Energy aware evaluation of LTE Hybrid-ARQ and Modulation/Coding Schemes

Jan Dohl and Gerhard Fettweis
Vodafone Chair Mobile Communication Systems
Technische Universität Dresden, D-01062 Dresden, Germany
{jan.dohl,fettweis}@inf.et.tu-dresden.de
http://www.ifn.et.tu-dresden.de/MNS

Abstract—**Mobile internet devices are among the fastest growing markets in consumer electronics. A big issue with these devices is the low battery capacity that requires the user to recharge the device regularly. This paper aims to reduce the energy consumption of LTE modems used in future devices. For this, the LTE HARQ process is modeled as a Markov chain and a simple generic model for energy consumption is introduced that divides energy consumption in a static and a dynamic part. Using these and the results of link-level simulations it is found that throughput maximization will also maximize the energy efficiency if the static energy consumption dominates. However, if the energy consumption also contains a dynamic part, the energy efficiency will severely degrade unless another optimization approach is used.**

I. INTRODUCTION

Mobile internet devices like tablet PCs, eReaders, smartphones or netbooks are currently gaining a lot of popularity. Big names like Apple, Amazon, Nokia, Google and Microsoft have entered these markets and provide the devices as well as content. As a result of this development, the demand for high-bandwidth mobile internet connectivity rises continuously. It is therefore inevitable that future devices will contain LTE modems besides common interfaces like Wi-Fi, Bluetooth or USB. However, using modern mobile standards requires a lot of power and the development of new batteries cannot keep up which results in reduced operation time [1]. In order to reduce the power requirements of mobile devices, all components undergo research in order to increase their energy efficiency. This paper focuses on certain aspects of the LTE modem.

One of the most prevalent use cases for end user devices is the consumption of media. In this use case the LTE modem is mostly utilized for downloading data from the network and therefore the downlink needs to be especially energy efficient. A lot of applications like downloading periodical magazines, newspapers and radio/TV shows can be scheduled to off-peak hours and maximum throughput is only of subordinate priority. Therefore, in this paper we analyze the energy efficiency of LTE downlink transmissions from the perspective of the mobile terminal.

In LTE, data can be transmitted using numerous Modulation and Coding Schemes (MCS) [2] [3], each with different properties regarding data throughput and error protection. Furthermore, LTE uses a combination of Hybrid Automatic Repeat Request (HARQ) on a low layer and plain ARQ on a high layer in order to cope with erroneous transmissions [4].
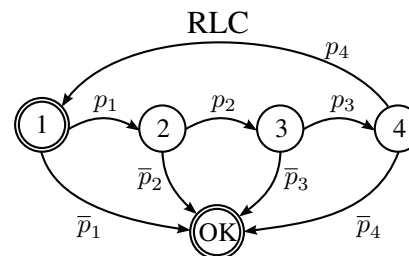

Fig. 1. HARQ process as Markov model, $N_r = 4$

HARQ unites the concepts of forward error correction and ARQ by combining faulty packets with the retransmissions yielding coding and repetition gains [5].

The paper is structured as follows. In section II we present an analytical way based on Markov chains that allows us to study the performance of HARQ using arbitrary cost functions. In order to look into energy efficiency, a model for the energy consumption of the LTE modem is required. In our project, no power data of the LTE modem is available at the time of writing. Therefore, we use a parameterized model containing dynamic and a static energy consumption as in [6]. The well-known model is presented in section III. In accordance with other publications we define the energy efficiency as the amount of successfully received information bits per unit energy [7]. In section IV we formulate the optimization problem using the previous results. Sections V and VI contain a description of the link-level simulator that was used to generate the parameters for the Markov model in section II and the results of the energy efficiency optimization. The paper is concluded in section VII.

