, questioned the appropriateness of these assumptions. Considering our results on polarisation diversity and power control, together with practically available maximum transmission-to-processing ratios ρ_{max} and practically encountered fading figures m, it seems that most of the published results are relevant only in a very limited range of practical situations. If we ever come close to the required transmission-to-processing ratios, there is a good chance that channel inversion and polarisation diversity are better alternatives energy-wise than cooperation.

---

[6]Our calculations give the corresponding threshold in (7.19), and one can always set up a scenario in which the transmission costs are totally dominant.

7.3 Polarisation diversity or cooperative MIMO

In Chapter 5 we saw that a receiver utilising dual-branch polarisation diversity could facilitate a sizeable reduction in transmission energy at a possibly small increase in processing energy. The threshold transmission-to-processing ratios $\tilde{\rho}$ were smaller than the optimistic thresholds (7.21) and (7.22) for cooperative MIMO. Let us therefore study the choice between SISO transmissions with dual-branch receiver diversity and cooperative MIMO with simpler, single-antenna nodes. We assume that the diversity order gain of the polarisation scheme $\Delta = 2$, but that there is no array gain (this would correspond performance-wise to the switched diversity scheme in Chapter 5). Consequently, the transmission-wise difference between 1×2 MIMO and the polarisation diversity reception is the array gain achieved by the cooperative scheme – observe that this holds regardless of the fading figure m – yielding a radiated-energy gain of $G_{\text{cM12}} = 2$. For larger MIMO systems the gain will be larger. In Figure 7.3 we compare the two alternatives under different processing costs for the polarisation diversity scheme. We show the results against the SISO transmission-to-processing ratio ρ_{S} and, on the lower horizontal axis, against the polarisation diversity transmission-to-processing ratio

$$\rho_{\text{D}} = \frac{\rho_{\text{S}}}{G_{\text{D}}^{1/g}}, \tag{7.25}$$

where G_{D} is the diversity gain achieved by dual branch receiver diversity (D) with respect to single branch reception. In the present case we use the median fading figure $m_{\text{med}} = 1.88$ and the corresponding gain is then $G_{\text{D}} = 2.74$, see Chapter 5 and (5.4). The results reveal that only for transmission-to-processing ratios $\rho_{\text{S}} > 10$ can we expect a cooperative approach to outperform the direct transmission approach. Extending the cooperation to a 1×3 MIMO transmission does not help much due to the increased processing energy from cooperation. In terms of the transmission-to-processing ratio ρ_{D} the threshold is approximately $\rho_{\text{D}} = 6$.

From this short analysis we draw the subsequent conclusion

> Simple polarisation receiver diversity is preferred over cooperative MIMO-STBC and its use moves the benefits from cooperative MIMO to larger transmission-to-processing ratios ρ.
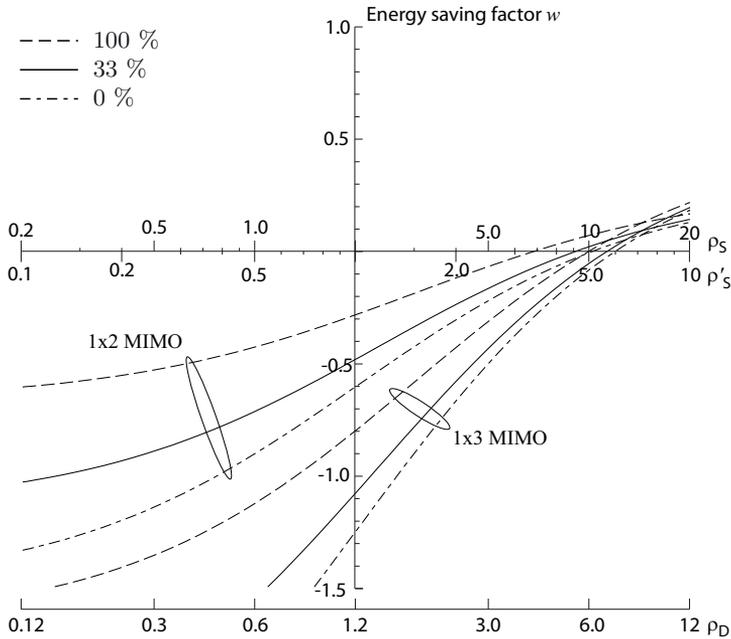
Figure 7.3: Energy saving w, as a function of the SISO transmission-to-processing ratio ρ_S, achieved by 1×2 and 1×3 cooperative MIMO over a SISO system employing dual-branch polarisation receiver diversity; the additional processing energy is assumed to be 0, 33 or 100 percent. Results for 1×2 MIMO are independent of the fading figure m, while the results for 1×3 MIMO pertain to $m_{\mathrm{med}} = 1.88$. The lower horizontal axis is the transmission-to-processing ratio ρ_D corresponding to the polarisation receiver diversity scheme. In all cases $\overline{B} = 10^{-3}$, $\alpha = 1$ and $g = 2$.

7.4   Multi-hop or cooperative MIMO

We have noted earlier, in Chapter 6, that cooperative schemes are penalised by the abrupt increase in processing energy consumption caused by the involvement of additional nodes in the transmission.[7] This has made us favour "gradual techniques" such as channel inversion and adaptive error correcting codes, or inherently processing-cheap techniques such as switched receiver diversity, over the cooperative schemes such as multi-hopping. Nevertheless, at some stage the transmission energy becomes dominant enough to render the cooperative schemes attractive energy-wise. There is a natural question at this stage: should we resort to multi-hopping or cooperative MIMO?

---

[7]The exception to the rule being cases when substantial aggregation/fusion benefits are readily available, reducing the processing costs of cooperation.

There is a quite rich literature on cooperative relaying, including multi-hop MIMO, and we wish to point out some differences between the present work and the most common approaches to this subject. The existing literature on cooperative techniques mostly treats the capacity/throughput aspect under transmit power constraints, without considering processing costs; see for instance Laneman and Wornell (2003) and Coso et al. (2007). Important and interesting as these results are, they do however address only half the problem of communication energy-efficiency when they leave out the processing part. Even if communication is possible at vanishing *transmit power*, the *total* energy consumption is not necessarily small; some papers advocate massive cooperation (hundreds of nodes) since it can reduce transmit costs, but they entirely neglect the (massive) processing costs, see for instance Bajwa et al. (2005). There are studies including processing costs – see for instance Jayaweera (2006) – but they are, to the best of our knowledge, all concerning the case of dominating transmission energy (large ρ). Unfortunately, the dominance of the transmission energy is most often implicit, hidden in a distance or similar. Therefore, our impression is that the existing literature might give an overly optimistic view as to the possible energy-wise benefits from cooperative diversity. Additionally, as we discuss briefly at the end of this chapter, there are a number of practical issues that are frequently overlooked. Our assessment is focussed on the impact of the processing energy, and is by no means comprehensive in terms of cooperative diversity schemes. Admittedly, there are many details regarding cooperative diversity that we leave out when concentrating on orthogonal space time block coding, but we believe that they are just details in comparison with the larger, first-order, energy effects we are including.

Before proceeding, we make the following assumptions.

Dense network. If we are to ignore the transmission costs (and the inverse diversity effects) during local MIMO cooperation – before and after the long-range MIMO transmission – the cooperating nodes must lie in close proximity of one another (here close is with respect to the long-range distance d). Thereby we feel it is reasonable to invoke the simplifying assumption that the relay nodes in the N-hop scheme are optimally positioned at fractions d/N of the total distance.

Equal amount of aggregation and fusion. We do not include packet aggregation or data fusion in the comparison since both schemes should have similar possibilities to benefit from them; if sampled data are redundant in one case, it ought to be it in the other.

Energy constrained destination. For simplicity we assume that the destination node R has the same energy consumption as the other nodes, meaning that $\alpha_R = \alpha$.

Both techniques, multi-hop and MIMO, reduce transmission energy, but they do so by very different means and are successful in diametrically opposed fading scenarios. Cooperative MIMO draws its benefits from diversity and the resulting resilience to fading, while multi-hopping suffers from "inverse diversity", see Definition 6.1 on page 171, and must rely on propagation loss reduction. By use of the energy consumption of uncoded multi-hop as given in Section 6.2, equation (6.8), we obtain

$$
\begin{aligned}
\breve{\mathcal{E}}_{\text{tot,NH}} &= \frac{\mathcal{E}_{\text{tot,NH}}}{\mathcal{E}_{\text{Pt}}} \\
&= \frac{N(\mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{Pr}} + \mathcal{E}_{\text{T,NH}})}{\mathcal{E}_{\text{Pt}}} \\
&= N(1 + \alpha) + N^{1 + \frac{1/m - \kappa}{g}} \rho_{\text{S}}.
\end{aligned}
\tag{7.26}
$$

By the use of the cooperative MIMO counterpart $\breve{\mathcal{E}}_{\text{tot,cM}}$, the righthand side in (7.16), we find that the fraction of energy saved by the use of cooperative MIMO with respect to multi-hop is given by

$$
\begin{aligned}
w_{\text{cM}} &= \frac{\breve{\mathcal{E}}_{\text{tot,NH}} - \breve{\mathcal{E}}_{\text{tot,cM}}}{\breve{\mathcal{E}}_{\text{tot,NH}}} \\
&= \left(N^{1 + \frac{1/m - \kappa}{g}} - G_{\text{cM}}^{-1/g} \right) \frac{\rho_{\text{S}}}{N(1 + \alpha) + N^{1 + \frac{1/m - \kappa}{g}} \rho_{\text{S}}} \\
&\quad + \frac{\left(N + 1 - u(n_{\text{t}} - 2) - \frac{n_{\text{t}}}{R_{\text{c}}} - n_{\text{r}} \right) + \left(N + 1 - n_{\text{t}} - \left(1 + \frac{1}{R_{\text{c}}} \right) n_{\text{r}} \right) \alpha}{N(1 + \alpha) + N^{1 + \frac{1/m - \kappa}{g}} \rho_{\text{S}}},
\end{aligned}
\tag{7.27}
$$

where the first and last terms represents the difference in transmission energy and processing energy respectively. For each ρ_{S} we choose the energy-optimal number of hops and the energy-optimal $n_{\text{t}} \times n_{\text{r}}$ cooperative MIMO scheme.

Over the range of transmission-to-processing ratios we consider here, our numerical searches have not resulted in more than two transmit nodes, $n_{\text{t}} = 2$, even under the favourable assumption that the space-time block code rate $R_{\text{c}} = 1$ for all n_{t} (it can be shown that $R_{\text{c}} < 1$ for all $n_{\text{t}} > 2$ (Larsson and Stoica, 2003, App. B)). We therefore make the following simplifying assumption.
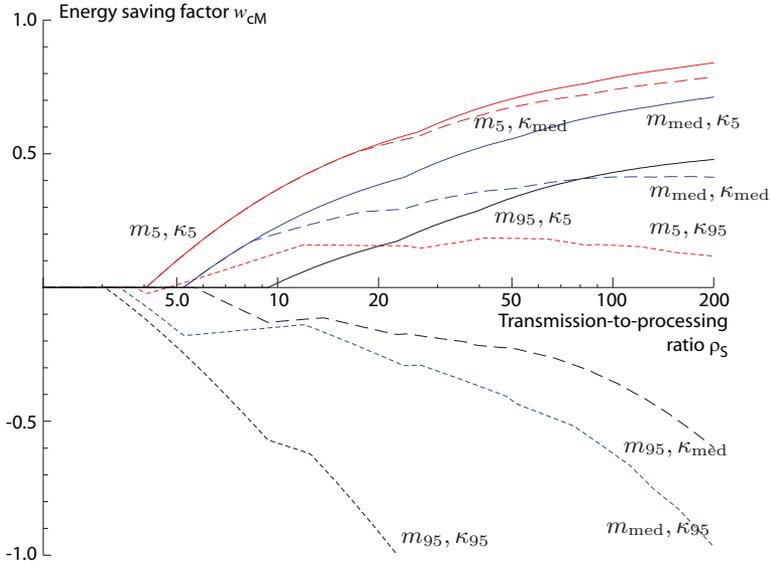
Figure 7.4: *Comparison between N-hop and $n_t \times n_r$ MIMO, shown as the energy saving w_{cM} of (7.27). The flat portion of the curves correspond to single-hop SISO. All curves were calculated for $\alpha = 1$ and $g = 2$.*

Assumption 7.1 *An orthogonal space-time block code with code $R_c = 1$ can be found for all the MIMO systems of present interest.* ∎

As usual we include the uncertainties in the Nakagami fading figure m and the propagation loss exponent κ by use of the probability distributions given in (2.85) and (2.79). More precisely, we use the absolute-error optimal estimates, the medians $m_{med} = 1.88$ and $\kappa_{med} = 3.32$, together with the 5th and 95th percentiles $m_5 = 1.05$, $\kappa_5 = 2.00$, $m_{95} = 17.5$ and $\kappa_{95} = 5.12$. We study the energy saving over a large range of SISO single-hop transmission-to-processing ratios[8] ρ_S in Figure 7.4, using combinations of the percentiles to illustrate the uncertainty. First, there is a striking spread in the results, but this was not unexpected because of the diametrically opposed impact that fading has on multi-hop and cooperative MIMO respectively. When fading is not severe, multi-hop can outperform cooperative MIMO (the curves that are below the horizontal axis), but it seems that cooperative MIMO is slightly

---

[8]Remember that the concentration of the transmission-to-processing ratios for multi-hop and cooperative MIMO, ρ_{NH} and ρ_{cM}, will keep the actual transmission-to-processing ratio below ρ_S. However, the figure becomes too difficult to interpret if we use ρ_{NH} or ρ_{cM} on the horizontal axis – there will be abrupt changes in the transmission-to-processing ratios as observed in Figures 6.4 and 6.5.

more energy efficient overall. For instance, when we consider the median values m_{med} and κ_{med}, the negative effect that the fading has on multi-hop outweighs the gain from shorter hops and cooperative MIMO offers a lower energy consumption.

7.4.1 Polarisation diversity receivers

The inherent problem of the multi-hop approach is its sensitivity to fading, manifested in Figure 7.4, and we have previously advocated the use of polarisation receiver diversity to counter fading, see Chapter 5. Consider therefore a dual-branch receiver with diversity order gain $\Delta = 2$, array gain $A_{\mathrm{D}} = 1$ (see Definitions 3.4 and 3.1) at an additional receive processing cost of 10 percent

$$\alpha_{\mathrm{D}} = \frac{\mathcal{E}_{\mathrm{Pr,D}}}{\mathcal{E}_{\mathrm{Pt}}} = \frac{1.10\mathcal{E}_{\mathrm{Pr}}}{\mathcal{E}_{\mathrm{Pt}}}1.10\alpha. \tag{7.28}$$

This would roughly correspond to a simple switched diversity (SwD) receiver with 10 percent extra transmissions. Note that if we were to use for instance maximum ratio combining, the array gain A_{D} would be larger but also the additional processing cost. Our results do not cover all possible alternatives but ought to provide useful indications.

