Generalised link-layer adaptation with higher-layer criteria for energy-constrained and energy-sufficient data terminals

Virgilio Rodriguez #1, Rudolf Mathar #2

[#] Institute for Theoretical Information Tech., RWTH Aachen Aachen, Germany

{ rodriguez, ² mathar}@ti.rwth-aachen.de

Abstract-It has long been recognised that a wireless communication system can be more efficient if link-layer parameters such as modulation order are adapted to the channel conditions. The most common optimising criterion is spectral efficiency (bps/Hertz) subject to very low bit-error constraint. But this criterion is not appropriate for practical communication networks fitted with strong data-packet error detection and a mechanism for selective packet re-transmission. In such scenario, packetoriented criterion for link adaptation, such as the "goodput", seems more appropriate. Herein we follow recent literature in performing link adaptation under "goodput" oriented criterion. But we focus on axiomatic properties of the packet-success rate function (PSRF). The benefits are generality, robustness and technical accuracy. In particular, our analysis applies to arbitrary combinations of link-layer parameters, as long as the resulting PSRF is an S-curve. We obtain a robust result: a set of possible link configurations can be ranked by the slope of a tangent line from the origin to the graph of a scaled version of the PSRF. Some link configurations can dominate others under all scenarios. We consider separately energy-limited (e.g., batteryfed) and energy-sufficient (e.g., vehicular) terminals, and provide graphical illustration of our results.

I. INTRODUCTION

The importance of (adaptively) optimising the link layer configuration of a wireless communication system has long been recognised. The adaptation to a time-varying channel of such parameters as constellation size, symbol rate and coding can result in increased efficiency. In particular, modulation adaptation with the aim of maximising performance subject to a maximal bit-error constraint has received significant attention in the literature[1], [2], [3], [4]. In these and most studies, the focus is the symbol: for example, to choose bits per symbol in order to maximise bps/Hertz (spectral efficiency), while holding the bit error rate under a specified (very low) level.

More recently, it has been recognised that packet-oriented (OSI layer-2) link adaptation can be more appropriate under certain interesting scenarios [5]. For example, in practical communication networks, the delivery of medium-access-control (MAC) packets is (idealised as) "guaranteed". In a common scheme, binary data is packetised, strong error-detecting codes (e.g., cyclic-redundancy codes (CRC)) are added to each data packet, and the system is fitted with an automated repeat-request (ARQ) mechanism. Sufficiently long CRC codes (e.g.,

16 bits) perform extremely well in practise, and can be assumed to provide perfect error detection. Likewise, the feedback ARQ channel can be designed to performed close to ideally. Thus, the system can be idealised as transmitting each packet the necessary number of times until it is received correctly. Under these conditions, the packet-error probability at the input of the error detector, and hence the bit-error probability can be relatively high, because a packet received in error is simply retransmitted. Too many retransmissions will, of course, depress the system's performance. But aiming for an extremely low bit error rate (none or very few retransmissions) may also be resource-wasteful. The resource manager should focus on post-retransmission performance (e.g., the "net throughput" after considering retransmissions). Therefore, the common criterion of maximising spectral efficiency subject to an extremely low bit-error constraint does not seem appropriate for the ARQ-based system. Thus, [5] seeks to (adaptively) configure the link layer parameters — that is, parameters such as packet and CRC lengths, number of bits per symbol, and symbol rate — with the aim of maximising a measure of throughput ("goodput").

More recently, on the basis of models from [3], [6] seeks adaptive rate and power policies for network utility maximisation (NUM) with an average powerconstraint. The network utility is taken as the sum of the utility achieved by each individual link, which is measured through a simple formula of the rate through the link. The NUM-based modulation adaptation policies reported by [6] are significantly different from those derived with a focus on spectral efficiency.