II. HARQ MODEL

In LTE, the ARQ is implemented in a two staged model. The HARQ process runs on a low layer and therefore allows for very quick retransmissions. After N_r faulty transmissions the packet is reported to the upper layers as failed. A plain ARQ process in the Radio Link Control (RLC) layer is then used for retransmission which also allows the transmitter to change the MCS and other parameters. The concept is explained in [4].

For analysis the HARQ/ARQ process is modeled as a Markov chain. This is shown in Fig. 1 for up to four retransmissions. The process starts with transmission "1", which fails

with probability p_1, resulting in a second transmission denoted by "2". Note that due to the HARQ, the failure probabilities are usually not equal. After a maximum of N_r failed transmissions the high level ARQ process restarts the transmission. The process ends if the "OK"-state is reached which indicates a successful transmission.

The given Markov representation of the HARQ/ARQ process is a simplified model that is only valid under the following assumptions. We assume that the HARQ feedback channel and the packet error detection both work error free. We further assume that there are adequate measures in place to guarantee that the HARQ processes at transmitter and receiver are always synchronized. In order to model the RLC process as a transition back to state "1" we have to assume that the transmission parameters are not changed during RLC ARQ.

The resulting model is a special case of a Markov chain which has exactly one absorbing state. In our model this is the "Ok"-state. For the further analysis we need to know how often each state is passed on average until the absorbing state is reached. The state transition matrix $\mathbf{T}$ of the model is given by

$$\mathbf{T} = \begin{pmatrix} 0 & p_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & p_2 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & 0 & p_{N_r-1} \\ p_{N_r} & 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \quad (1)$$

and the initial state probability vector $\mathbf{t}$ is given as

$$\mathbf{t} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix}, \quad (2)$$

indicating that state "1" is always the starting state of the chain. Let x_n be a random variable (RV) whose realizations describe how often the n-th state is passed until the process ends. Then we can merge these RVs into a random vector $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_{N_r} \end{pmatrix}$ whose expectation $E[\mathbf{x}]$ is given as [8]

$$E[\mathbf{x}] = \mathbf{t}(\mathbf{I} - \mathbf{T})^{-1} \quad (3)$$

where $\mathbf{I}$ denotes the identity matrix.

Let $\mathbf{c} = \begin{pmatrix} c_1 & c_2 & \cdots & c_{N_r} \end{pmatrix}^T$ be an arbitrary cost vector whose elements c_n contain the cost that is incurred each time state n is reached. The total cost $r_{\mathbf{c}}$ of the process is then another random variable and the average cost $E[r_{\mathbf{c}}]$ is:

$$E[r_{\mathbf{c}}] = \mathbf{t}(\mathbf{I} - \mathbf{T})^{-1}\mathbf{c} \quad (4)$$

A special cost vector is $\mathbf{c} = \mathbf{1}^T$ where $\mathbf{1}$ denotes a vector containing only ones. This cost vector yields the average amount of transmissions $r = \mathbf{t}(\mathbf{I} - \mathbf{T})^{-1}\mathbf{1}^T$ until the process succeeds.

III. ENERGY MODEL

We assume that the reception circuitry of the LTE modem has a constant power dissipation of P_s while enabled. We further assume that the power dissipation during sleep mode and when being switched on or off can be neglected. Thus, only the active mode is considered and the static energy

consumed during the reception of one OFDM symbol is given as

$$E_s = P_s \cdot T_{sym} \quad (5)$$

where T_{sym} is the duration of one OFDM symbol. On the other hand, the dynamic power depends on how much computational effort is required to process one OFDM symbol. The major source of computational complexity in LTE is the turbo decoder whose complexity depends linearly on the amount of code bits per OFDM symbol N_c and the number of turbo iterations N_{ti}. Assuming that processing one bit requires an energy of E_{bit} per iteration, the dynamic energy consumption per OFDM symbol E_d can be written as

$$E_d = E_{bit} \cdot N_{ti} \cdot N_c. \quad (6)$$

The total energy consumption E to process one OFDM symbol is then given as the sum of these components:

$$E = E_s + E_d. \quad (7)$$

IV. ENERGY OPTIMIZATION

In LTE, different MCS are available which differ greatly in terms of data throughput and error protection. The goal of the optimization shall be to find the MCS that yields the highest energy efficiency. First, we define the vector $\mathbf{b}$ whose elements completely describe a certain MCS. We do not define $\mathbf{b}$ in detail in order to keep the following discussion as generic as possible.