We now compare the use of multi-hop and polarisation receiver diversity on the one hand, with cooperative MIMO with single-antenna radios on the other. The energy consumption, as usual normalised by the transmit processing energy per bit, becomes

$$\begin{aligned}
\breve{\mathcal{E}}_{\mathrm{tot,NH,D}} &= \frac{\mathcal{E}_{\mathrm{tot,NH,D}}}{\mathcal{E}_{\mathrm{Pt}}} \\
&= \frac{N\left(\left(\mathcal{E}_{\mathrm{Pt}} + \mathcal{E}_{\mathrm{Pr}} + \left(\frac{\mathcal{E}_{\mathrm{T,NH}}}{G_{\mathrm{D}}}\right)^{1/g}\right)\right)}{\mathcal{E}_{\mathrm{Pt}}} \\
&= N(1 + \alpha_{\mathrm{D}}) + N\left(\frac{N^{1/m-\kappa}}{G_{\mathrm{D}}}\right)^{1/g}\rho_{\mathrm{S,1H}},
\end{aligned} \tag{7.29}$$

where $\rho_{\mathrm{S,1H}}$ is the SISO single-hop transmission-to-processing ratio and G_{D} is the dual-branch receiver diversity gain which depends on the fading figure m and the target bit error rate B (see Chapter 5). Figure 7.5 shows the expected impact that the multi-hop scheme performs significantly better in severe fading if receiver diversity can be exploited. Note that when the curves are below $w_{\mathrm{cM}} = 0$, the multi-hop, receiver diversity, approach is better, while the cooperative MIMO, single antenna, approach is preferable otherwise. The extra
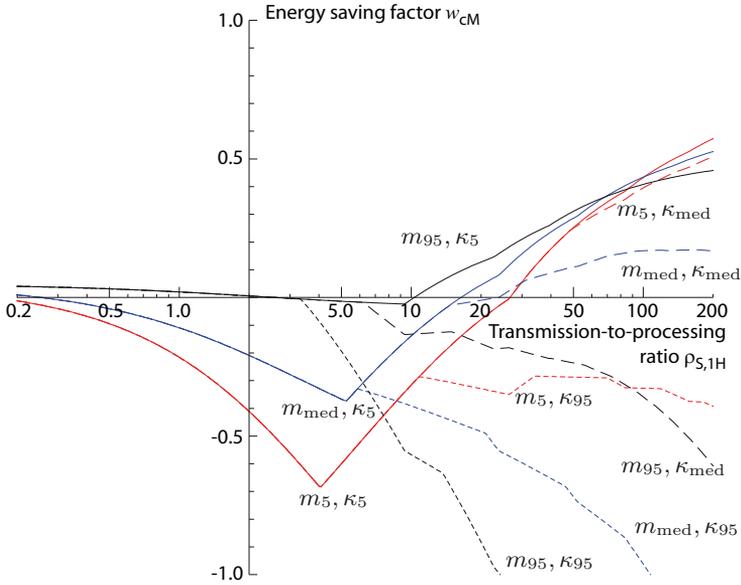
*Figure 7.5: Comparison between cooperative MIMO with single-antenna
nodes, and N-hop communication with dual-branch receiver diversity (no
array gain assumed). The receiver diversity alleviates the multi-hop scheme
from much of its inverse diversity burden and gives it a slight advantage
over cooperative MIMO.*

processing incurs a slight penalty for small transmission-to-processing ratios.
Regarding Figure 7.4 we feel that the following observations are motivated.

- For transmission-to-processing ratios $\rho_{S,1H} < 5$, the comparison is es-
  sentially between SISO with, and SISO without, polarisation receiver
  diversity. This is because 1×1 MIMO, that is to say SISO, is included
  as a special case of MIMO in the calculations. The sharp turns in the
  solid curves correspond to the switch from SISO to 1×2 MIMO.

- The polarisation receiver diversity scheme has a significantly smaller
  processing cost than the cooperative schemes. Hence, it provides an
  advantage over cooperative MIMO for transmission-to-processing ra-
  tios $\rho_{S,1H} < 10$, and remains a good alternative up to $\rho_{S,1H} \approx 20$.
  Upon comparing Figure 7.5 with Figure 7.4 the difference is evident:
  polarisation receiver diversity improves the multi-hop energy efficiency
  significantly for severe fading, that small m.

7.5 Concluding remarks

Cooperative MIMO based on orthogonal space-time block codes provides a reduction in transmission energy which is substantial for severe fading, but the cooperative approach achieves this at a relatively large processing energy cost. It is therefore advisable to consider less processing intensive diversity schemes before resorting to cooperation, and our results show that polarisation receiver diversity makes cooperative MIMO significantly less attractive.

Regarding longer distances, when dual-branch receiver diversity is not enough to sustain energy efficient single-hop, our our analysis shows that the choice between multi-hop and cooperative MIMO is not easily settled. The reason is that the propagation loss exponent κ and the Nakagami fading figure m have significant impact on the outcome, and under our present uncertainty we can not draw a general and firm conclusion. If we consider single-antenna radios only, see Figure 7.5, the cooperative MIMO scheme outperforms multi-hop for most – but not all – scenarios, mainly because the multi-hop transmission energy is significantly increased by the inverse diversity effect in Definition 6.1. The use of dual-branch receiver diversity results in a substantial reduction of this inverse diversity effect, see Figure 7.4. Multi-hop communication needs some form of processing efficient diversity to perform well, and our suggestion is to then use polarisation receiver diversity.

Ignored facts and simplifying assumptions. To put our results in perspective, we wish to stress some neglected issues and the influence of the simplifying assumptions we have utilised.

Channel inversion. As shown in Chapter 4, the use of channel inversion in place of a fixed link margin can reduce the total energy consumption for transmission-to-processing ratios $\rho > 1/2$. Therefore, given that a nodes ρ_{max} is large enough, the use of channel inversion will reduce the relative benefits from cooperative MIMO. Especially if we combine channel inversion through transmit power control with receiver diversity, large gains ought to be achieved, and multi-hop would then be less affected by severe fading.

Node positions. The present analysis assumes ideal placement of cooperative nodes, both for multi-hop and for MIMO-STBC. If the network is not very dense with respect to the end-to-end distance, the lack of appropriate collaborators can constitute a problem. The use of widely

spaced cooperative MIMO nodes may introduce non-negligible inverse diversity effects, and we have already studied the effects of position uncertainty on multi-hop in Chapter 6.

Synchronisation. We have presupposed that the MIMO nodes are perfectly synchronised in their transmissions and receptions. Achieving a high degree of synchronisation is not easy, and any errors can only be negative for the cooperative MIMO performance Jagannathan et al. (2004).

Appendix 7.A MIMO-STBC bit error rate

According to Wennström (2002), orthogonal space-time block coding over $n_t n_r$ identical and independent Nakagami-m fading channels achieves an array gain n_r. Therefore, the combined average received signal-to-noise ratio per bit is $\overline{\gamma}_{\mathrm{cM}} = n_r \overline{\gamma}$. Because of the fact that the $n_t n_r$ channels, each with a gamma distribution

$$p(\gamma|m, \overline{\gamma}, I) = \frac{m^m}{\Gamma(m)\overline{\gamma}^m} \gamma^{m-1} e^{-m\frac{\gamma}{\overline{\gamma}}} \tag{7.30}$$

for its signal-to-noise ratio per bit, are combined in a coherent fashion, the resulting instantaneous signal-to-noise ratio per bit γ_{cM} has the distribution

$$p(\gamma_{\mathrm{cM}}|n_t, n_r, m, \overline{\gamma}, I) = \frac{(mn_t n_r)^{mn_t n_r}}{\Gamma(mn_t n_r)(n_r \overline{\gamma})^{mn_t n_r}} \gamma^{mn_t n_r - 1} e^{-mn_t n_r \frac{\gamma_{\mathrm{cM}}}{n_r \overline{\gamma}}}. \tag{7.31}$$

Wennström (2002) shows that the diversity order gain is $n_t n_r$, which exactly corresponds to going from (7.30) to (7.31). Now, by use of the bit error rate expression for (square) quadrate amplitude modulation given in (4.59) on page 125, we can find the average bit error rate by the use of mathematical software (Wolfram Research Inc., 2007), or tabulated integrals (Gradshteyn and Ryzhik, 2000);

$$
\begin{aligned}
\overline{B} &= \int_0^\infty B(\gamma_{\mathrm{cM}}) p(\gamma_{\mathrm{cM}}|n_t, n_r, m, \overline{\gamma}, I) \mathrm{d}\gamma_{\mathrm{cM}} \\
&= (-1)^{-mn_r n_t} \frac{2}{b\sqrt{\pi}} \frac{\sqrt{2^b} - 1}{\sqrt{2^b}} \frac{\Gamma\left(mn_r n_t + \frac{1}{2}\right)}{\Gamma(mn_r n_t)} \\
&\quad \times \beta\left(-\frac{2\left(2^b - 1\right) mn_t R_c}{3b\overline{\gamma}}, mn_r n_t, \frac{1}{2} - mn_r n_t\right).
\end{aligned} \tag{7.32}
$$

Chapter 8

Network Measurement Capacity

O UR focus has thus far been on energy consumption, but is energy efficiency really the adequate metric for the network as a whole? Certainly, in many sensor networks energy will constitute a limited resource which must be spent judicially if long-lasting network functionality is to be sustained at a sufficient level of quality. Energy efficiency is then a positive feature of, say, a transmission scheme. But, as we noted in the beginning of Chapter 6, aiming only for energy-efficiency may lead to unbalanced work load among the sensor nodes with degraded network operation as a consequence. Intuitively, we want to avoid severely unbalanced load in a sensor network due to the detrimental effects it may have, for example through rapid energy depletion in certain regions of the network. On the other hand, we realise that *requiring a perfect load balance* may cause very energy-inefficient operation. Indeed, as Perillo et al. (2004) find in their study of the energy imbalance problem, "energy balancing can be achieved only at the expense of gross energy inefficiencies".

At the heart of the matter is the fact that energy is not a *central* sensor-network resource, but a *distributed* resource with energy 'quanta' residing in individual nodes. By contrast, the performance of the *network* is generally perceived in holistic terms, and we do not actually care about the individual nodes as long as the network as a whole operates satisfactorily.

The topic of this chapter is the development of our notion of *measurement capacity* as a network-wide common resource. The measurement capacity is based on the *number of different sequences of events/requests* that the network can respond to at a given energy budget, and it thereby entails both energy efficiency and load balancing. In other words, it is a measure

of *the network's readiness to respond to any one of a large class of different events and/or measurement requests.* It is the size of the class of possible events/requests that defines the measurement capacity, and in this sense it bears resemblance to the capacity of a (noiseless) channel as its readiness to transmit any one of a large class of messages.[1] We present here the first steps in the development of measurement capacity and later in the chapter we use it for assessment of multi-hopping in shadowed and non-shadowed environments. We finally make an initial assessment of the benefits from heterogeneous hierarchical networks in terms of measurement capacity.

8.1 Measurement capacity as the event multiplicity

Consider an M node network with a topology, sensing task, communication structure and energy allocation that allows node m to make n_m measurements (including all processing and transmission cost). The network can perform $N = \sum n_m$ measurements in total and we can relate this situation to our initial discussion of energy efficiency and load balancing as follows:

- A design strategy based on energy efficiency would implicitly address maximisation of N, the total number of measurements.

- A design strategy based solely on even load distribution among nodes would aim for identical n_m, whatever the consequences for N.

Our proposal is to consider *the number of different measurement/event sequences that the network can meet successfully.* This number will measure the network's readiness to respond to different tasks, its flexibility to meet an unknown future of events/requests. The number of different ways to distribute N measurements such that node m perform n_m of them is given by the multinomial coefficient

$$W = \frac{N!}{n_1! n_2! \ldots n_M!},\tag{8.1}$$

see for instance (Sivia, 1996). The reader might at this point anticipate where we are heading by recalling the Wallis derivation of the Shannon entropy (see Section 2.3.2 on page 37). By the use of the Stirling approximation in (2.38)

---

[1] We have borrowed the words from Jaynes (2003, p. 629).

on page 38 we can, for large n_m, write

$$
\log\left(W\right) \approx N \log(N) - \sum_{m=1}^{M} n_m \log(n_m)
$$

$$
= -N \sum_{m=1}^{M} \overline{n}_m \log(\overline{n}_m),
$$

(8.2)

where $\overline{n}_m = n_m/N$ are the normalised numbers of measurements at each node.[2]

The $\log(W)$ term can be viewed as the size of the class of all measurement sequences that the network can handle, and in (8.2) we see that it can be approximated by the entropy $-\sum \overline{n}_m \log(\overline{n}_m)$ of the measurement distribution $\overline{n}_m$, weighted by the total number of measurements N.

Generalisation from nodes to measurement cells. Depending on the application and the sensor deployment, nodes may or may not constitute the adequate "space" on which we define the number of measurement n_m. In some scenarios, for instance when several nodes are placed very close to each other and cover the same spatial area of interest, it is more appropriate to think of the network as consisting of *M cells* from each of which n_m measurements can be obtained: the cells constitute the definition of our "measurement space" (we elaborate on the choice of cells in Section 8.1.2).

Definition 8.1 *The* measurement capacity *of an M cell sensor network is*

$$
\mathcal{C} \equiv -N \sum_{m=1}^{M} \overline{n}_m \log(\overline{n}_m),
$$

(8.3)

where n_m is the number of measurements obtainable from cell m and N is the total number of measurements: $\overline{n}_m = n_m/N$ and $N = \sum_{m=1}^{M} n_m$. ∎

---

[2]To be precise, the Stirling approximation contains an additional term, yielding the more accurate approximation

$$
\log\left(W\right) \approx -N \sum_{m=1}^{M} \overline{n}_m \log(\overline{n}_m) - \frac{1}{2} \sum_{m=1}^{M} \log\left(\frac{n_m}{N^{1/M}}\right).
$$

As N grows very large, the *absolute error* in (8.2) will grow without bound, but the *relative error* is small since the first term in the above expression – the term we use in (8.2) – grows much faster with N than the second – the absolute error term which we have discarded.

Observe that nothing prevents cells and nodes from being equivalent in some applications – indeed, it seems reasonable that deployments actually should be conducted in a manner that yields one node per cell.

Remark 8.1 *Remember that our definition of the measurement capacity rests, through our use of the Stirling approximation (2.38), on the assumption that all n_m are large. For small n_m we should go back to the multinomial term W in (8.1). Observe that if one $n_m = 0$, its corresponding factor $n_m! = 0! = 1$ in the denominator of the multinomial coefficient (8.1) and the problem is essentially reduced to $M - 1$ cells (or nodes). Cells (nodes) close to complete exhaustion thus contributes very little to the measurement capacity.*

For illustration, let us compare the measurement capacity metric $\mathcal{C}$ with the two related approaches of load balancing and minimisation of energy consumption.

Load balancing. The maximum entropy for $\overline{n}_m$ is attained by a uniform distribution – a balanced load – of the normalised measurement capability $\overline{n}_m = 1/M$ (see Section 2.3.3). The entropy factor $-\sum \overline{n}_m \log(\overline{n}_m)$ in (8.3) thus captures the intuitive feeling that it is good to distribute the load uniformly over the nodes (cells) in the network: an even distribution of $\overline{n}_m$ leads to a greater "readiness" of the network.

Energy efficiency. Adoption of an energy efficiency strategy would mean maximisation of N: make the design choices that achieves the maximum number of measurements per unit energy. The measurement capacity (8.3) anticipates the benefits from energy efficiency through the N factor: more measurements increase the network's "readiness" to encounter many different scenarios.

In $\mathcal{C}$ we have a metric that combines an even load distribution with energy-efficiency in a way that measures the network's readiness to handle many different event sequences.

8.1.1   Properties of the measurement capacity

Before we proceed, we shall point out four important properties of the measurement capacity $\mathcal{C}$ in (8.3).

More energy increases measurement capacity. If we double the energy resource at each node such that $n'_m = 2n_m$, the network measurement capacity is doubled, $\mathcal{C}' = 2\mathcal{C}$. *There is a linear increase in*

$\mathcal{C}$ with N as long as the distribution $\overline{n}_m$ does not change. Thereby, if the distribution $\overline{n}_m$ is kept constant throughout the network's period of operation, there is through N a direct correspondence between the measurement capacity $\mathcal{C}$ and energy efficiency.