The present work has the same aims and scope of [5]; that is, it seeks optimal link-adaptation policies under packet-focused criteria (e.g., "goodput"). But the technical approach herein is substantially different. The development in [5] follows an algebraic approach that requires full knowledge of an explicit formula ("equation") for the packet success rate function (PSRF), such as $[1 - P_b(\gamma_s, b)]^{L/b}$ (discussed further below). Such specific formulae are valid only under strong assumptions, and/or major simplifications. By contrast, we take an axiomatic approach: we postulate that the packet success rate function have some basic properties that are reflected on the general geometric shape of its graph, and make our derivations on the basis of those properties. The benefits of our approach are generality, robustness, and technical accuracy. Our analysis is relevant for *any* arbitrary combination of linklayer parameters, provided that the PSRF satisfies some simple properties, directly connected to the geometrical characteristics of its graph. And our derivations follow from those properties, and avoid certain simplifications and idealisations necessitated by the algebraic analysis, but that seem controversial (for example, [5] treats all relevant parameters as continuous variables, including the number of bits per M-QAM symbol, and takes derivative with respect to them in search of the optimal configuration).

Below, we describe our system model, after giving further detail on [5]'s. Then, we discuss optimisation criteria appropriate for link layer configuration involving data traffic, considering separately energy-limited and energy-sufficient terminals. Subsequently, we provide the core of our analysis and main results with a focus on the energy-limited terminal. In the immediately following section, we show that the preceding analysis also applies to the energy-sufficient data terminal that can adjust its symbol rate. Then we illustrate and discuss our results. An appendix provides some additional technical details.

II. SYSTEM MODEL

A. Previous formulation

Reference [5] focuses on a single communication link, and M-QAM modulation and defines a terminal's throughput by

$$T = \frac{L - C}{L} bR_s f(b, \gamma_s, L) \tag{1}$$

where L, C, b, R_s , and γ_s are, respectively, the packet length, CRC length, bits/symbol, symbol rate, and signal-to-noise ratio *per symbol.* f yields the packet-success rate, and in [5]'s development is assumed equal to $[1 - P_b(\gamma_s, b)]^{L/b}$ where P_b denotes the symbol-error probability. For P_b approximate closed-form expressions from the literature are used.

B. Present formulation

- N_0 is the average Gaussian noise spectral density
- E is the energy budget, when applicable
- \hat{P} is the power constraint, if any
- *H* is the channel gain, and $h := H/N_0$
- p = HP is the received power
- R_s is the symbol rate
- b bits per symbol
- σ_s is the signal-to-noise ratio (SNR) per symbol
- L-bit packets carrying L C information bits are used.
- For a given combination of the *relevant* parameters, a, $F(x; \mathbf{a})$ is the packet-success rate function (PSRF) (i.e., 1 minus the packet error rate), which gives the probability of correct reception of a data packet as a function of the symbol SNR, x. For example, in the notation of [5] and implicitly assuming independent bit errors $F(x; \mathbf{a}) = [1 P_b(x, b)]^{L/b}$.



Figure 1. An S-curve, S(x), xS'(x) (solid bell curve), and the tangenu (tangent line from origin). The inflexion (and hence the peak of S'(x) which is *not* shown) occurs at x = i, and the genu (knee) at (K,k). S(x)/x (dash, scaled) is maximised at x = K. The interpretation of x and S(x) varies depending on the context.



Figure 2. (Axes intentionally left unmarked). The S shape is very versatile : Besides the S-shape proper (U_3) , it also covers the "ramp" (U_2) , the ("mostly") concave (U_1) , and the ("mostly") convex curve (U_4) . The "ramp" of an S-curve (see U_3) can approximate a line over a range of interest.

For some technical reasons, f(x; a) := F(x; a) - F(0; a) replaces F [7] (e.g., for the previous example, f(x; a) = [1 - P_b(x, b)]^{L/b} - [1 - P_b(0, b)]^{L/b}). For any a, f is assumed to satisfy Definition A.1; that is, its graph as function of the *per symbol* signal-to-noise ratio (SNR) has the S-shape shown in fig. 1, which is a very mild assumption (see fig. 2). Our analysis does *not* rely on any specific PSRF. The relevant parameters depend on specific characteristics of the communication system, and technical assumptions made by the analyst. Further details on the technical properties of f are in the appendix.