At the time of writing this paper, neither P_s or E_{bit} are known. In order to use the model presented in the last section, we introduce the following assumptions which allow a generic approach to the analysis. We note that in order to select the most energy efficient MCS, the actual values of E are not important as long as the ratios are correct for different MCS. Furthermore, the values of N_{ti} and N_c depend on the selected MCS and therefore become functions of $\mathbf{b}$. Let $v \in [0\ldots1]$ denote the ratio of static to total energy consumption. We define

$$P_s = v \quad (8)$$

and

$$E_{bit}(\mathbf{b}) = \frac{1 - v}{N_{max}} \quad (9)$$

where $N_{max} = \max[N_{ti}(\mathbf{b}) \cdot N_{ti}(\mathbf{b})]$. Note that N_{max} is a normalization that is merely used for convenience because otherwise the values of E_s and $E_d(\mathbf{b})$ would be very different which makes the comparison less intuitive. Using these definitions the total energy consumption becomes

$$E(\mathbf{b}) = v \cdot T_{sym} + (1 - v) \cdot \frac{N_{ti}(\mathbf{b}) \cdot N_c(\mathbf{b})}{N_{max}}. \quad (10)$$

For $v = 1$ the static energy consumption dominates and for $v = 0$ the dynamic energy consumption dominates. For real systems, a value in between the two extreme cases will describe the ratio of energy consumption.

The state transition probabilities p_n which model the failure probabilities in each transmission of the HARQ process also
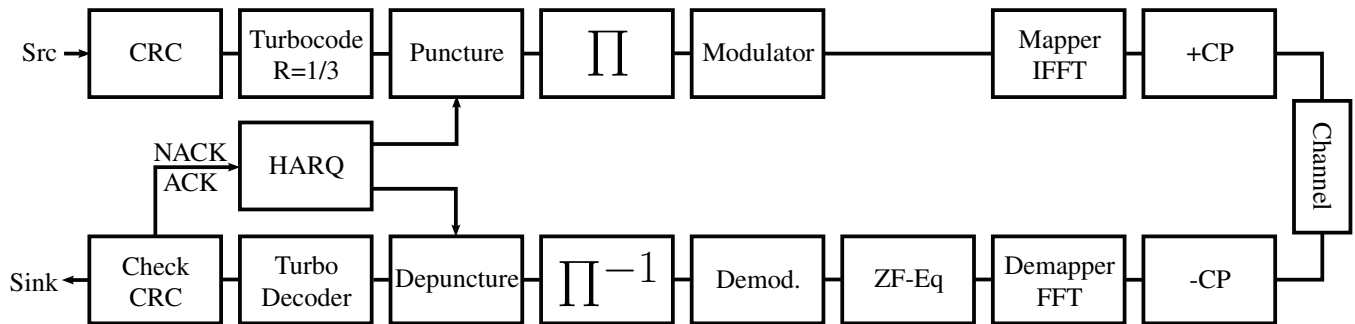
Fig. 2.   Link level system model

strongly depend on the current MCS which makes the state transition matrix dependent on $\mathbf{b}$. The cost vector $\mathbf{c}$ in eq. (4) contains the energy cost per OFDM symbol and also depends on the current MCS. We assume that the consumed energy is constant for each transmission so that $\mathbf{c}$ can be written as $\mathbf{c}(\mathbf{b}) = E(\mathbf{b}) \cdot \mathbf{1}$. Thus, the average energy consumed for successfully receiving and processing one OFDM symbol with the MCS described by $\mathbf{b}$ is given by

$$\overline{e}(\mathbf{b}) = \mathrm{E}[r_{\mathbf{c}}(\mathbf{b})] = E(\mathbf{b})\tau(\mathbf{I} - \mathbf{T}(\mathbf{b}))^{-1}\mathbf{1} \qquad (11)$$
$$= E(\mathbf{b})r(\mathbf{b}) \qquad (12)$$

where $r(\mathbf{b})$ denotes the average amount of retransmissions.