The cost of a measurement. All measurements will reduce the remaining measurement capacity, but the single measurement which reduces the measurement capacity the most is the one performed by the sensor with the smallest n_m. The use of the sensor with the largest n_m will cause the smallest reduction in measurement capacity. All measurements cause the same decrease in the total number N, but exploiting the sensor with smallest n_m accentuates the existing unbalance.

Hierarchical capacity. Forming a larger network by combining K networks with respective sensing capacities $\mathcal{C}_k$ will increase the measurement capacity to

$$\mathcal{C} = \mathcal{C}'_K + \sum_{k=1}^{K} \mathcal{C}_k, \tag{8.4}$$

where

$$\mathcal{C}'_K = N \log(N) - \sum_{k=1}^{K} N_k \log(N_k) \tag{8.5}$$

is the measurement capacity of the super network made of K sub-networks with N_k measurements each, $N = \sum_k^K N_k$, and

$$\mathcal{C}_k = N_k \log(N_k) - \sum_{m=1}^{M_k} n_{m_k} \log(n_{m_k}) \tag{8.6}$$

are the sensing capacities of the sub-networks, where $N_k = \sum_{m_k=1}^{M_k} n_{m_k}$. Here, M_k is the number of cells (nodes) in sub-network k, and n_{m_k} is the number of measurements that cell (node) $m_k = 1, 2, \ldots, M_k$ in sub-network k can perform. We can confirm the relation in (8.4) by

the use of (8.5) and (8.6),

$$
\begin{aligned}
\mathcal{C} =& \mathcal{C}'_K + \sum_{k=1}^{K} \mathcal{C}_k \\
=& \left(N \log(N) - \sum_{k=1}^{K} N_k \log(N_k) \right) \\
& + \sum_{k=1}^{K} \left(N_k \log(N_k) - \sum_{m=1}^{M_k} n_{m_k} \log(n_{m_k}) \right) \quad (8.7) \\
=& N \log(N) - \sum_{k=1}^{K} \left(\sum_{m=1}^{M_k} n_{m_k} \log(n_{m_k}) \right) \\
=& N \log(N) - \sum_{m=1}^{M} n_m \log(n_m),
\end{aligned}
$$

where the last line is precisely the measurement capacity $\mathcal{C}$ in Definition 8.1. *The measurement capacity of combined networks is larger than the sum of their individual sensing capacities.*

Single cell networks. If there is only one cell left with measurements, cell k say, then $n_k = N$ and all other $n_{m \neq k} = 0$. Consequently, measurement capacity is zero,

$$
\mathcal{C} = N \log(N) - n_k \log(n_k) = 0. \quad (8.8)
$$

At first, this might seem an undesirable property of the metric, because then the number of measurements N has no importance. Let us therefore consider the hierarchical capacity (8.4) above. From the large-scale perspective in which there is only one cell, the measurement capacity $\mathcal{C}'_K$ in (8.5) *is* zero because only one sequence of measurements can be performed. However, if the macro-cell has internal cells $k = 1, 2, \ldots, K$, it is only the "macroscopic" capacity $\mathcal{C}'_K$ in (8.4) that is zero. The subcell capacity $\mathcal{C}_k$ in (8.6) is non-zero and the total capacity is thus non-zero. Once we get down to an individual sensor node, it is only the "internal cell space" of that sensor node left. For instance, if we have one sensor node that detects the presence of a target, then it has two internal cells and a non-zero measurement capacity. *It is only if the sensors have only one response (one output) that the measurement capacity becomes zero.*[3]

---

[3] And such sensor are of course useless.

This last point regarding the single cell network and "internal cells" raises the question of how to define the cells appropriately.

8.1.2 Defining the correct measurement space

There are apparently connections between our measurement capacity and the Shannon entropy. With regard to the entropy of a probability distribution, and its maximisation in inference problems, it is important to note the role of the hypothesis space with respect to which the entropy is defined. We encountered this subtlety in Example 2.3 on the Poisson distribution, where an apparently correct application of the principle of maximum entropy was lead astray by our use of the wrong hypothesis space[4]. The following remark is worth contemplating:

> [T]he simple, unqualified term 'entropy' is meaningless; it is always defined with respect to some basic 'measure' and the result of maximizing it depends not only on the constraints, but also on the measure. The difficulty in applying maximum entropy to problems outside thermodynamics is not in deciding what constraints should be applied, but in deciding what is the underlying measure – or, as I prefer to call it, what is the hypothesis space on which our entropies are defined?

> Jaynes (1985)

Jaynes' words appear in his paper on entropy and search theory where he points to the cogency of the sizes of the searched cells when devising an optimal target search plan on information theoretic grounds. The choice of cell sizes amounts to a choice of hypothesis space, and therefore it is not only the probability of finding the target in a cell that is important: the size of each cell must be taken into account when planning the search.

In the present context of measurement capacity we face a similar situation since the capacity $\mathcal{C}$ results from our definition of the number of measurements per cell n_m – that is our definition of the measurement space. This is emphasised by considering the hierarchial relationship in (8.4) and the discussion of single-cell capacity above. Just like for inference problems, where the question of the correct hypothesis space is open ended and can not be settled once and for all, the question of the correct *measurement space* can not be given an all-encompassing answer. In some cases the node space

---

[4]The hypothesis space we choose to work on implicitly defines our ignorance measure q(x) which is subsequently used in the MaxEnt procedure, Section 2.3.3.

will be perfectly adequate as a measurement space (each node forming a cell), but in others it will definitely not be a good choice (for instance when measuring a distributed field with all sensors at the same time).

We believe that it is the definition of the measurement space that is the key to successful use of $\mathcal{C}$ in sensor network design, but it is also the most difficult step, and it is application specific. Therefore, unless otherwise stated, *it is henceforth simply assumed that the measurement space we use is adequate for the application.*

8.1.3 Connecting energy and measurements

The measurement capacity is defined in terms of the number of measurements per cell n_m, but these numbers are of course dependent on several properties of the sensor network. For example, the energy consumed by the collection of a sensor reading, the energy consumed by wireless communication of the data, the energy available in the cell's nodes' batteries and the topology of the network all influence the numbers n_m.

Let us for simplicity start by assuming that each node in cell m reports directly to the central sink at a per-measurement energy cost $s_m > 0$, including the communication costs that may vary significantly across the network.[5] We then have a network with the measurement capacity

$$\begin{aligned}
\mathcal{C} &= N \log(N) - \sum_{m=1}^{M} n_m \log(n_m) \\
&= \left(\sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \right) \log \left(\sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \right) - \sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \log \left(\frac{\mathcal{E}_m}{s_m} \right),
\end{aligned} \tag{8.9}$$

where $\mathcal{E}_m \gg s_m$ is the energy resource in the mth cell. Consider now the problem of determining the optimum distribution of a total amount of energy $\mathcal{E}_{\text{tot}}$ among the cells. This corresponds to some extent to the uneven deployment approach to counteract load imbalance, see for instance Rivas et al. (2006) or Prasad and Agrawal (2007). We formulate the optimisation problem

$$\max_{\{\mathcal{E}_m\}} \left(\sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \right) \log \left(\sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \right) - \sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \log \left(\frac{\mathcal{E}_m}{s_m} \right) \tag{8.10}$$

---

[5]Observe that no cooperation among nodes is assumed for simplicity. Inclusion of cooperative transmissions, such as multi-hopping, makes the expression much more complicated due to the interconnections between the number of measurements at the nodes. We consider the multi-hop case later in Section 8.2.

such that

$$\sum_{m=1}^{M} \mathcal{E}_m = \mathcal{E}_{\text{tot}}.$$ (8.11)

We can readily address this constrained maximisation problem by the use of Lagrange multipliers, whose application lead to the system of $M+1$ equations

$$\mathcal{E}_k = s_k \left(\sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \right) e^{-\lambda s_k} \quad , \qquad k = 1, 2, \ldots M$$ (8.12)

$$\sum_{m=1}^{M} \mathcal{E}_m = \mathcal{E}_{\text{tot}}.$$ (8.13)

Lacking an analytical solution for the $\mathcal{E}_m$, we can still gain some insight into the optimum solution by noting that (8.12) can be restated as

$$\overline{n}_k = \frac{\mathcal{E}_k}{s_k} \left(\sum_{m=1}^{M} \frac{\mathcal{E}_m}{s_m} \right)^{-1} = e^{-\lambda s_k} \quad , \qquad k = 1, 2, \ldots M.$$ (8.14)

From this we conclude that energy should be distributed so that the fractions of all measurements thereby allocated to the cells decay exponentially with the sensing cost s_m. *This fact highlights the compromise between obtaining many measurements, large N, and spreading them evenly, even-sized n_m's.* If we instead, hypothetically, would have a fixed number of measurements to distribute, the largest measurement capacity would be achieved by a uniform distribution. The above energy distribution is in exact correspondence with the Shannon capacity for a noiseless channel when the symbols have different transmission time s_m (Jaynes, 2003, pp. 630–632).

Before we, armed with the measurement capacity in (8.3), return to the assessment of multi-hop as an efficient transmission technique, let us outline work related to our measurement capacity.

8.1.4   Related work.

We have found three broad topics that are relevant to discuss in the present context.

Network lifetime.   A commonly used network-wide metric is the network's lifetime, proposed by Chang and Tassiulas (1999) to be the time of operation until the death of the first node. This network lifetime metric is

widely used and does in some sense combine both energy efficiency and energy balance.[6] There are a number of other, related, lifetime metrics, defined on the number of functioning nodes, the network's coverage, the connectivity of the network, etc. A survey is given by Dietrich and Dressler (2006) and the interested reader is referred to their report for further reading. Cheng et al. (2008) have recently performed a quite thorough study of deployment strategies under the sensor network lifetime metric of Chang and Tassiulas (1999). They conclude, among other things, that "A good sensor network deployment strategy is one that achieves both energy balance and energy efficiency".

While the network lifetime metric captures important aspects of the sensor network's ability, it has some shortcomings. First, a sensor network is not completely useless as soon as one node stops functioning, except possibly in some very rare application (robustness against node failures is moreover important and a research topic of its own). If, say, 99 out of 100 nodes are operational we can likely still put the network to good use, although the network has inexorably become *less useful* as nodes have stopped functioning. Second, the metric focuses on the worst case only and does not encourage purposeful use of the other nodes' resources as long as they fare better than the most short-lived node. Mhatre et al. (2005) hold that it is "important to ensure... that almost all the nodes expire at about the same time" because under the lifetime metric all residual energy is considered wasted energy. Chen and Zhao use the *expected* lifetime "measured as the average amount of time until the network is considered nonfunctional" (Swami et al., 2007, Ch. 5). Also here residual energy is a concern and it is argued that toward the end of the lifetime it therefore becomes important the balance the load in the network.

In our measurement capacity $\mathcal{C}$ we have found a metric that combines an even load distribution with energy-efficiency, and we believe that it makes precise the conclusion by Cheng et al. (2008) that "a good sensor network deployment strategy is one that achieves both energy balance and energy efficiency". Good designs strike a balance between even load and energy efficiency. Comparing measurement capacity with network lifetime, we note the following difference: the measurement capacity allows heavy load on individual nodes (cells) if the gains are large enough. More precisely, one or a few nodes (cells) can be "sacrificed" to make it easier for the rest of the

---

[6]Quite commonly, the lifetime metric proposed by Chang and Tassiulas (1999) is modified to the be the time of operation until a certain percentage of the nodes run out of energy.

network: if the decrease in the entropy factor $-\sum \overline{n}_m \log(\overline{n}_m)$ is outweighed by an increase in N the sacrifice is accepted, see (8.3). By contrast, when one node is very heavily loaded, the network lifetime metric does not care about any other nodes, and they must support the heavily loaded node at almost any cost.

Sensing capacity. The concept of sensing capacity, as proposed by Rachlin et al. (2005) and Aeron et al. (2007), concerns only the sensing aspect. The two notions of sensing capacity given by Aeron et al. (2007) and Rachlin et al. (2005) are similar but not identical, and for brevity, we focus here on the latter as described in Swami et al. (2007, Ch. 4). Rachlin et al. (2005) models the network's region of interest as divided into smaller spatial regions – let us call them cells – in which targets are either absent or present. Each sensor node observes one or several cells *jointly*, and the sensor output is therefore possibly affected by several targets at a time. For instance, vibrational sensors can detect signals from several vehicles at the same time. The question that the sensing capacity of Rachlin et al. (2005) answers is: what is the minimum number of sensors required to achieve a certain distortion (a certain detection probability)?

In one sense, our measurement capacity and the described sensing capacity operate at the same level, the cells which form the underlying space of interest. In another sense, the two metrics are on different levels as the sensing capacity of Rachlin et al. (2005) concerns the achievable sensing performance per sensor, while our measurement capacity concerns in how many ways the network can be used to perform sensing, including sensing and communication costs. It might be that the two approaches can be joined into one more general metric; we provide in Appendix 8.A a few tentative generalisations of our measurement capacity $\mathcal{C}$ that could facilitate a combined metric.

Bounds on communication capacity. Complementary to Rachlin et al. (2005) and Aeron et al. (2007), there is also work that concerns mainly the *communication* constraints in wireless sensor networks, see the chapter by Gastpar in Swami et al. (2007, Ch. 2). The main topic is how to communicate the sensed information most efficiently from the sensors under a constraint for the distortion in the reconstruction of the desired quantity; which are the ultimate limits for this network communication. The power constraint typically concerns only radiated power, and as it has been shown that the source-channel separation theorem does not hold, so-called analog techniques

have been proposed in place of the digital one we commonly consider. The basic approach is to amplify the observations at each sensor (no regular source coding) and transmit them from all sensors coherently to achieve two benefits. First, the sensor noise is averaged out in the channel. Second, the received signal-to-noise power ratio is increased by the coherent combining of the signals (Bajwa et al., 2005, Gastpar et al., 2006). The communication bounds can of course be important, but in our framework it is total energy consumption that is in focus.

8.2   Measurement capacity in multi-hop networks

In Chapter 6 (and partly also in Chapter 7) we assessed the energy efficiency of multi-hopping as compared to direct transmission (and cooperative MIMO). One main finding was that many of the present node radios, due to insufficient output power, would be forced to use multi-hopping at shorter ranges than motivated by energy-efficiency. We noted already then that the energy-efficiency viewpoint, although yielding very informative results, misses the aspect of load balance, but we deferred its treatment to this chapter. An alternative approach, which we adopted in (Björnemo et al., 2006), is to study energy efficiency and network load imbalance in parallel. However, it has not been clear how these two metrics should be combined to give a fair overall judgement, nor is it clear how the network lifetime metric combines these metrics. As the choice of multi-hop in place of single-hop alters both the overall energy consumption *and* the distribution of the load, it appears that a network-wide assessment must include both aspects.

Now it seems that our notion of measurement capacity provides a method for combining both aspects and shows us a constructive way to assess the network-wide impact of the transmission choice. In this section we therefore readdress the choice between single-hopping and multi-hopping under the utility criterion of measurement capacity as given by Definition 8.1. In short, we find the following:

- There is no distinct threshold transmission-to-processing ratio ρ between multi-hop and single-hop: the measurement capacity expression in (8.3) is 'well-behaved' and continuous in the $\overline{n}_m$. Hence, there is a gradual transition from the use of direct to-the-sink transmissions to the use of alternative routes via other nodes as the transmit conditions change.

- Sensing energy consumption (the energy it costs to take one measure-

ment) plays a determining role in the choice between multi-hop and single-hop. Counter-intuitively, it is when sensing requires little energy relative to the communication that single-hop is preferable, while multi-hop is favoured by energy intense sensors. Simplifying matters a little, the measurement capacity metric cares about measurements, not energy, and when a relay operation costs many measurements we should refrain from it.