For convenience, x may be used as a *generic* function argument (as in "the graph of $g(x) = \sqrt{x}$ is concave").

C. Information transferred over a period of interest

The total number of information bits that a terminal operating with packet-success rate $f(\cdot; \mathbf{a})$ and SNR held at σ_s can transfer over time period τ is given by

$$\tau \frac{L-C}{L} b R_s f(\sigma_s; \mathbf{a}) \tag{2}$$

III. LINK ADAPTATION CRITERIA FOR DATA TRAFFIC

If the terminal of interest has reliable information about future channel states it could consider it. In practise, such information is difficult to obtain. Thus the terminal opts to manage its resources as if the present channel state were permanent (that is, it is the most reliable estimate of the future state). As channel state information is updated periodically, the terminal re-calculates its optimum which may lead to a change in resource allocation and/or link configuration.

We focus on a terminal that has a long stream of (delaytolerant) data to transfer. Such a terminal's performance is determined by the total number of information bits that it can transfer over a period of interest. The period of interest is determined by the status of its energy supply: limited (battery) versus unlimited (power grid, vehicular applications, etc).

- For an energy-limited data terminal with plenty of data to transfer, the natural period of interest is the total time of operation by the time its energy runs out; that is, its "battery life". Thus, the terminal manages its resources with the aim of maximising the total number of bits it can transfer over the lifetime of its battery charge ("bits per Joule" (bpJ) maximisation).
- For a terminal with an unlimited energy supply, and a long stream of (delay-tolerant) bits to transfer a natural period of interest is the time unit. Thus, such terminal makes choices with the aim of maximising the total number of information bits that it transfers per unit of time ("net" throughput or "goodput").

IV. ENERGY-EFFICIENT LINK CONFIGURATION

Lemma IV.1 For a given link configuration, \mathbf{a} , a terminal with energy budget E and a long stream of data to transfer allocates its resources seeking to maximise an expression of the form $S(x; \mathbf{a})/x$ where x denotes the terminal's signal-to-noise ratio and S satisfies Definition A.1.

Proof: If the terminal operates with constant power P, and symbol rate R_s , it achieves SNR of hP/R_s , and its battery lasts $\tau = E/P$.

By (2), over τ the terminal transfers

$$\frac{E}{P}\frac{L-C}{L}bR_sf(h\frac{P}{R_s};\mathbf{a}) = Eh\frac{L-C}{L}b\frac{f(x;\mathbf{a})}{x} := Eh\frac{S(x;\mathbf{a})}{x}$$
(3)

with $x = hP/R_s$ and $S(x; \mathbf{a}) = ((L - C)/L)bf(x; \mathbf{a})$

For a given link configuration, b(L - C)/L is a constant. Thus, S satisfies Definition A.1 because f does.

Remark IV.1 Equation (3) implies that the terminal will choose an optimal ratio P/R_s . Whether the symbol rate is adjustable or not, the bpJ criterion will not specify the value of each variable, only their ratio. If the symbol rate is adjustable, then regardless of its power limit and channel

state, the terminal can in principle reach any desired SNR by sufficiently lowering its symbol rate.

Lemma IV.2 Consider the problem of maximising S(x))/xsubject to $0 \le x \le X$, where S satisfies Definition A.1. Then, (i) there is a unique tangent line from the origin to the graph of S, denoted as c^*x and called the tangenu, with tangency point, genu, $(x^*, S(x^*))$, with $c^* = S(x^*)/x^*$ (ii) S(x))/x is strictly quasi-concave, and (iii) its unique maximiser is $\min(x^*, X)$.

Remark IV.2 Figure 1 shows the tangenu (tangent from (0,0)) and genu (tangency point) for the S-curve $S(x) = [1 - \exp(-x/2)/2]^{80} - 2^{-80}$, as well as graphs of xS'(x) and S(x)/x.