While minimizing $\overline{e}(\mathbf{b})$ over $\mathbf{b}$ minimizes the energy consumption for the reception of one OFDM symbol it is an unfair metric as the amount of information per OFDM symbol also depends on the current MCS. Therefore, $N_i(\mathbf{b})$ shall be the amount of information bits per OFDM symbol for MCS $\mathbf{b}$. We define the amount of information bits per unit energy as

$$i_{\overline{e}}(\mathbf{b}) = \frac{N_i(\mathbf{b})}{\overline{e}(\mathbf{b})} \ . \qquad (13)$$

The optimal MCS $\mathbf{b}_{opt}$ is the one which maximizes (13) so that we can specify the optimization problem as follows:

$$\mathbf{b}_{opt} = \arg\max_{\mathbf{b}}[i_{\overline{e}}(\mathbf{b})] \qquad (14)$$

An equivalent problem is to minimize the energy consumption per information bit that is given by the inverse of $i_{\overline{e}}(\mathbf{b})$.

A. Extreme cases

We now give some theoretical results for the cases of dominating static power ($v = 1$) and dominating dynamic power ($v = 0$). For the case $v = 1$ we get

$$E(\mathbf{b}) = T_{sym} \qquad (15)$$

and

$$i_{\overline{e}}(\mathbf{b}) = \frac{N_i(\mathbf{b})}{r(\mathbf{b})T_{sym}} \qquad (16)$$

which can be interpreted as the effective amount of information bits per second. Therefore, setting $v = 1$ leads to a classical rate maximization.

On the other hand, for $v = 0$ we get

$$E(\mathbf{b}) = \frac{N_{ti}(\mathbf{b}) \cdot N_c(\mathbf{b})}{N_{max}} \qquad (17)$$

and

$$i_{\overline{e}}(\mathbf{b}) = \frac{a \cdot N_{max}}{N_{ti}(\mathbf{b}) \cdot r(\mathbf{b})} \qquad (18)$$

where we assume that the ratio $a = \frac{N_i(\mathbf{b})}{N_c(\mathbf{b})}$ of information bits to coded bits is constant for all MCS. This is not true for the data sent over the actual air interface. However, the receiver depunctures the incoming data and the decoder works on the depunctured data stream that always has the rate of the mothercode. The rates of the MCS become irrelevant in the optimization of energy efficiency and the amount of turbo iterations as well as repetitions is minimized. We therefore expect that mainly low rate MCS are selected in order to reach a high probability of requiring only one transmission and turbo iteration for smaller SNR.

V. LINK LEVEL SIMULATOR

A link level simulator was employed to obtain the probabilities $p_n(\mathbf{b})$ of the state transition matrix $\mathbf{T}(\mathbf{b})$. There is a finite set of simulated MCS $\mathbf{b}$ and finding $\mathbf{b}_{opt}$ is then simplified to calculating $i_{\overline{e}}(\mathbf{b})$ for all possible MCS and selecting the MCS that yields the largest $i_{\overline{e}}(\mathbf{b})$. The system model used for the link-level simulations is shown in Fig. 2. The mother code is a rate 1/3 turbo code as in [2]. Afterwards, the code is punctured to obtain the code rates 1/3, 1/2 and 3/4. The modulator maps the bits to symbols using QPSK, 16-QAM or 64-QAM constellations. After mapping the symbols to carriers the symbol is sent over the mobile channel using usual OFDM methods. After demapping, the effects of the frequency selective channel are equalized using a frequency domain zero forcing equalizer under the assumption of perfect channel knowledge. The turbo decoder in the receiver consists of two BCJR-decoders that exchange extrinsic information about the systematic bits in a turbo loop. Afterwards, the data block is checked for correctness using the CRC checksum. If the transmission was faulty the transmitter will send a repetition. Because the HARQ is designed to have short delays between transmissions, we assume that the channel does not change between transmissions, i.e. there is no temporal diversity gain from retransmissions.