- Shadowing by in-network obstacles can promote the use of multi-hop in a stronger manner than short hops. Multi-hop provides in shadowed networks alternative paths that can save transmission energy.

- Hierarchical solutions with heterogeneous nodes can increase a network's measurement capacity significantly.

8.2.1 Optimised multi-hop routing

Consider a two dimensional wireless sensor network comprising M identical sensor nodes and one central sink to which all the measurements are sent, see Figure 8.1. We assume the application is such that the measurement space coincides with the individual nodes. Sensor nodes are numbered $m = 1, 2, \ldots, M$ while the sink is referred to as the zeroth node, $m = 0$. Let

$$
\begin{aligned}
(a_m, b_m) &\equiv \text{Cartesian location of node } m, \\
(a_0, b_0) &\equiv (0,0), \text{sink is located in the Cartesian origin}, \\
d_{km} &\equiv \sqrt{(a_k - a_m)^2 + (b_k - b_m)^2} \text{ is the inter-node distance}, \\
n_m &\equiv \text{the number of measurements taken by node } m, \\
z_{km} &\equiv \text{number of measurements sent from node } k \text{ to node } m, \\
\mathcal{E}_m &\equiv \text{energy available to node } m, \\
\mathcal{E}_0 &\equiv \infty.
\end{aligned}
\tag{8.15}
$$

Next, we state the objective (O) of the optimisation, its constraints (C) and the network and node parameter assumptions (A).

O: Objective function

For a given deployment of M nodes at positions $\{(a_m, b_m)\}_{m=0}^{M}$ we now wish to find the routing pattern z_{km}, $k = 0, 1, 2, \ldots, M$ and $m = 0, 1, 2, \ldots, M$,
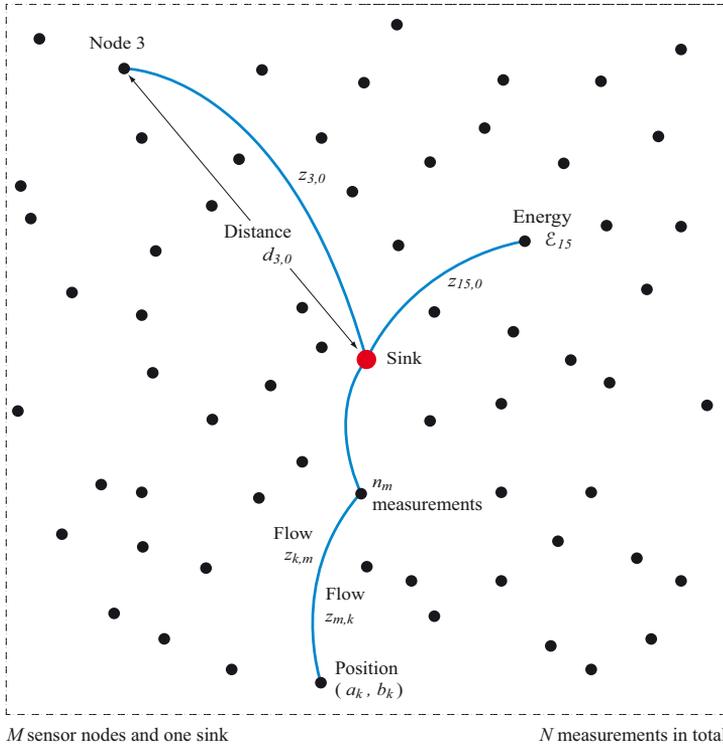
Figure 8.1: Illustration of the two-dimensional network considered for op-
timised multi-hop routing with the parameters in (8.15).

which maximises the measurement capacity $\mathcal{C}$ in (8.3),

$$\text{O:} \qquad \max_{z_{km}} N \log (N) - \sum_{m=1}^{M} n_m \log (n_m), \qquad (8.16)$$

where $N = \sum n_m$. These n_m are related to the routing pattern z_{km} through
the following constraints.

C: Constraints on routing and energy

The constraints on the optimisation are the following:

Non-negative flows. Flows of data (measurements) are non-negative,

$$\text{C1:} \qquad z_{km} \geq 0. \qquad (8.17)$$

Sources and sinks. All sensor nodes are sources and can not absorb measurements,

$$\text{C2a:} \qquad n_m = \sum_{k=1}^{M} (z_{mk} - z_{km}) \geq 0. \qquad (8.18)$$

The central sink only receives measurements and has zero outgoing flow,

$$\text{C2b:} \qquad z_{0m} \equiv 0, \qquad (8.19)$$

but absorbs all other flows,[7]

$$\text{C2c:} \qquad n_0 = \sum_{k=1}^{M} z_{m0} = N. \qquad (8.20)$$

Energy. Sensing (performing a measurement), sending and forwarding consumes energy and the activities are constrained by $\mathcal{E}_m$, the energy resource of node m. Therefore, feasible routing patterns do not violate

$$\text{C3:} \qquad \mathcal{E}_m \geq n_m \mathcal{E}_{\text{S}} + \sum_{k=1}^{M} z_{km} \mathcal{E}_{\text{Pr}} + \sum_{k=1}^{M} z_{mk} \left(\mathcal{E}_{\text{Pt}} + \mathcal{E}_{\text{T},mk} \right), \qquad (8.21)$$

where $\mathcal{E}_{\text{S}}$ is the sensing energy per measurement, $\mathcal{E}_{\text{Pr}}$ is the radio reception processing energy, $\mathcal{E}_{\text{Pt}}$ is the radio transmission processing energy and $\mathcal{E}_{\text{T},mk}$ is the node m to node k transmission energy (see Section 2.1 for definitions).

By maximising (8.16) under the constraints C1-C3 in (8.21)-(8.18) we will gain insight into how different sensing and communication costs affect the optimum routing choice with respect to the measurement capacity C in (8.3), and under what circumstances multi-hopping might provide benefits.

A: Assumptions and parameters

The optimum routing solution will depend on the energy parameters in (8.21), the characteristics of the communication channel and the sensor node positions.

---

[7]The constraint in (8.20) is actually superfluous as the sink, naturally, is the only place measurements can finally go.

Energy ratios. The energy trade-off between sensing, forwarding and direct transmission depends only on the relative sizes of $\mathcal{E}_S$, $\mathcal{E}_{Pr}$, $\mathcal{E}_{Pt}$ and $\mathcal{E}_{T,km}$. Recall therefore from (2.7) the transmission-to-processing ratio

$$\rho_{km} = \frac{\mathcal{E}_{T,km}}{\mathcal{E}_{Pt}},\tag{8.22}$$

which here is different for every node pair (k, m). All our numerical results will however be given with reference to a single transmission-to-processing ratio, without subscript, namely the largest node-to-sink ratio

$$\rho = \max(\rho_{k0}),\tag{8.23}$$

corresponding to the node having the worst single-hop transmission conditions. Also recall from (2.9) the receiver-to-transmitter processing ratio α, which here will be assumed to be one,

$$\text{A1:} \qquad \alpha = \frac{\mathcal{E}_{Pr}}{\mathcal{E}_{Pt}} = 1,\tag{8.24}$$

respectively. The additional, and important, ratio that we consider is the sensing-to-processing ratio

$$\xi \equiv \frac{\mathcal{E}_S}{\mathcal{E}_{Pt}}.\tag{8.25}$$

Energy-wise, we finally include the power amplifier degradation model (2.14) which states that savings in the radiated energy are diminished by decreased efficiency,

$$\mathcal{E}_T = \mathcal{E}_{T,\max} \left(\frac{\mathcal{E}_{rad}}{\mathcal{E}_{rad,\max}} \right)^{1/g},\tag{8.26}$$

where we will use the exponent $g = 2$ as in previous chapters.

Radio channel. For brevity, we will exclude the impact of small-scale fading and assume that the channels are static over time: the small scale fading gain of (2.22) is constant,

$$\text{A2:} \qquad x_f \equiv 1.\tag{8.27}$$

In addition, also for brevity, we restrict ourselves to the median propagation loss exponent

$$\text{A3:} \qquad \kappa = \kappa_{med} = 3.315,\tag{8.28}$$

see Section 2.4.2. The uncertainty regarding the degree of fading and the propagation loss exponent is of course still present, but to simplify our present exposition we suppress it for the moment and ask the reader to bear this in mind. We will on the other hand include shadowing effects, which in a static-channel scenario amounts to inclusion of stationary objects causing spatial variations in the channel gain. The possibility to bypass such obstacles will affect the usefulness of multi-hopping, and when we now consider two-dimensional sensor networks this aspect becomes important. In summary, referring to the channel model (2.22) on page 29 and Assumptions 2.7 and 2.8, we will use the time-invariant node k to node m channel gain

$$x_{km} = x_{\mathrm{s},km} \left(\frac{d_{km}}{d_0} \right)^{\kappa_{\mathrm{med}}}, \tag{8.29}$$

where $x_{\mathrm{s},km}$ is the log-normal shadowing and d_0 is a reference distance in the propagation loss model. All shadowing effects $x_{\mathrm{s},km}$ are assigned independent probability distributions for simplicity, but observe that this might be unrealistic because long transmissions can be assigned a larger channel gain x_{km} than a short transmission along the same line. We comment further on this when considering the numerical results.

Normalised distances. Since we only consider one propagation loss exponent we can disregard the scaling effects that actual distance has on the normalised relay-node density and the possibilities to multi-hop – analysed in Section 6.5.2. We therefore normalise the network area to the unit square and work with normalised node positions

$$\text{A4:} \qquad \begin{aligned} -1 &\le \; \breve{a}_m \; \le 1, \\ -1 &\le \; \breve{b}_m \; \le 1, \end{aligned} \tag{8.30}$$

corresponding to the absolute positions a_m and b_m in (8.15). For this reason, we need not specify d_0 in (8.29) and can simply use

$$\text{A5:} \qquad x_{km} = x_{\mathrm{s},km} \breve{d}^{\kappa_{\mathrm{med}}}, \tag{8.31}$$

to capture the effects from propagation loss and shadowing.

Continuous approximation. While the hop pattern z_{km} in the present formulation of the problem should be discrete-valued, we have used a real-valued approximation

$$\text{A6:} \qquad z_{km} \in \mathbb{R}, \tag{8.32}$$

to speed up the optimisation. Thanks to the large n_m – typically $n_m >$ 1000 in the settings we have tested – the continuous approximation is conjectured to have a minor impact on the results.

Miscellaneous. No packet aggregation or data fusion is assumed. No power control or diversity scheme is needed because of the static channel. No error correcting codes are used.

Optimisation procedure

We have used a series of local optimisations $i = 1, 2, \ldots K$ starting from small transmission-to-processing ratios $\rho(1) < 1$, which are known to yield a single-hop solution, stepping up through intermediate transmission-to-processing ratios $\rho(i)$, ending in $\rho(K)$. In each step i, the previous optimum solution $z_{km,\text{opt}}(i-1)$ has been used as the initial search point. In this way we have obtained information on the gradual increase in multi-hopping. The Mathematica search routine FindMaximum was used for the local searches, but we have also checked the first and the last results $z_{km,\text{opt}}(1)$ and $z_{km,\text{opt}}(K)$ with the global search routine NMaximize and they have always agreed. See Wolfram Research Inc. (2007) for more details about FindMaximum and NMaximize.

All optimisation results given below were obtained with

$$\rho(i) = 0.25(i-1), \qquad i = 1, 2, \ldots, 200, \tag{8.33}$$

yielding a range of transmission-to-processing ratios $\rho \in [0.25, 50]$. Remember that $\rho(i)$ always refers to the node with the worst single-hop conditions, be it due to long distance $\breve{d}_{k0}$ or severe shadowing $x_{\text{s},k0}$ in (8.31).

8.2.2 Trading off sensing and communication energies

Consider first the sensing-to-processing ratio ξ in (8.25) and let us study its impact on the optimum hop pattern z_{km}. We postpone the shadowing effects till the next section, and let

$$x_{\text{s},km} \equiv 1 \tag{8.34}$$

for now, see (8.31). In Figure 8.2 we show the optimum hopping patterns z_{km} pertaining to a 36 node randomly deployed network, $M = 36$, with initial battery energies corresponding to $10^5 \, \mathcal{E}_{\text{Pt}}$,

$$\mathcal{E}_m = 10^5 \mathcal{E}_{\text{Pt}}. \tag{8.35}$$

We study the optimum hopping patterns for transmission-to-processing ratios $\rho \in [0.25, 50]$ and sensing-to-processing ratios $\xi \in \{1, 10, 100\}$, but the results for $\xi = 1$ are not shown as multi-hopping was never observed for this sensing-to-processing ratio. The transmission-to-processing ratio ρ refers to the node farthest away from the sink, so all other nodes will have smaller transmission-to-processing ratios. Arrows are shown for all data flows that constitute more than one thousandth of the nodes total number of measurements, that is, for

$$z_{km} \geq \frac{n_k}{1000}. \tag{8.36}$$

This is to suppress all paths that are not in use, but for which the numerical optimisation still returns a non-zero value: typically these z_{km} are of the order of 10^{-7} due to numerical errors.

We observe the following:

- For a sensing-to-processing ratio $\xi = 100$, there is a steady tendency to increase the number of hops as ρ increases. For $\rho = 10$, seven nodes exclusively utilise relays (the first hop was observed for $\rho = 6.25$). When $\rho = 50$, there is basically a full multi-hop structure (only one far-away node still sends part of its data directly to the sink). These results are consistent with the energy-efficiency results in Chapter 6 where the use of multi-hop, under similar transmission conditions, turned out to be preferable when transmission-to-processing ratios exceeded $\rho \approx 7$, see Figure 6.1 and Figure 6.2.

- For a sensing-to-processing ratio $\xi = 10$, there is no or very little multi-hopping for transmission-to-processing ratios $\rho < 10$, but for $\rho = 20$ multi-hop routes start to open up. However, they are often used in conjunction with a single-hop route, showing some 'hesitation' to use multi-hop. Even more strikingly, for $\rho = 50$ the node in the lower left corner actually gives up its multi-hop route and reverts back to direct transmissions only.

- For a sensing-to-processing ratio $\xi = 1$, the single-hop approach was optimal throughout the range of transmission-to-processing ratios studied. For this reason the results are not included in Figure 8.2.

Clearly, *dominating transmission energy is not enough to motivate the use of multi-hop*. The sensing-to-processing ratio ξ has very strong influence on the choice of routing pattern z_{km}.