Theorem IV.1 Consider a set of I link layer configurations each identified by a combination of parameters \mathbf{a}_i ; that is, $\mathbf{a}_1, \ldots, \mathbf{a}_I$. With each i associate the function $S(x; \mathbf{a}_i) = b(L - C)f(\cdot; \mathbf{a}_i)/L$ and denote the abscissa of its genu as x_i^* . $S(x_i^*; \mathbf{a}_i)/x_i^* \ge S(x_j^*; \mathbf{a}_j)/x_j^*$ implies that if the terminal can reach both SNR values x_i^* and x_j^* then it can achieve at least as many bits per Joule with link configuration \mathbf{a}_i as it can with configuration \mathbf{a}_j . Furthermore, there is a bits-per-Joule-optimal configuration j^* defined by $S(x_{j^*}^*; \mathbf{a}_{j^*})/x_{j^*}^* =$ $\max_i \{S(x_i^*; \mathbf{a}_i)/x_i^*\}$; that is, $S(x; \mathbf{a}_{j^*})$ has the steepest tangenu among the functions $S(x; \mathbf{a}_i)$.

Proof: By Lemma IV.1, for any $\mathbf{a_i}$, the terminal seeks to maximise $EhS(x; \mathbf{a_i})/x$.

By Lemma IV.2, the maximum occurs at the genu (tangency point) of $S(x; \mathbf{a}_i)$.

Then, the terminal's maximal bits-per-Joule performance with link configuration *i* is $EhS(x_i^*; \mathbf{a_i})/x_i^*$.

Since Eh is the same value for all the considered configurations, between two considered configurations, the one with higher ratio $S(x_i^*; \mathbf{a_i})/x_i^*$ is bpJ-better.

Therefore, the bpJ-optimal configuration has the highest ratio $S(x_i^*; \mathbf{a_i})/x_i^*$ (which is the slope of its tangenu).

Remark IV.3 Theorem IV.1 indicates that a set of link configurations can be sorted in order of energy efficiency. Hence, there is a bpJ-ideal configuration in the set which can be identified graphically because the tangenu of its associated S-curve is the steepest (has the greatest slope) among those considered (see Figure 3 for a case in which the considered configurations have common L, C, b, and Figure 4 for a more general case). A terminal with an adjustable symbol rate can always reach the optimal operating point $x_{j^*}^*$ of the most efficient configuration \mathbf{a}_{j^*} (see Remark IV.1). However, a power-limited terminal with a fixed symbol rate may be unable to reach $x_{j^*}^*$ and consequently may need to choose the most efficient of those configurations whose optimal operating points can actually be reached. As discussed in section VI concerning Figure 4, the ideal configuration \mathbf{a}_{j^*} could still outperform the next one if the highest reachable SNR is sufficiently close to x_{j*}^* .

Corollary IV.2 If $S(x_i^*; \mathbf{a}_i)/x_i^* \ge S(x_j^*; \mathbf{a}_j)/x_j^*$ and $x_i^* \le x_j^*$ then \mathbf{a}_i dominates \mathbf{a}_j in the sense that \mathbf{a}_i provides superior performance, and \mathbf{a}_i 's optimal operating point is reachable whenever \mathbf{a}_i 's optimal operating point is.

Proof: Follows directly (see also Remark IV.3).

Remark IV.4 By Corollary IV.2, a link parameter combination can be permanently eliminated if whenever its implied optimal operating point can be reached, there is another combination of parameters that performs better and whose optimal operation point can also be reached.

V. LINK CONFIGURATION FOR MAXIMAL THROUGHPUT

Lemma IV.1 is the key to the development in section IV. It relies on the fact that battery life equals E/P, which does not apply for the energy-sufficient terminal. For this terminal, the natural period of interest is the time unit. And it is intuitively clear that it will set its power to the highest available level. However, if this terminal has an adjustable symbol rate, then a result analogous to Lemma IV.1 can be proved.

Lemma V.1 For a given link configuration, \mathbf{a} , an energysufficient terminal with adjustable symbol rate, power limit \hat{P} , and a long stream of data to transfer seeks to maximise an expression of the form $S(x; \mathbf{a})/x$, where x denotes the terminal's signal-to-noise ratio and S satisfies Definition A.1.