TABLE I
SIMULATION PARAMETERS

Modulation schemes	QPSK, 16-QAM, 64-QAM
HARQ Schemes	Type I, Type III
Retransmissions	4
Code rates	1/3, 1/2, 3/4
Turbo iterations	1, 2, 4, 8
Channel	Urban Macro (SCME)
Realizations	500
OFDM subcarriers	1024
Used subcarriers	601
T_{sym}	$71.4\mu s$

Fig. 4.   Information bits per unit energy for $v = 1$

Fig. 3.   Information bits per unit energy for $v = 0$

Fig. 5.   Throughput and energy efficiency characteristics for low rate MCS (dashed curves) and high rate MCS (solid curves)

The channel is modeled using the tapped delay-line model of the SCME Urban Macro channel [9]. $N = 500$ channel realizations are generated and simulated per MCS yielding the failure probabilities $p_{k,n}(\mathbf{b})$ where $k \in [1..4]$ is the transmission index and $n \in [1..N]$ is the index of the channel realization. Using these results the average failure probabilities $p_k(\mathbf{b})$ in Fig. (1) are then calculated as

$$p_1(\mathbf{b}) = \frac{1}{N}\sum_{n=1}^{N} p_{1,n}(\mathbf{b}) \tag{19}$$

and

$$p_k(\mathbf{b}) = \frac{\sum_{n=1}^{N}\prod_{l=1}^{k} p_{l,n}(\mathbf{b})}{\sum_{n=1}^{N}\prod_{l=1}^{k-1} p_{l,n}(\mathbf{b})}. \tag{20}$$

Here, $p_k(\mathbf{b})$ denotes the probability of failure in state k given that all previous transmissions have also failed, corresponding to the Markov description of the HARQ process.

Table I summarizes the simulation parameters. It should be noted that no scheduling or power allocation over the subcarriers happens. All available subcarriers are being used for the transmission with equal power per subcarrier independently of possible deep fades in the transfer function of the channel.

VI. RESULTS

Fig. 3 and 4 show how many information bits per unit energy are received correctly for the extreme cases $v = 0$ and $v = 1$. It shows that the analytic prediction we did in
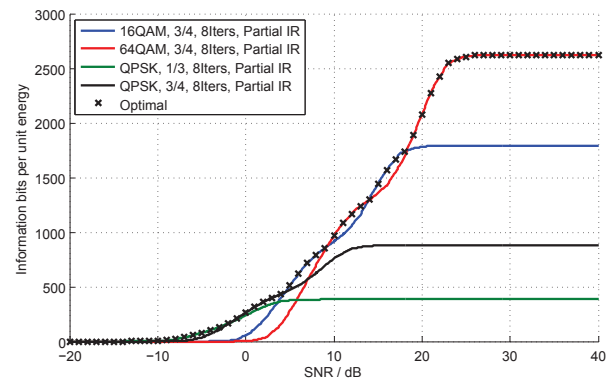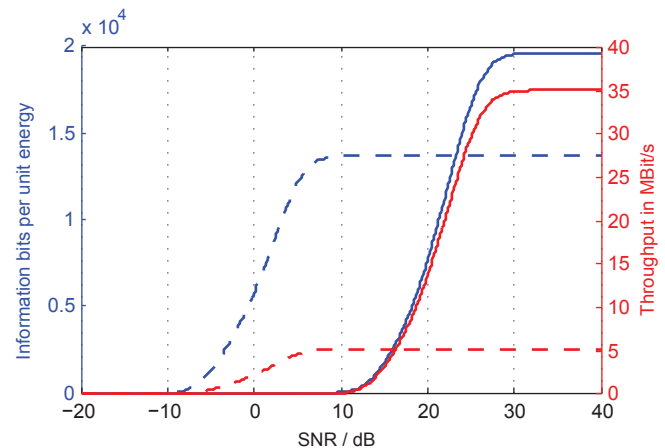
section IV-A is correct. For dominating static power the MCS with high rates are selected for very small SNRs already. It also solely uses MCS with the maximum amount of iterations because those become reliable for smaller SNR. On the other hand, for dominating dynamic power the lowest rate MCS with QPSK, rate 1/3 and 1 iteration is selected for most of the time as predicted.