(a) $\xi = 100, \rho = 10$

(b) $\xi = 10, \rho = 10$

(c) $\xi = 100, \rho = 20$

(d) $\xi = 10, \rho = 20$

(e) $\xi = 100, \rho = 50$

(f) $\xi = 10, \rho = 50$
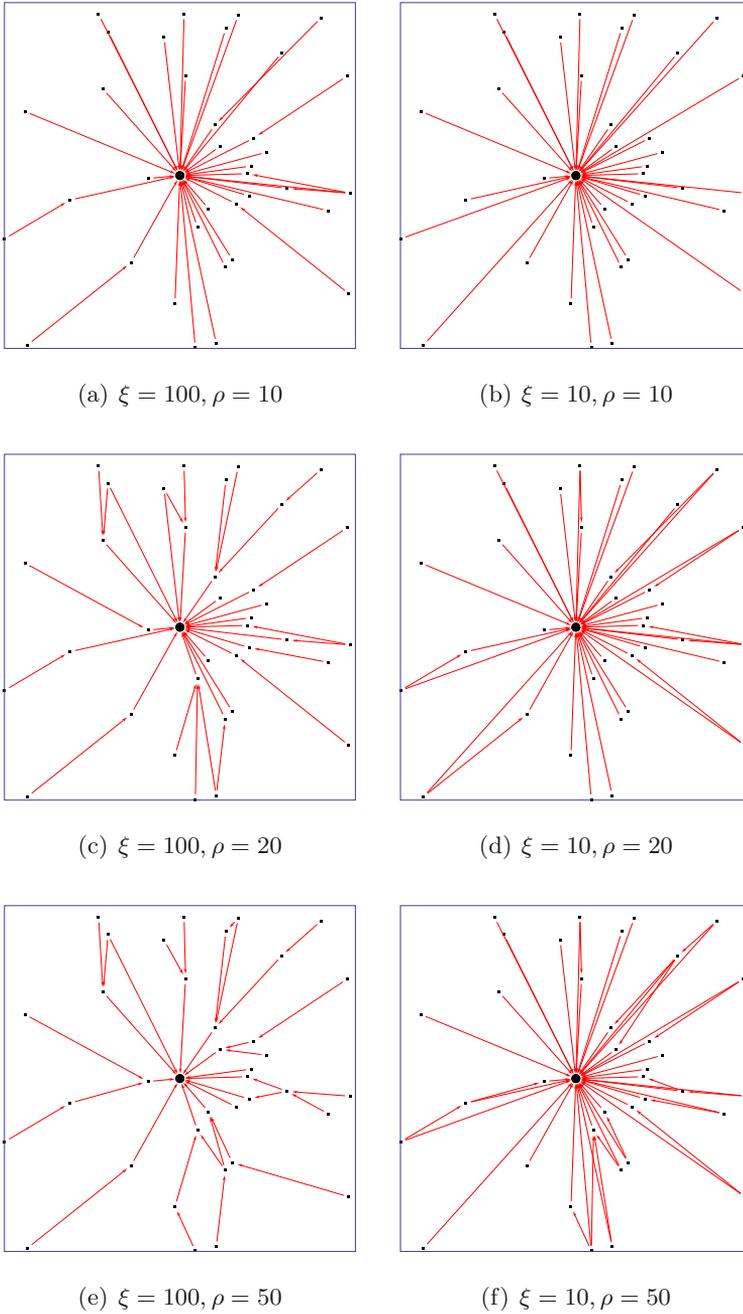
Figure 8.2: Optimised hop patterns z_{km} for a randomly deployed $M = 36$ node network. No shadowing, propagation loss exponent $\kappa = 3.315$, power amplifier degradation exponent $g = 2$.

(a) $\xi = 100$

(b) $\xi = 10$
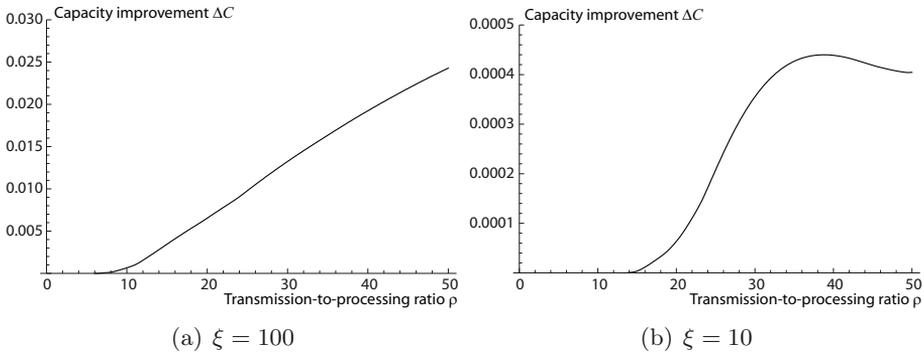
Figure 8.3: Improvement $\Delta\mathcal{C}$ in measurement capacity $\mathcal{C}$ with respect to strict single-hopping, see (8.37), achieved by optimised multi-hop routing. Observe the different scales on the horizontal axes in the two figures.

The relative improvement in measurement capacity $\mathcal{C}$ achieved by optimised routing with respect to a single-hop structure (SH) is

$$\Delta\mathcal{C} \equiv \frac{\mathcal{C}_{\mathrm{opt}} - \mathcal{C}_{\mathrm{SH}}}{\mathcal{C}_{\mathrm{SH}}}. \tag{8.37}$$

In Figure 8.3 we show the achieved improvements $\Delta\mathcal{C}$ in measurement capacity, as a function of the transmission-to-processing ratio ρ, for the two sensing-to-processing ratios $\xi \in \{10, 100\}$ The relative increase in measurement capacity for $\xi = 1$ was mostly negative and only a product of numerical errors, typically $\Delta\mathcal{C} \ll 10^{-10}$, and therfore these results are not presented. In effect, the results in Figure 8.3 confirm what we found from the hopping patterns, that the sensing-to-processing ratio ξ is highly influential. Especially for $\xi = 10$ is this evident: when the transmission-to-processing ratio ρ grows larger than ξ its multi-hop promoting effect diminishes. The reader might react to the seemingly minute increases $\Delta\mathcal{C}$, but one must understand that $\mathcal{C}$ is typically large and even a small relative improvement can be significant. For instance, $\mathcal{C} \approx 10^6$ for $\xi = 10$ and $\rho = 1$, so the difference $\mathcal{C}_{\mathrm{opt}} - \mathcal{C}_{\mathrm{SH}}$ was on the order of $10^2 - 10^3$.

The explanation to the, perhaps, counter-intuitive results[8] lies in the fact that our measurement capacity $\mathcal{C}$, loosely speaking, tells us that the number of measurements n_m are important, not energy per se. Therefore, when sensing energy $\mathcal{E}_{\mathrm{S}}$ is relatively small, a large cost in terms of measurements is incurred to any node acting as a relay. So even if a far-away node, to

---

[8]At least they were a surprise to the author, and a lot of unnecessary debugging was undertaken to find the cause of what was first believed to be an error in the calculations, but turned out to be an important aspect of the problem.

which transmission of a measurement is quite costly, can save more *energy* than the relay has to spend, the *measurement cost* is far larger than the relay is allowed to accept by the objective function. Here, the goal of energy efficiency, or rather, measurement efficiency, takes precedence of the goal of balancing.

8.2.3   Shadowing and bypassing obstacles

We have now seen that the transmission-wise energy gains achieved by short hops are not enough to motivate multi-hopping when the sensing-to-processing ratio ξ is small relative to the transmission-to-processing ratio ρ. But in many sensor network applications we can expect to have obstacles shadowing the transmission paths between sensors, and as such shadowing of the communication signals can be of significant magnitude we anticipate that the possibility to bypass obstacles can be used successfully. Let us therefore study the effects of shadowing on the optimum hopping patterns z_{km}.

In Section 2.4.3 we quantified our uncertainty regarding the log-normal shadowing effects. More specifically, under the log-normal shadowing framework of Assumption 2.7, we assigned a probability distribution (2.81) for the shadowing standard deviation σ_{dB} with median $\sigma_{\mathrm{dB,med}} = 3.88$, 10th percentile $\sigma_{\mathrm{dB,10}} = 1.40$ and 90th percentile $\sigma_{\mathrm{dB,90}} = 8.36$. By the use of these three values we can study the impact of shadow fading over a range of likely values of σ_{dB}, the parameter determining the severeness of the shadowing.

In Figure 8.4 we find optimum hop patterns for a sensing-to-processing ratio $\xi = 10$. This value of ξ was chosen because it most clearly highlights the impact of the sensing energy on the hopping patterns, see Figure 8.2. For comparability, all results were obtained using the same shadowing pattern but scaled according to the standard deviation σ_{dB}. The left column in Figure 8.4, panels 8.4(a), 8.4(c) and 8.4(e), displays the results for a transmission-to-processing ratio $\rho = 6$. The right column in Figure 8.4, panels 8.4(b), 8.4(d) and 8.4(f), displays the results for a transmission-to-processing ratio $\rho = 50$. Observe that, *due to shadowing the worst conditions for direct transmission occurs at different nodes for different σ_{dB}.* Therefore, ρ refers to different nodes: for small shadowing effects – see panels 8.4(a) and 8.4(b) – it is the distance from the sink that dominates, while large shadowing effects can dominate the impact of distance – see panels 8.4(e) and 8.4(f).

We observe the following:

- Shadowing effects lower the threshold for multi-hop because the nodes with shadowed direct paths z_{k0} to the sink can use multi-hop paths
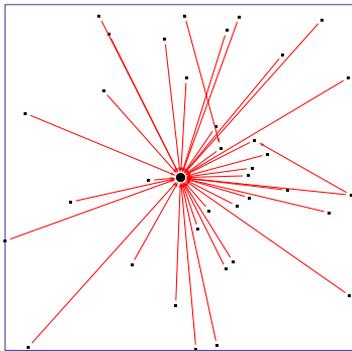
(a) $\xi = 10, \rho = 6, \sigma_{\mathrm{dB}} = 1.40$
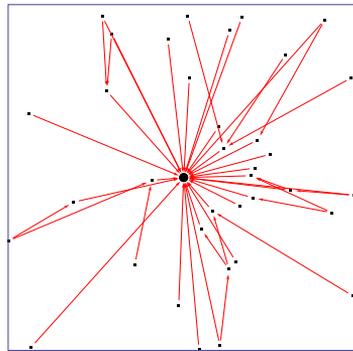
(b) $\xi = 10, \rho = 50, \sigma_{\mathrm{dB}} = 1.40$
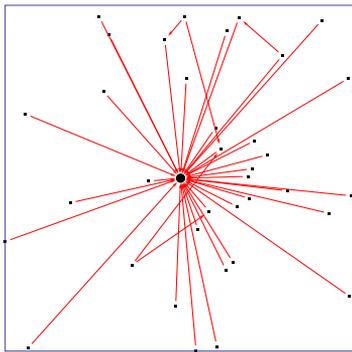
(c) $\xi = 10, \rho = 6, \sigma_{\mathrm{dB}} = 3.88$
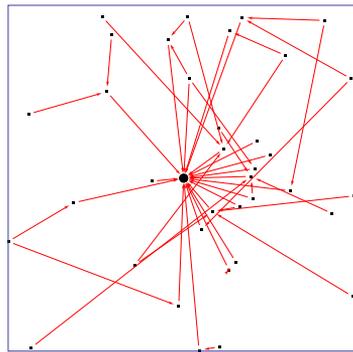
(d) $\xi = 10, \rho = 50, \sigma_{\mathrm{dB}} = 3.88$
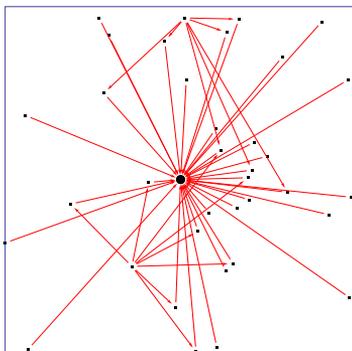
(e) $\xi = 10, \rho = 6, \sigma_{\mathrm{dB}} = 8.36$
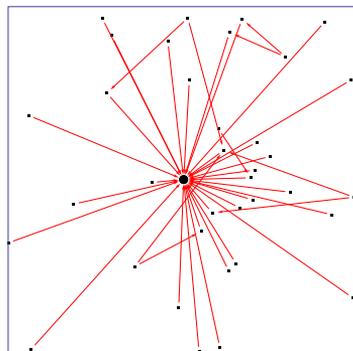
(f) $\xi = 10, \rho = 50, \sigma_{\mathrm{dB}} = 8.36$

Figure 8.4: *Optimised hop patterns* z_{km} *for a randomly deployed* $M = 36$ *node network. Propagation loss exponent* $\kappa = 3.315$, *power amplifier degradation exponent* $g = 2$.

z_{km} with significantly better channel gains $x_{km} \gg x_{k0}$, see (8.31). The transmission gains are then substantially larger than in a non-shadowed environment: it is simply the possibility of bypassing obstacles that multi-hopping can exploit. This bypassing effect becomes more pronounced the amount of shadowing, as given by σ_{dB}, increases.

• The results show the balancing effect built into the measurement capacity criterion (8.16), an effect that did not appear in the non-shadowed case. The nodes suffering from severe shadowing are helped to avoid that their 'cells' are under-covered which would decrease the measurement capacity too much. In particular, this effect is evident in panel 8.4(e) where two nodes utilise several neighbours to improve the situation.

The relative capacity increase $\Delta \mathcal{C}$, see (8.37), achieved by optimum routing is larger in shadowed than in non-shadowed environments. We observe this by comparing Figure 8.5 with Figure 8.3(b), where the non-shadowed result for sensing-to-processing ratio $\xi = 10$ is given. There is indeed a significant difference and the multi-hop scheme is apparently coming to much better use in shadowed environments, although it still requires quite large transmission-to-processing ratios ρ to show significant benefits.

8.2.4   Summary of optimised multi-hop routing

As far as the measurement capacity $\mathcal{C}$ of (8.3) is adequate for the application at hand, we have support for the following conclusions regarding multi-hop in energy-constrained wireless sensor networks.

1. The sensing-to-processing ratio ξ in (8.25) plays a decisive role in the choice between single-hop and multi-hop. It is when ξ is small compared to the communication processing costs that multi-hop performs the worst. Forwarding data is then too costly in terms of measurements, which under the measurement capacity metric is the currency of the trade, even if energy can be saved. Applications with low-power sensors are therefore, most likely, better served by single-hop structures.

2. Shadowing can make multi-hop much more attractive through the possibility to bypass, hop around, the shadowing objects. The technique then plays a "life-saving" role which is appreciated by the measurement capacity metric: to avoid severe imbalance it is beneficial to help the nodes suffering most from shadowed direct paths to the sink. Note
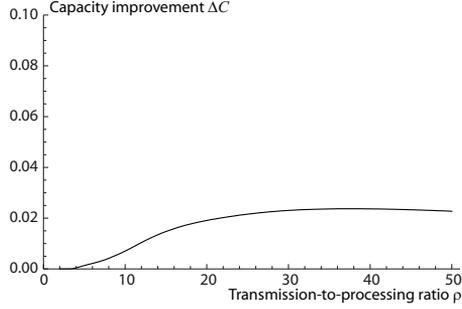
(a) $\xi = 10, \sigma_{\mathrm{dB}} = 1.40$
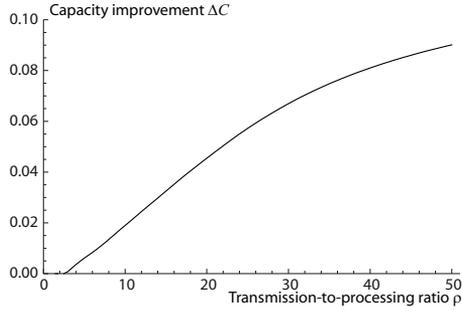
(b) $\xi = 10, \sigma_{\mathrm{dB}} = 3.88$
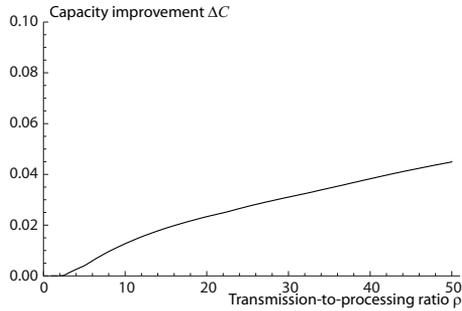
(c) $\xi = 10, \sigma_{\mathrm{dB}} = 8.36$

Figure 8.5: *Improvement* $\Delta\mathcal{C}$ *in measurement capacity* $\mathcal{C}$ *achieved with respect to strict single-hopping by optimised multi-hop routing.*

however that when the shadowing becomes too severe, like in Figure 8.5(c), the gain decreases. This is probably because it then takes too much resources to help the shadowed nodes. In any case, the measurement capacity gains achieved by multi-hop are significantly larger in shadowed than in non-shadowed environments.