Proof: The terminal operates with constant power \hat{P} . If it sets its symbol rate to R_s , it achieves SNR of $h\hat{P}/R_s$.

By (2), over a time unit the terminal transfers

$$\frac{L-C}{L}bR_s f\left(h\hat{P}/R_s;\mathbf{a}\right) = \hat{P}h\frac{L-C}{L}b\frac{f(x;\mathbf{a})}{x} := \hat{P}h\frac{S(x;\mathbf{a})}{\overset{(4)}{x}}$$

with $x = hP/R_s$ and $S(x; \mathbf{a}) = ((L - C)/L)bf(x; \mathbf{a})$ Therefore, for given link configuration, the throughputmaximising terminal will seek to maximise $S(x; \mathbf{a})/x$.

For a given link configuration, b(L - C)/L is a constant. Thus, S satisfies Definition A.1 because f does.

Remark V.1 Lemma V.1 takes the place of Lemma IV.1 and makes evident that the rest of the results and discussion in section IV also apply to an energy-sufficient terminal with adjustable symbol rate and a long stream of data to transfer.

VI. DISCUSSION

We have followed the rationale and motivation of [5] leading to the conclusion that packet-oriented link adaptation (for example based on OSI layer-2) — as opposed to symbol/ spectral-efficiency oriented adaptation — can be more appropriate for complex multi-layered networks. However, [5] took an algebraic approach that requires highly specific closed-form expressions, and involves certain approximations and analytical steps that seem controversial. By contrast, we



Figure 3. The link-layer configuration whose associated S-curve (a multiple of the PSRF) has the "steepest" tangenu (red, solid) is best. But if the terminal cannot reach the corresponding genu (x^*) , it may need to configure its link layer differently. Bits per symbol are the same for these configurations (as those in [9]'s Table I).



Figure 4. Link layer configuration with different bits per symbol. The combination of parameters corresponding to S_2 is better whenever SNR x_{12} (intercept of tangenu T_1 and S_2) can be reached. Configuration 3 is eliminated, because whenever x3 is feasible so is x2, and S_2 performs better (i.e., $S_2(x2)/x2 > S_3(x3)/x3$). S_1 is used if x_{12} cannot be reached.

undertook an axiomatic approach. We observed that the critical information from any given combination of link configuration parameters is fully reflected in the packet-success rate function (PSRF) of the symbol SNR. Then we postulated that the graph of this function is a smooth increasing bounded curve, and grounded our analysis on analytical geometry. This assumption is very robust, because, as shown by Figure 2, the S-shape is very versatile, including curves that are sufficiently close to many common shapes. Thus, the graph of any real PSRF, if not an exact S-curve, should be close enough to one such curve. Furthermore, our analysis is not tied to the myriad of assumptions that are necessary to obtain a PSRF with a remotely tractable algebraic form. An our development is in fact "exact" in the sense that no approximation is used, and its technical steps seem uncontroversial.

In spite of the above, our results are quite clear, simple, and can in fact be implemented. The main result is that we can rank an arbitrary collection of link layer configurations - each characterised by a vector of parameters - by performing two rather simple operations: plotting the graph of a scaled version of the PSRF, and drawing a tangent line from the origin to the resulting graph. The graph may be the result of exact or approximate analysis, or of simulations or experimental measurements. Once the graph is available, the simple operation of drawing a tangent line (or the equivalent simple computer algorithm) can be performed. Thus, each combination of parameters can be summarised by two values: the optimal operating point (SNR) — found at the tangency point (genu) – and the slope of the tangent line (tangenu). A greater slope (steeper tangent) is better. Some configurations can be eliminated, while others that are in principle suboptimal may be used when the optimal SNR of a better one cannot be reached.