A very interesting metric is the loss of throughput when the MCS is selected according to optimal energy efficiency. On the other hand, the loss of energy efficiency compared to the optimal energy efficiency is of interest when the throughput is optimized. The following curves will have a lot of discontinuities and we first explain the reasons for them. Fig. 5 shows the efficiency and throughput curves for two MCS. It can be easily seen that the SNRs where the curves intersect are different for energy efficiency and throughput. Therefore, if we optimize for throughput but then plot the resulting energy efficiency the curve will have a discontinuity at the position where the MCS is switched. This can be seen in Fig. 6. Obviously, a similar effect can be seen when optimizing for energy efficiency and then plotting the resulting throughput.

Fig. 7 shows the throughput loss, compared to the maximum throughput, when optimizing for energy efficiency. We added a new criteria to the optimization that selects the highest rate MCS in cases where several MCS have equal energy
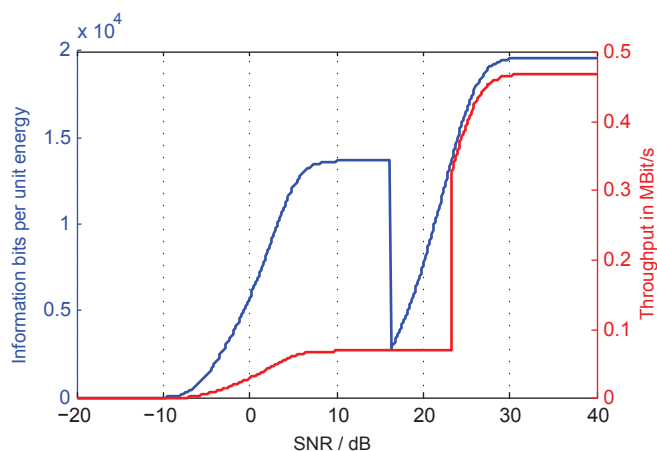
Fig. 6. Resulting throughput when optimizing for energy efficiency (red line) and resulting energy efficiency when optimizing for throughput (blue line) for MCS from Fig. 5
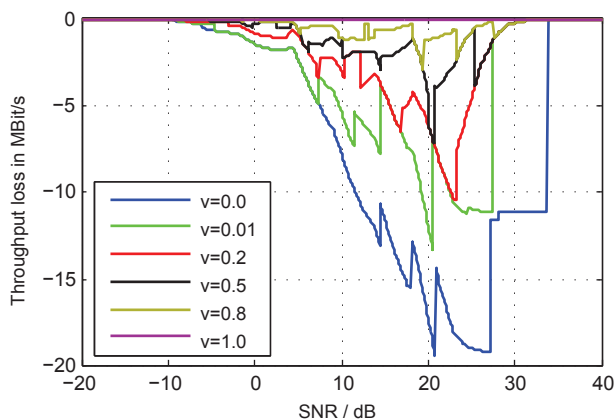