8.3 Heterogeneous and hierarchical sensor networks

The larger the network becomes spatially, the less efficient it will become. In the present study, the measurement capacity $\mathcal{C}$ in (8.3) decreases consistently with increasing energy consumptions for sensing $\mathcal{E}_S$, radio processing $\mathcal{E}_{Pr}$ and $\mathcal{E}_{Pt}$, and transmission $\mathcal{E}_T$ (see (8.21)). All energy costs reduce the number of remaining measurements per node and hence the measurement capacity, and large networks will suffer from large communication costs no matter how we choose to look at it. Several authors have studied the energy imbalance problem and found, like Perillo et al. (2004), that the imbalance problem can not be alleviated in large many-to-one networks at less than severely energy inefficient operation. From the slightly different viewpoint of throughput in an any-to-any network, Gupta and Kumar (2000) found that the throughput per node (user) invariably diminishes with an increasing number of nodes.

If the problem can not be alleviated, can it be countered? Considering many-to-one data gathering networks, one commonly proposed solution is uneven energy, or node, deployment to match the energy resources to the communication load (Prasad and Agrawal, 2007, Rivas et al., 2006). Another, and we believe more promising, approach is to use a second layer of nodes which act as sinks to the sensor nodes, forming smaller subnetworks. Such hierarchical network structures employing heterogeneous node types has been studied in several network contexts, but we defer our short discussion of these till Section 8.3.2 and now briefly assess the impact that a hierarchical structure[9] may have on the network's measurement capacity $\mathcal{C}$.

8.3.1 Measurement capacity in hierarchical heterogeneous networks

Consider a network of 100 nodes spread uniformly over a square region with the sink in the middle. Let the measurement capacity it can achieve by optimised multi-hop routing be denoted $\mathcal{C}_{flat}$. Then divide, conceptually, the

---

[9]Observe that by hierarchical we mean a structure in layers, or tiers. The use of different node types in the upper and lower layers makes it a heterogeneous hierarchical network.

network into four quadrants and place a sink in the middle of each of the four sub-networks. Denote their respective sensing capacities $\mathcal{C}_k$, $k = 1, 2, 3, 4$. According to the hierarchical measurement capacity formulas in (8.4)-(8.6) we can express the hierarchical measurement capacity

$$\mathcal{C}_{\text{hier}} = \mathcal{C}'_K + \sum_{k=1}^{4} \mathcal{C}_k, \tag{8.38}$$

where $\mathcal{C}'_K$ is the measurement capacity of the four-cell super-network, see (8.5). Initial optimisation show that the improvement in measurement capacity is significant. For the above example we have found that

$$\Delta\mathcal{C} = \frac{\mathcal{C}_{\text{hier}} - \mathcal{C}_{\text{flat}}}{\mathcal{C}_{\text{flat}}}$$
$$\approx \begin{cases} 0.3 & , \quad \rho_{\text{hier}} = 5, \\ 0.6 & , \quad \rho_{\text{hier}} = 16, \end{cases} \tag{8.39}$$

where ρ_{hier} is the transmission-to-processing ratio. More simulations need to be performed to verify the validity these figures, but we feel it is safe to say that the improvement is clear (especially in comparison with the odd percent achieved by optimised multi-hop routing, see Figure 8.3). It is not surprising that the four extra support nodes improve the network's performance, but our aim is not to make a fair comparison but to illustrate the possible improvements in measurement capacity achievable by hierarchical networks.[10] Below we will argue that heterogeneous hierarchical sensor network structures are strongly supported by our results, but before we do so we note an important consequence of a hierarchical structure with single-hop sub-networks:

> In a hierarchical network structure where the low-level sensor nodes report directly to the local sink, *energy efficiency is reinstated as the proper assessment metric of the communication schemes.* Hence, our point-to-point results for energy efficient communication under processing costs have direct bearing on the design choices for local node-to-sink communication.

---

[10]Mhatre et al. (2005) has performed a study of hierarchical heterogeneous sensor networks where the monetary cost was constrained, and this led to trade-offs between having many small and cheap sensor nodes and having the support from the more powerful nods. A comparison in their spirit would be more fair, but is outside our present scope.

8.3.2 Our results support heterogeneous hierarchical networks

Let us begin with some commonly noted benefits that heterogeneous hierarchies can bring, and then look at the obvious objections and how our results relate to these matters.Gnawali et al. (2006)

- Routing is obviously simplified considerably

- Synchronisation and sleep scheduling is simplified because each sensor has only one sink to synchronise to.

- Data (measurement) processing can be more energy efficient by exploiting processing hierarchies Tsiatsis et al. (2005).

- Imbalances in communication burden is easier to counter through smart deployment because upper layer nodes are fewer in numbers.

- Communication protocols can be chosen for inter-operability in higher levels and for energy-efficiency on lower levels.

- Hierarchical structures naturally support scalability better than flat network structures do.

- Sensing can be performed at different levels of sophistication and energy cost, yielding overall benefits Tsiatsis et al. (2005).

Many of the benefits mentioned above are due to the fact that we essentially *add* more capability and resources though the local sinks – the higher level nodes – and *concentrate* much of the responsibility and activity to them. This fact may have negative consequences.

Cost. The monetary costs of the higher levels should be included in an overall analysis to judge the achievable gains from hierarchies at a given monetary cost. The question is if the expensive higher levels nodes are worth the resulting decrease in the number of low level nodes. Mhatre et al. (2005) have studied this trade-off.

Robustness. A possible drawback with heterogeneous hierarchical networks is that failure of one higher level node can cause significant damage to the network as the coverage of a large area relies on its operation. Robustness of flat networks to node failures is often put forward as a major advantage of their distributed nature.

Energy. The local sinks can generally be expected to have limited energy resources, and considering the heavy load they must carry this limitation might prove prohibitive to heterogeneous hierarchical networks.

The mentioned problems must be seriously considered by anyone thinking about a heterogeneous hierarchical network. Our general comment to these matters, before turning to the more specific results in our research, are the following:

Sensing-communication decoupling. Due to the poor scaling behaviour of communication performance in large flat networks, there will be a conflict between the sensing application and the wireless communication when networks grow large. Communication aspects can impose a deployment which from the sensing point of view is highly sub-optimal. For instance, the problem of unbalanced communication load can be countered by deploying more sensors in the most loaded areas (Prasad and Agrawal, 2007, Rivas et al., 2006), but such an uneven node density is almost surely not optimal in terms of sensing performance. The use of heterogeneous and hierarchical networks nicely decouples the low-level sensing task from the overall communication problems, and also simplifies adequate deployment of communication nodes as they are significantly less in numbers. This fact might, in conjunction with easier maintenance of higher level nodes, result in lowered costs in the long run.

Predictability. The communication performance, and hence also the sensing performance, of flat networks with a multi-hop structure is very difficult to predict. Hierarchical networks are much more predictable in this respect, and this may be utilised to improve robustness. Other aspects are given by Gnawali et al. (2006).

Energy harvesting. Higher level nodes that are decoupled from the low level sensing task can be positioned in a way that improves their energy harvesting possibilities, that is their possibility to live from ambient energy resources such as sunlight, vibrations and heat. This may lower the maintenance cost to an acceptable level.

Communication hierarchies

Considering the results we have presented in Chapters 4 through 7, we observe the following regarding heterogeneous hierarchical networks.

Transmission-to-processing ratio. As we have observed repeatedly, the energy optimal choice of transmission strategy tends to concentrate the transmission-to-processing ratio

$$\rho = \frac{\mathcal{E}_{\mathrm{T}}}{\mathcal{E}_{\mathrm{Pt}}} \tag{8.40}$$

to values in the range $\rho \in [0.5, 10]$. This concentration loosely defines, for a given, hardware-specific, transmit processing energy $\mathcal{E}_{\mathrm{Pt}}$ a maximum network size in which single-hopping will be favourable to multi-hopping. Therefore, as soon as the network becomes larger than this size, multi-hopping should be applied, but then we might as well hop to a local sink.

Match capabilities to needs. Power control, error correcting codes and receiver polarisation diversity saves energy at the transmitter while the largest energy costs has to be payed by the receiver. In a heterogeneous hierarchical network the receiver is most of the time a higher level node, saving energy for a lower level node. Lower level nodes are generally harder to reach, if it is even possible, to refill energy due to their placement, and also the node type present in the greatest numbers. Hierarchies thus open up for a better match between the needs and the capabilities at each level: low level nodes can use the simplest, most processing efficient, techniques while the local sinks possibly use more sophisticated schemes to improve the performance. Moreover, communication on higher levels is almost by definition suited for processing intensive communication schemes like multi-hopping and cooperative MIMO because of the longer transmission distances.

Multi-hop and aggregation. We saw in Chapter 6 that one exception to the rule that multi-hop requires large transmission-to-processing ratios to outperform single-hop was applications where significant data fusion effects could be exploited. For a hierarchical network this possibility remains a good source of energy saving, but this time to the higher level nodes (without incurring any energy costs at the lowest level).

We conclude from this that heterogeneous hierarchical data gathering networks facilitates a match between communication needs and energy resources. In the design of such a network the results of Chapters 4 through 7 will come to good use. Paraphrasing on the early "smart dust" description of sensor nodes, we would now prefer to promote the design principle of "dumb dust and smart network structures". Let the sensors sense, and the other nodes communicate.

Measurement capacity

The results of the present chapter, based on the notion of measurement capacity, lead us to the following remarks pertaining to heterogeneous hierarchical networks:

Increased measurement capacity. Our initial results show a significant increase in measurement capacity by the introduction of local sinks. This is achieved by a combination of reduced and more uniform transmission costs throughout the low-level network. And, it is the low-level nodes that do the sensing.

Scalable hierarchical scheduling. The measurement capacity $\mathcal{C}$ in (8.3) is a combined resource pertaining to the network as a whole. This resource metric is therefore suitable in optimal scheduling of measurements, because we can determine the resource cost of performing certain tasks. Now, considering the hierarchical property (8.4) of the metric, it appears suitable for use in hierarchical scheduling. Higher levels schedule according to the measurement capacity $\mathcal{C}_k$ and the number of measurements N_k of the cells below them, and the local sink schedules according the number of measurements n_{m_k} that each node has left.

Sink diversity and shadowing. Shadowing will affect the measurement capacity negatively by reducing the number of measurements n_m that a shadowed node can report. We saw in Figures 8.4 and 8.5 how multihop under these circumstances facilitated a sizeable increase in the measurement capacity by opening alternative paths around obstacles. In a hierarchical network, the local sinks will, if deployed at suitable positions, provide path diversity for the sensor nodes. Especially the sensor nodes located midway between local sinks will benefit from such path diversity.

8.4   Summary

The concept of measurement capacity given in Definition 8.1 offers, to the best of our knowledge, an entirely new way of assessing network performance. It is suited for scenarios in which the network can be used in many different, unpredictable ways. Be it dictated by the sensing application, event detection for instance, or by the user(s) having different needs at different times. By measuring the 'volume' of the space of possible scenarios to which the network can respond satisfactorily, the measurement capacity $\mathcal{C}$ constitutes

a network-wide metric of how well-armed the network is to meet upcoming events/requests.

The key to successful use of the measurement capacity in sensor network design is the definition of the 'cells', the present counterpart to the hypothesis space of an inference problem, since a bad choice of cells can never be compensated for. We are unsure as to how large the range of suitable applications is, and this uncertainty stems solely from the problem with cell definitions. Once, and if, a good choice of cells have been made we are confident that the use of the measurement capacity will yield sensible results. In its relative simplicity, the foundation of the concept of measurement capacity in terms of multiplicities to us appears as its major strength.

Two generalisations should be investigated. First, there will in reality be uncertainty regarding the number of measurements n_m per cell, and this uncertainty should be taken into account. Probability theory as logic is well suited for this. Second, in Appendix 8.A we outline a generalisation towards multi-cell measurements.

By using the measurement capacity metric to assess multi-hop routing and the impact of sensing energy consumption, we reached the counter-intuitive conclusion that small sensing costs tend to favour the use of single-hop structures, because the forwarding cost in terms of measurements then reduces the network's measurement capacity. The main benefit from multi-hop appears to be the possibility to jump around obstacles in shadowed environments. Under shadowing, the capacity gains from optimised multi-hop routing over single-hop was significantly larger than under non-shadowed conditions.

We also found that a substantial increase in measurement capacity can be achieved by a second layer of nodes acting as local sinks. This was not surprising, but together with the previous results in this thesis, and a wealth of results in the literature, we think that there is a strong support for heterogeneous hierarchical sensor network structures.

In the immediate future, we deem an application of measurement capacity to optimal search problems in hierarchical networks an interesting research topic. This is because measurement capacity is the network-wide resource which we can find ways to distribute optimally in terms of searches performed by sensors. Without a central resource metric, the allocation of search effort is harder to address, and to the best of our knowledge no other adequate resource metric has been proposed.

Appendix 8.A   Possible generalisations

Similar to Rachlin et al. (2005), we also consider an underlying space of cells on which our capacity metrics are defined. But, they consider the possibility of simultaneous events in different cells, while we consider sequences of single-cell measurements. Let us point to two generalisations of our measurement capacity $\mathcal{C}$ in (8.3) toward the model used by Rachlin et al. (2005).

Alternative 1. We still consider a network which can perform n_m measurements in each cell $m = 1, 2, \ldots, M$, and a total number of measurements $N = \sum n_m$. At each instant of interest, we could according to Definition 8.1 of the measurement capacity $\mathcal{C}$ only make one measurement, but let us now consider all possible distributions of the measurements over N instants (we can think of N cells in time). No more than one measurement per node and time instant is possible. In each cell, there is then

$$W_m = \frac{N!}{(N - n_m)! n_m!} \tag{8.41}$$

different ways to perform measurements. Together, the network can respond to

$$W = \prod_{m=1}^{M} \frac{N!}{(N - n_m)! n_m!} \tag{8.42}$$

different measurement distribution. Applying Stirlings approximation (2.38) we can after some manipulations find that

$$\log(W) \approx N \left[-\sum_{m=1}^{M} \overline{n}_m \log(\overline{n}_m) - \sum_{m=1}^{M} (1 - \overline{n}_m) \log(1 - \overline{n}_m) \right], \tag{8.43}$$

where $\overline{n}_m = n_m / N$ as before. In addition to the measurement capacity $\mathcal{C}$, which is the first term, we get an additional "entropy" term defined on $1 - \overline{n}_m$. The metric in (8.43) incorporates the possibility to measure simultaneously in many cells. However, it also comprises measurement instants when *no* measurement is taken by any node.

Alternative 2. If we would only like to consider instants when one or more nodes perform measurements, that is to say we consider an event-driven rather than a time-driven scenario, we must incorporate the constraint that at least one measurement per instant is performed. Unfortunately, we can not find a tractable expression for this case, but we believe that it would adequate for many sensing scenarios.

Chapter 9

Concluding Remarks

THE results in Chapters 3 through 8 provide relevant guidelines for the design of wireless sensor network.

Radio design. We have repeatedly seen that the optimal choices of transmission strategy tend to equalise the transmission processing energies, so that the transmission-to-total-processing ratio ρ' is neither excessively small, nor large. We have also found that cooperative techniques, that is multi-hopping and cooperative MIMO, become attractive for transmission-to-total-processing ratios $\rho' > 5$, approximately. Hence, *radios should be designed with a maximum transmission-to-total-processing ratio $\rho'_{\max} > 5$ to avoid having to resort to cooperative techniques when they are energy-inefficient.*

Transmission technique ordering. The most basic transmission is an uncoded, fixed power, single-antenna, node-to-node single-hop transmission. This choice is suitable for transmission-to-total-processing ratios $\rho' < 0.1$. Increasing the transmission-to-total-processing ratio ρ', corresponding in a broad sense to increasing the transmission distance, our results show that transmit power control, polarisation receiver diversity and error correcting codes become attractive for $\rho' \in [0.1, 0.5]$. Their energy gains are fairly consistent for transmission-to-total-processing ratios $\rho' > 0.5$, and substantial for $\rho' > 1$. Increasing distance further, the three mentioned techniques can be used to balance transmission and processing costs until the transmission-to-total-processing ratio $\rho' \approx 5$. Then the energy efficient, non-cooperative, range of transmission-to-processing ratios end, and multi-hop and cooperative MIMO become attractive alternatives for $\rho' > 5$. Of course, the coopera-

tive techniques are used in combination with the three non-cooperative techniques. Observe that the use of transmit power control, error correcting codes and receiver diversity extends the *distance range* over which non-cooperative techniques are preferable, but keep the transmission-to-total-processing ratio ρ' limited. To do this efficiently the non-cooperative transmission techniques must be adaptive, but the adaptivity can probably be coarse grained with small amounts of feedback.

The impact of uncertainties. Different techniques are preferable under different circumstances, and we are generally uncertain regarding the exact conditions a sensor network will encounter. Hence, we can not find sharp transmission-to-processing boundaries at which the considered techniques ought to be applied. However, for transmission-to-total-processing ratios $\rho' < 5$ we can with confidence conclude that the non-cooperative techniques are more energy-efficient than the the cooperative techniques – the exception being applications for which significant data fusion and/or packet aggregation is possible through the cooperation.

Hierarchical (tiered) networks. Our use of the network measurement capacity – the number of possible measurement sequences a network has capability to perform – revealed that the measurement-to-radio-processing ratio has strong influence on the suitability of multi-hopping. When the measurement energy is small with respect to the radio processing energy, multi-hop cooperation costs too many measurements and should be refrained from. We conjecture that this is true also for cooperative MIMO. Only in shadowed environments did multi-hopping obtain a sizeable gain in measurement capacity by circumventing obstacles and thereby reduce energy consumption. However, the increase in measurement capacity achieved by the use of a second layer of nodes, a second tier, was an order of magnitude larger than achieved by optimised multi-hopping. A heterogeneous hierarchical network obtains the large increase by increasing the number of measurements *and* by equalising the load in the lower layer. Further, the combined use of heterogeneous hierarchical networks and the measurement capacity as a resource metric will facilitate efficient hierarchical scheduling, exploiting the hierarchic properties of the measurement capacity.

Summary. In Figure 9.1 we provide an overview of our results and the design guidelines which, we conclude, emerge from the results. First, we note again the need for radios to have enough transmission power to cover
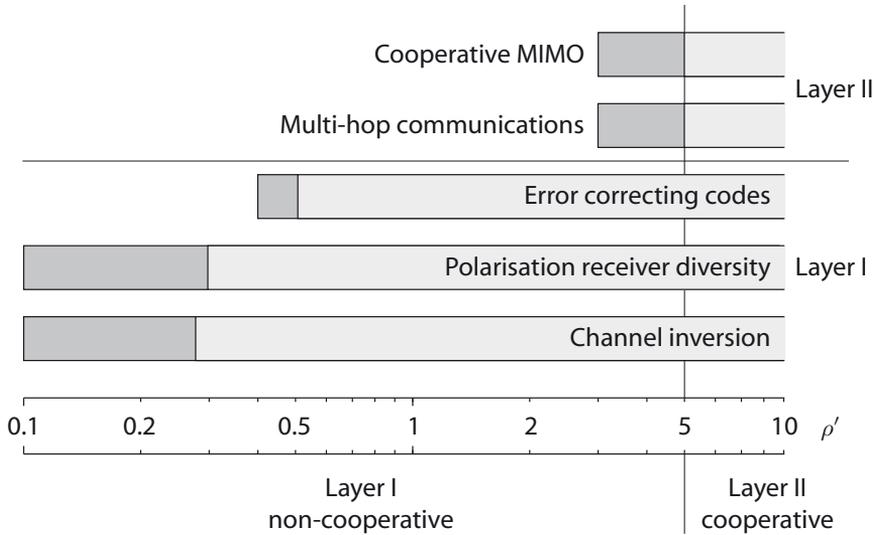
Figure 9.1: Energy efficient ranges of transmission-to-total-processing ratios
ρ' *for different transmission techniques. Light grey areas show ranges of*
certain savings, dark grey indicate uncertain savings. Layer I and Layer II
refers to hierarchical levels, indicating suitable transmission-techniques for
low-level and high-level communication.

the transmission-to-total-processing range up till $\rho' = 5$, where cooperative
techniques become useful. Otherwise, there will be an "inefficiency gap"
between the radio capability and efficient use of cooperative techniques. This
gap is largely left open by existing radios. Second, we observe that our results
also give a natural "size" of the sub-networks in a heterogeneous hierarchical
approach, defined by the range over which non-cooperative transmission is
most energy efficient. Larger communication distances are taken care of
by higher layers which possibly can make efficient use of the cooperative
techniques. Voilà, we have a decoupling of the low-level sensing task from
the cooperative communication in a structure that still facilitates the use
of data fusion and packet aggregation. Summarising, our overall design
principle for energy constrained wireless sensor networks is well described
in the following way.

Dumb dust but smart networks

Bibliography

A. Abdi and M. Kaveh. On the utility of gamma PDF in modeling shadow fading (slow fading). In *Vehicular Technology Conference, 1999 IEEE 49th*, volume 3, pages 2308–2312, Houston, TX, USA, July 1999. doi: 10.1109/VETEC.1999.778479.

A. F. Abouraddy and S. M. Elnoubi. Statistical modeling of the indoor radio channel at 10 GHz throughpropagation measurements .i. narrowband measurements and modeling. *IEEE Transactions on Vehicular Technology*, 49(5):1491–1507, September 2000. ISSN 0018-9545. doi: 10.1109/25.892532.

A. A. Abu-Dayya and N. C. Beaulieu. Analysis of switched diversity systems on generalized-fading channels. *IEEE Transactions on Communications*, 42(11):2959–2966, November 1994. ISSN 0090-6778. doi: 10.1109/26.328977.

Aerocomm. Datasheet ZigBee, 2008. http://www.aerocomm.com/docs.

S. Aeron, M. Zhao, and V. Saligrama. On sensing capacity of sensor networks for a class of linear observation models. In *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, pages 388–392, August 2007. doi: 10.1109/SSP.2007.4301286.

I. F. Akyildiz, Weilian Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102–114, August 2002. ISSN 0163-6804. doi: 10.1109/MCOM.2002.1024422.

P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik. Survey of channel and radio propagation models for

wireless mimo systems. *EURASIP J. Wirel. Commun. Netw.*, 2007(1):56–56, 2007. ISSN 1687-1472. doi: http://dx.doi.org/10.1155/2007/19070.

J. B. Anderson and A. Svensson. *Coded Modulation Systems*. Kluwer Academic Publishers, Norwell, MA, USA, 2002. ISBN 0306472791.

D. Aronsson. Channel estimation and prediction from a bayesian perspective, 2007. http://www.signal.uu.se/Publications/pdf/l071.pdf.

W. Bajwa, A. Sayeed, and R. Nowak. Matched source-channel communication for field estimation in wireless sensor networks. In *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 44, Piscataway, NJ, USA, 2005. IEEE Press. ISBN 0-7803-9202-7.

S. Bandyopadhyay and E. J. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, volume 3, pages 1713–1723, March/April 2003.

D. Beauvarlet and K. L. Virga. Measured characteristics of 30-GHz indoor propagation channels with low-profile directional antennas. *IEEE Antennas and Wireless Propagation Letters*, 1:87–90, 2002. ISSN 1536-1225. doi: 10.1109/LAWP.2002.802553.

E. Björnemo, M. Johansson, and A. Ahlén. Two hops is one to many in an energy-limited wireless sensor network. In *International Conference on Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, April 2006.

E. Björnemo, A. Ahlén, and M. Johansson. On the energy-efficiency of cooperative MIMO in Nakagami-fading wireless sensor networks. In *41st Asilomar Conference on Signals, Systems, and Computers*, Asilomar, USA, November 2007.

A. Bohdanowicz, G. J. M. Janssen, and S. Pietrzyk. Wideband indoor and outdoor multipath channel measurements at 17GHz. In *Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th*, volume 4, pages 1998–2003, Amsterdam, Netherlands, 1999. doi: 10.1109/VETECF.1999.797287.

P. K. Bondyopadhyay. The first application of array antenna. In *Phased Array Systems and Technology, 2000. Proceedings. 2000 IEEE Interna-*

tional Conference on, pages 29–32, Dana Point, CA, USA, 2000. doi: 10.1109/PAST.2000.858903.

R. J. C. Bultitude, S. A. Mahmoud, and W. A. Sullivan. A comparison of indoor radio propagation characteristics at 910 MHzand 1.75 GHz. *IEEE Journal on Selected Areas in Communications*, 7(1):20–30, January 1989. ISSN 0733-8716. doi: 10.1109/49.16840.

J. Chang and L. Tassiulas. Routing for maximum system lifetime in wireless ad-hoc networks. In *in 37-th Annual Allerton Conference on Communication, Control, and Computing*, 1999.

Y. H. Chee. *Ultra Low Power Transmitters for Wireless Sensor Networks*. PhD thesis, University of California, Berkeley, Berkeley, USA, 2006. http://bwrc.eecs.berkeley.edu/Publications/2006/index.htm.

J. Cheng and N. C. Beaulieu. Maximum-likelihood based estimation of the nakagami m parameter. *IEEE Communications Letters*, 5(3):101–103, March 2001. ISSN 1089-7798. doi: 10.1109/4234.913153.

Z. Cheng, M. Perillo, and W. B. Heinzelman. General network lifetime and cost models for evaluating sensor network deployment strategies. *IEEE Transactions on Mobile Computing*, 7(4):484–497, April 2008. ISSN 1536-1233. doi: 10.1109/TMC.2007.70784.

R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the LambertW function. *Advances in Computational Mathematics*, 5 (1):329–359, Dec 1996. ISSN 1019-7168. doi: 10.1007/BF02124750.

A. Del Coso, U. Spagnolini, and C. Ibars. Cooperative distributed MIMO channels in wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 25(2):402–414, February 2007. ISSN 0733-8716. doi: 10.1109/JSAC.2007.070215.

E. Costa and S. Pupolin. M-QAM-OFDM system performance in the presence of a nonlinearamplifier and phase noise. *IEEE Transactions on Communications*, 50(3):462–472, March 2002. ISSN 0090-6778. doi: 10.1109/26.990908.

R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946. doi: 10.1119/1.1990764. http://link.aip.org/link/?AJP/14/1/1.

S. Cui, A.J. Goldsmith, and A. Bahai. Energy-efficiency of MIMO and coop-
erative MIMO techniques in sensor networks. *IEEE Journal on Selected
Areas in Communications*, 22(6):1089–1098, August 2004. ISSN 0733-
8716. doi: 10.1109/JSAC.2004.830916.

S. Cui, A.J. Goldsmith, and A. Bahai. Energy-constrained modulation op-
timization. *IEEE Transactions on Wireless Communications*, 4(5):2349–
2360, September 2005. ISSN 1536-1276. doi: 10.1109/TWC.2005.853882.

M. Di Renzo, F. Graziosi, R. Minutolo, M.Montanari, and F. Santuci. The
ultra-wide bandwidth outdoor channel: From measurement campaign to
statistical modelling. *Journal Mobile Networks and Applications*.

I. Dietrich and F. Dressler. On the lifetime of wireless sen-
sor networks. Internal report, Department of computer sci-
ence, University of Erlangen, 2006. http://www7.informatik.uni-
erlangen.de/ dressler/publications/pdf/dietrich2006lifetime.pdf.

A. Dunkels and J. P. Vasseur. http://www.ipso-
alliance.org/Pages/DocumentsAndWhitePapers.aspx, 2008.

G. Durgin, T. S. Rappaport, and H. Xu. Measurements and models for ra-
dio path loss and penetration loss inand around homes and trees at 5.85
GHz. *IEEE Transactions on Communications*, 46(11):1484–1496, Novem-
ber 1998. ISSN 0090-6778. doi: 10.1109/26.729393.

D. Estrin, L. Girod, G. Pottie, and M. Srivastava. Instrumenting the world
with wireless sensor networks. In *Acoustics, Speech, and Signal Processing,
2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference
on*, volume 4, pages 2033–2036, Salt Lake City, UT, USA, 2001. doi:
10.1109/ICASSP.2001.940390.

N. Filiol, N. Birkett, J. Cherry, F. Balteanu, C. Gojocaru, A. Namdar,
T. Pamir, K. Sheikh, G. Glandon, D. Payer, A. Swaminathan, R. Forbes,
T. Riley, S. M. Alinoor, E. Macrobbie, M. Cloutier, S. Pipilos, and T. Vare-
las. A 22 mW Bluetooth RF transceiver with direct RF modulation and
on-chip IF filtering. In *Solid-State Circuits Conference, 2001. Digest of
Technical Papers. ISSCC. 2001 IEEE International*, pages 202–203, San
Francisco, CA, USA, 2001. doi: 10.1109/ISSCC.2001.912604.

M. Gastpar, M. Vetterli, and P. L. Dragotti. Sensing reality and communi-
cating bits: a dangerous liaison. *IEEE Signal Processing Magazine*, 23(4):
70–83, July 2006. ISSN 1053-5888. doi: 10.1109/MSP.2006.1657818.

S. S. Ghassemzadeh, R. Jana, C. W. Rice, W. Turin, and V. Tarokh. Measurement and modeling of an ultra-wide bandwidth indoor channel. *IEEE Transactions on Communications*, 52(10):1786–1796, October 2004. ISSN 0090-6778. doi: 10.1109/TCOMM.2003.820755.

O. Gnawali, K.-Y. Jang, J. Paek, M. Vieira, R. Govindan, B. Greenstein, A. Joki, D. Estrin, and E. Kohler. The tenet architecture for tiered sensor networks. In *SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 153–166, New York, NY, USA, 2006. ACM. ISBN 1-59593-343-3. doi: http://doi.acm.org/10.1145/1182807.1182823.

A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, USA, 2005. ISBN 0521837162.

A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, November 1997. ISSN 0018-9448. doi: 10.1109/18.641562.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Academic Press, San Diego, USA, 2000. ISBN 0122947576.

L. Gu and J. A. Stankovic. Radio-triggered wake-up capability for sensor networks. In *Real-Time and Embedded Technology and Applications Symposium, 2004. Proceedings. RTAS 2004. 10th IEEE*, pages 27–36, May 2004. doi: 10.1109/RTTAS.2004.1317246.

P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000. ISSN 0018-9448. doi: 10.1109/18.825799.

M. Haenggi. The impact of power amplifier characteristics on routing in random wireless networks. In *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, volume 1, pages 513–517, December 2003. doi: 10.1109/GLOCOM.2003.1258290.

H. Hashemi. The indoor radio propagation channel. *Proceedings of the IEEE*, 81(7):943–968, July 1993. ISSN 0018-9219. doi: 10.1109/5.231342.

H. Hashemi, M. McGuire, T. Vlasschaert, and D. Tholl. Measurements and modeling of temporal variations of the indoorradio propagation channel. *IEEE Transactions on Vehicular Technology*, 43:733–737, August 1994. ISSN 0018-9545. doi: 10.1109/25.312774.