Fig. 7. Throughput loss when optimizing energy efficiency

efficiency. It can be seen that the massive losses for $v = 0$ quickly decline even for very small values like $v = 0.01$. It can also be seen that the losses vanish completely for the high SNR region. The reason for this is that for each MCS the average amount of retransmissions will approach one in the high SNR region

$$\lim_{SNR \to \infty} [r(\mathbf{b})] = 1 \qquad (21)$$

even for just one turbo iteration. If we exclude all MCS where $N_i(\mathbf{b}) > 1$ then the dynamic energy consumption $E_d(\mathbf{b})$ and as a result the total energy consumption $E(\mathbf{b})$ become constants. This means that the energy efficiency optimization equals a rate maximization in the high SNR region for all v.

Fig. 8 shows the energy efficiency, compared to the optimal energy efficiency in that scenario, when optimizing for throughput. It can be seen that throughput optimization leads to a significant waste of energy even if there is high static energy consumption. Similar to the previous case, the energy efficiency becomes ideal in the high SNR region.

VII. CONCLUSIONS

In this paper we present a way to describe and analyze HARQ processes using a simple yet easily extendable Markov
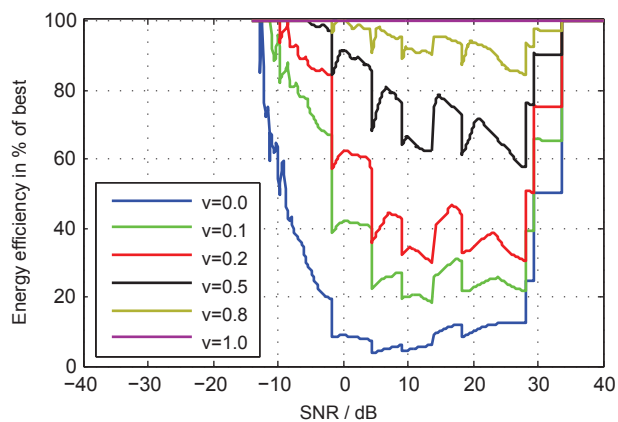

Fig. 8. Resulting energy efficiency when optimizing for throughput

model. Due to the lack of real power measurements of a working LTE modem we assume a model that contains static and dynamic energy consumption with parameterizable ratio. This allows for an interesting analysis of the extreme cases. We simulate the most common MCS for LTE and optimize them for energy efficiency. It shows that throughput optimization leads to significant waste of energy if the static energy consumption is not clearly dominating.

For the regular download of large amounts of data during off-peak periods it can therefore make sense to use energy optimal transmission methods. Maximum throughputs are not required for this use case but the energy savings will conserve the battery.

REFERENCES

[1] Y. Neuvo, "Cellular phones as embedded systems," *2004 IEEE International Solid-State Circuits Conference (IEEE Cat. No.04CH37519)*, pp. 32–37, 2008.
[2] T. Specification, "TS 136 212 - V8.7.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding (3GPP TS 36.212 version 8.7.0 Release 8)," 2009.
[3] T. Specification, "TS 136 211 - V9.0.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211 version 9.0.0 Release 9)," 2010.
[4] M. Meyer, H. Wiemann, M. Sagfors, J. Torsner, and J.-f. Cheng, "ARQ Concept for the UMTS Long-Term Evolution," *IEEE Vehicular Technology Conference*, pp. 1–5, Sep. 2006.
[5] K. C. Beh, A. Doufexi, and S. Armour, "Performance Evaluation of Hybrid ARQ Schemes of 3GPP LTE OFDMA System," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, 2007, pp. 1–5.
[6] G. A. S. Machado, *Low-power HF microelectronics: a unified approach*. IET, 1996.
[7] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.
[8] V. R. G. Latouche, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial Mathematics, 1999.
[9] D. S. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, vol. 5, 2005, pp. 3132 – 3136 Vol. 5.