A. Healey, C. H. Bianchi, and K. Sivaprasad. Wideband outdoor channel sounding at 2.4 GHz. In *Antennas and Propagation for Wireless Communications, 2000 IEEE-APS Conference on*, pages 95–98, Waltham, MA, USA, 2000. doi: 10.1109/APWC.2000.900150.

W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, January 2000.

S. L. Howard, C. Schlegel, and K. Iniewski. Error control coding in low-power wireless sensor networks: when is ECC energy-efficient? *EURASIP J. Wirel. Commun. Netw.*, 2006(2):29–29, 2006. ISSN 1687-1472. doi: http://dx.doi.org/10.1155/WCN/2006/74812.

M. Ilyas and I. Mahgoub. *Handbook of Sensor Networks – Compact Wirless and Wired Sensing Systems*. CRC Press, Boca Raton, Florida, USA, 2005.

S. Jagannathan, H. Aghajan, and A. Goldsmith. The effect of time synchronization errors on the performance of cooperative MISO systems. In *Global Telecommunications Conference Workshops, 2004. GlobeCom Workshops 2004. IEEE*, pages 102–107, November/December 2004. doi: 10.1109/GLOCOMW.2004.1417557.

G. J. M. Janssen, P. A. Stigter, and R. Prasad. Wideband indoor channel measurements and BER analysis of frequencyselective multipath channels at 2.4, 4.75, and 11.5 GHz. *IEEE Transactions on Communications*, 44 (10):1272–1288, October 1996. ISSN 0090-6778. doi: 10.1109/26.539768.

S. K. Jayaweera. Virtual MIMO-based cooperative communication for energy-constrained wireless sensor networks. *IEEE Transactions on Wireless Communications*, 5(5):984–989, May 2006. ISSN 1536-1276. doi: 10.1109/TWC.2006.1633350.

E. T. Jaynes. Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630, May 1957a.

E. T. Jaynes. Information theory and statistical mechanics II. *The Physical Review*, 108(2):171–190, October 1957b.

E. T. Jaynes. On the rationale of of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, September 1982.

E. T. Jaynes. Entropy and search theory. In *Maximum Entropy and Bayesian Methods in Inverse Problems*, page 443, Dordrecht, Reidel, 1985. Presented June 1981 at the First Maximum Entropy Workshop at the University of Wyoming.

E. T. Jaynes. *Probability Theory – The Logic of Science*. Cambridge University Press, March 2003.

M. Johansson, E. Björnemo, and A. Ahlén. Fixed link margins outperform power control in energy-limited wireless sensor networks. In *International Conference on Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, April 2006.

S. Kullback. *Information Theory and Statistics*. Dover Publications, second edition, 1968.

J. N. Laneman. *Cooperative diversity in wireless networks: algorithms and architectures*. PhD thesis, 2002. Supervisor-Gregory W. Wornell.

J. N. Laneman and G. W. Wornell. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Transactions on Information Theory*, 49(10):2415–2425, October 2003. ISSN 0018-9448. doi: 10.1109/TIT.2003.817829.

E. G. Larsson and P. Stoica. *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, New York, NY, USA, 2003. ISBN 0521824567.

A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson. Wireless sensor networks for habitat monitoring. In *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97, New York, NY, USA, 2002. ACM. ISBN 1-58113-589-0. doi: http://doi.acm.org/10.1145/570738.570751.

W. Q. Malik and D. J. Edwards. UWB impulse radio with triple-polarization SIMO. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 4124–4128, November 2007. doi: 10.1109/GLOCOM.2007.784.

V. P. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff. A minimum cost heterogeneous sensor network with a lifetime constraint. *IEEE Transactions on Mobile Computing*, 4(1):4–15, January/February 2005. ISSN 1536-1233. doi: 10.1109/TMC.2005.2(410) 4.

S. Mikami, T. Takeuchi, H. Kawaguchi, C. Ohta, and M. Yoshimoto. An efficiency degradation model of power amplifier and the impact against transmission power control for wireless sensor networks. In *Radio and Wireless Symposium, 2007 IEEE*, pages 447–450, January 2007. doi: 10.1109/RWS.2007.351864.

R. Min and A. Chandrakasan. Mobicom poster: top five myths about the energy consumption of wireless communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 7(1):65–67, 2003. ISSN 1559-1662. doi: http://doi.acm.org/10.1145/881978.881998.

R. Min, M. Bhardwaj, S.-H. Cho, N. Ickes, E. Shih, A. Sinha, A. Wang, and A. Chandrakasan. Energy-centric enabling technologies for wireless sensor networks. 9(4):28–39, aug 2002. ISSN 1536-1284.

M. Nakagami. The m-distribution – a general formula of intensity distribution of rapid fading. *Statistical Methods of Radio Wave Propagation*, 1960.

S. Olariu and I. Stojmenovic. Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–12, April 2006. doi: 10.1109/INFOCOM.2006.296.

J. D. Parsons. *The Mobile Radio Propagation Channel*. John Wiley and Sons, Chichester, England, 2001. ISBN 047198857.

M. Perillo, Z. Cheng, and W. Heinzelman. On the problem of unbalanced load distribution in wireless sensor networks. In *Global Telecommunications Conference Workshops, 2004. GlobeCom Workshops 2004. IEEE*, pages 74–79, November/December 2004. doi: 10.1109/GLOCOMW.2004.1417552.

A. S. Y. Poon and M. Ho. Indoor multiple-antenna channel characterization from 2 to 8 GHz. In *Communications, 2003. ICC '03. IEEE International Conference on*, volume 5, pages 3519–3523, May 2003. doi: 10.1109/ICC.2003.1204108.

G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Commun. ACM*, 43(5):51–58, 2000. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/332833.332838.

P. Prasad and P. Agrawal. Energy efficient spatial distribution of nodes in a sensor network. In *Sarnoff Symposium, 2007 IEEE*, pages 1–5, Nassau Inn, Princeton, NJ,, April/May 2007. doi: 10.1109/SARNOF.2007.4567360.

J. G. Proakis. *Digital Communications*. McGraw Hill, New York, USA, 2001. ISBN 0072321113.

Y. Rachlin, R. Negi, and P. Khosla. Sensing capacity for discrete sensor network applications. In *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pages 126–132, April 2005. doi: 10.1109/IPSN.2005.1440911.

V. Raghunathan, C. Schurgers, Sung Park, and M. B. Srivastava. Energy-aware wireless microsensor networks. *IEEE Signal Processing Magazine*, 19(2):40–50, March 2002. ISSN 1053-5888. doi: 10.1109/79.985679.

R. Rajagopalan and P. K. Varshney. Data-aggregation techniques in sensor networks: a survey. *IEEE Communications Surveys & Tutorials*, 8(4): 48–63, / 2006. ISSN 1553-877X. doi: 10.1109/COMST.2006.283821.

T. S. Rappaport. Indoor radio communications for factories of the future. *IEEE Communications Magazine*, 27(5):15–24, May 1989. ISSN 0163-6804. doi: 10.1109/35.29535.

M. Rice and S. B. Wicker. Adaptive error control for slowly varying channels. *IEEE Transactions on Communications*, 42:917–926, 1994. ISSN 0090-6778. doi: 10.1109/TCOMM.1994.580200.

H. Rivas, T. Voigt, and A. Dunkels. A simple and efficient method to mitigate the hot spot problem in wireless sensor networks. In *Proceedings of Performance Control in Wireless Sensor Networks*, Coimbra, Portugal, 2006. http://eprints.sics.se/306/01/perf2006.pdf.

K. Romer and F. Mattern. The design space of wireless sensor networks. *IEEE Wireless Communications*, 11(6):54–61, December 2004. ISSN 1536-1284. doi: 10.1109/MWC.2004.1368897.

L. Rubio, J. Reig, and N. Cardona. Evaluation of Nakagami fading behaviour based on measurements in urban scenarios. *AEU-International Journal of Electronics and Communications*, 61(2):135–138, 2007.

A. Rustako, Y. Yeh, and R. Murray. Performance of feedback and switch space diversity 900 MHz FM mobile radio systems with rayleigh fading.

Communications, IEEE Transactions on [legacy, pre - 1988], 21(11):1257–
1268, November 1973. ISSN 0096-2244.

A. Saleh and R. Valenzuela. A statistical model for indoor multipath propagation. *IEEE Journal on Selected Areas in Communications*, 5(2):128–137,
February 1987. ISSN 0733-8716.

J. Salo, L. Vuokko, H. M. El-Sallabi, and P. Vainikainen. An additive
model as a physical basis for shadow fading. *IEEE Transactions on Vehicular Technology*, 56(1):13–26, January 2007. ISSN 0018-9545. doi:
10.1109/TVT.2006.883797.

M. Sartipi and F. Fekri. Source and channel coding in wireless sensor
networks using LDPC codes. In *Sensor and Ad Hoc Communications
and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, pages 309–316, October 2004. doi:
10.1109/SAHCN.2004.1381931.

S. Y. Seidel and T. S. Rappaport. 914 MHz path loss prediction models for
indoor wirelesscommunications in multifloored buildings. *IEEE Transactions on Antennas and Propagation*, 40(2):207–217, February 1992. ISSN
0018-926X. doi: 10.1109/8.127405.

C. E. Shannon. A mathematical theory of communication. *Bell System
Technical Journal*, 27:379–423, jul 1948a.

C. E. Shannon. A mathematical theory of communication. *Bell System
Technical Journal*, 27:623–656, oct 1948b.

A.U. Sheikh, M. Abdi, and M. Handforth. Indoor mobile radio channel at
946 MHz: Measurements and modeling. In *Vehicular Technology Conference, 1993 IEEE 43rd*, pages 73–76, Secaucus, NJ, USA, May 1993. doi:
10.1109/VETEC.1993.507014.

B. Shen and Q.-L. Zhang. A new method for analyzing the quantization effect
of ADC in broadband QAM receiver. In *Communications, Circuits and
Systems and West Sino Expositions, IEEE 2002 International Conference
on*, volume 2, pages 1262–1266, June/July 2002.

E. Shih, S.-H. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan. Physical layer driven protocol and algorithm design for energy-
efficient wireless sensor networks. In *MobiCom '01: Proceedings of the
7th annual international conference on Mobile computing and networking,*

pages 272–287, New York, NY, USA, 2001. ACM Press. ISBN 1-58113-422-3. doi: http://doi.acm.org/10.1145/381677.381703.

E. Shih, S. Cho, F. S. Lee, B. H. Calhoun, and A. Chandrakasan. Design considerations for energy-efficient radios in wireless microsensor networks. *J. VLSI Signal Process. Syst.*, 37(1):77–94, 2004. ISSN 0922-5773. doi: http://dx.doi.org/10.1023/B:VLSI.0000017004.57230.91.

M. K. Simon and M. S. Alouini. A unified performance analysis of digital communication with dualselective combining diversity over correlated rayleigh and nakagami-mfading channels. *IEEE Transactions on Communications*, 47(1):33–43, January 1999. ISSN 0090-6778. doi: 10.1109/26.747811.

D. S. Sivia. *Data Analysis – A Bayesian Tutorial*. Clarendon Press, 1996.

K. Sohrabi, B. Manriquez, and G. J. Pottie. Near ground wideband channel measurement in 800-1000 MHz. In *Vehicular Technology Conference, 1999 IEEE 49th*, volume 1, pages 571–574, Houston, TX, USA, July 1999. doi: 10.1109/VETEC.1999.778120.

K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie. Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications*, 7(5):16–27, October 2000. ISSN 1070-9916. doi: 10.1109/98.878532.

A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, editors. *Wireless Sensor Networks: Signal Processing and Communications*. October 2007. ISBN 978-0-470-03557-3.

V. Tralli. On the energy efficiency of multihop communication in wireless sensor networks. In *Wireless Communication Systems, 2005. 2nd International Symposium on*, pages 488–492, September 2005. doi: 10.1109/ISWCS.2005.1547749.

V. Tsiatsis, R. Kumar, and M. B. Srivastava. Computation hierarchy for in-network processing. *Mob. Netw. Appl.*, 10(4):505–518, 2005. ISSN 1383-469X. doi: http://doi.acm.org/10.1145/1160162.1160174.

R. Vaughan and J. B. Andersen. *Channels, Propagation and Antennas for Mobile Communications*. The Institution of Electrical Engineers, Stevenage, United Kingdom, 2003. ISBN 0852960840.

R. G. Vaughan. Polarization diversity in mobile communications. *IEEE Transactions on Vehicular Technology*, 39(3):177–186, August 1990. ISSN 0018-9545. doi: 10.1109/25.130998.

S. Verdu. Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory*, 48(6):1319–1343, June 2002. ISSN 0018-9448. doi: 10.1109/TIT.2002.1003824.

B. Vucetic. An adaptive coding scheme for time-varying channels. *IEEE Transactions on Communications*, 39(5):653–663, May 1991. ISSN 0090-6778. doi: 10.1109/26.87156.

M. Wennström. *On MIMO systems and adaptive arrays for wireless communication*. PhD thesis, Uppsala University, Uppsala, Sweden, 2002. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-2604.

D. D. Wentzloff. *Pulse-Based Ultra-Wideband Transmitters for Digital Communication*. PhD thesis, Massachusetts Institute of Technology, Massachusetts, USA, 2007. http://www.eecs.umich.edu/ wentzlof/publications.html.

D. D. Wentzloff, B. H. Calhoun, R. Min, Alice Wang, N. Ickes, and A. P. Chandrakasan. Design considerations for next generation wireless power-aware microsensor nodes. In *VLSI Design, 2004. Proceedings. 17th International Conference on*, pages 361–367, 2004. doi: 10.1109/ICVD.2004.1260947.

M. Z. Win, F. Ramirez-Mireles, R. A. Scholtz, and M. A. Barnes. Ultra-wide bandwidth (UWB) signal propagation for outdoor wirelesscommunications. In *Vehicular Technology Conference, 1997 IEEE 47th*, volume 1, pages 251–255, Phoenix, AZ, USA, May 1997. doi: 10.1109/VETEC.1997.596358.

Wolfram Research Inc. Mathematica edition: Version 6.0, 2007.

W. Wu, M. A. T. Sanduleanu, X. Li, and J. R. Long. 17GHz RF front-ends for low-power wireless sensor networks. *2007. BCTM '07. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pages 164–167, September/October 2007. ISSN 1088-9299. doi: 10.1109/BIPOL.2007.4351860.

O. Yang and W. B. Heinzelman. Sensor selection cost function to increase network lifetime with qos support. In *MSWiM '08: Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 370–374,

New York, NY, USA, 2008. ACM. ISBN 978-1-60558-235-1. doi: http://doi.acm.org/10.1145/1454503.1454565.

M. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, and S. Singh. Exploiting heterogeneity in sensor networks. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 878–890, March 2005. doi: 10.1109/INFCOM.2005.1498318.

P. Youssef-Massaad. Impact of processing energy on the capacity of wireless channels. Technical report, 2005.

X. Zhao, J. Kivinen, P. Vainikainen, and K. Skog. Propagation characteristics for wideband outdoor mobilecommunications at 5.3 GHz. *IEEE Journal on Selected Areas in Communications*, 20(3):507–514, April 2002. ISSN 0733-8716. doi: 10.1109/49.995509.

L.C. Zhong, J.M. Rabaey, and A. Wolisz. Does proper coding make single hop wireless sensor networks reality: the power consumption perspective. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 2, pages 664–669, March 2005. doi: 10.1109/WCNC.2005.1424587.

M. Zorzi and R. R. Rao. Coding tradeoffs for reduced energy consumption in sensor networks. In *Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium on*, volume 1, pages 206–210, September 2004.