Figures 3 and 4 illustrate the procedure. In Figure 3 the considered configurations involve the same number of bits per symbol and packet parameters, as with the binary modulation schemes from Table I of [9] (presumably, an ideal softwaredefined radio could switch from any such modulation scheme to another [10]). Figure 4 is relevant to link configurations involving the "order" of a family of modulation schemes and/or variation in packet size — such as choosing an M from the M-QAM family, as in the literature that followed [1]. In Figure 4, the combination of parameters corresponding to S_2 is better whenever its operating point x^2 can be reached. The combination of parameters associated with S_2 "dominates" that leading to S_3 , because S_3 has a less steep tangenu (i.e., performs worse) and a higher operating point (thus, whenever $S'_{3}s$ optimal operating point is feasible so is S_{2} 's, and S_{2} performs better). The combination of parameters leading to S_1 is not completely eliminated, because its optimal occurs at a lower SNR than S_2 's, hence S_1 may be usable when x_2 cannot be reached. But notice that if the highest reachable SNR is less than x^2 but greater than x_{12} (at the intercept of tangenu T_1 with S_2), then S_2 would still be preferable to S_1 (because for $x_{12} < x \leq x^2$, $S_2(x)/x$ is greater than $S_1(x^1)/x^1$; that is, the slope of a line from the origin to the point $(x, S_2(x))$ is greater than the slope of the tangenu of S_1).

The preceding procedure could be applied "off line" to generate tabular information, to be loaded in a terminal's memory, so that it can perform link layer re-configurations following simple table look-ups. Developing such tables is a possible avenue of further research (the exact and general closed-form expression for the BER for PAM and QAM modulations, with Gray code bit mapping, over an AWGN channel provided by [11] could be useful in such pursuit). Likewise, since we have focused on data traffic, the consideration of media traffic, such as voice or video, within our framework should receive attention in the future. Finally, our results could be embedded in a network model such as [6]'s.

ACKNOWLEDGEMENT

Financial support from the Deutsche Forschungsgemeinschaft (DFG), through the UMIC project, is gratefully acknowledged.

APPENDIX

In the analysis the following definitions are used.

Definition A.1 $S : \Re_+ \to [0, Y]$, is an S-curve with unique inflexion at x_f if (i) S(0) = 0, S is (ii) continuously differentiable, (iii) strictly increasing, (iv) convex over $[0, x_f)$ and concave over (x_f, ∞) , and (v) surjective (see Fig. 2).

Remark A.1 In Definition A.1, S is strictly increasing and also surjective (for each $y \in [0, Y]$ there is an $x \in \Re_+$ such that S(x) = y). Therefore, S must approach Y asymptotically as x goes to infinity (i.e., this follows from the definition).

Definition A.2 A function $\mathcal{B} : \Re_+ \to [0, Y]$ is a bell curve over \Re_+ if \mathcal{B} is continuous, surjective and has a global maximum at $X \in \Re_+$ such that $\mathcal{B}(X) = Y$, $0 \le x_1 < x_2 \le X \implies \mathcal{B}(x_2) > \mathcal{B}(x_1)$ and $X \le x_1 < x_2 \implies \mathcal{B}(x_2) < \mathcal{B}(x_1)$.

Remark A.2 In Definition A.2, \mathcal{B} is strictly increasing up to x = X and strictly decreasing thereafter. Since \mathcal{B} is also surjective, it must approach 0 asymptotically as x goes to infinity.

REFERENCES

- W. Webb and R. Steele, "Variable rate QAM for mobile radio," Communications, IEEE Transactions on, vol. 43, pp. 2223–2230, Jul 1995.
- [2] M. Pursley and J. Shea, "Adaptive nonuniform phase-shift-key modulation for multimedia traffic in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 18, pp. 1394–1407, Aug 2000.
- [3] S. T. Chung and A. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *Communications, IEEE Transactions on*, vol. 49, pp. 1561–1571, Sep 2001.
- [4] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *Wireless Communications, IEEE Transactions on*, vol. 4, pp. 2349–2360, Sept. 2005.
- [5] T. Yoo, R. J. Lavery, A. Goldsmith, and D. J. Goodman, "Throughput optimization using adaptive techniques." Available: http://wsl.stanford. edu/Publications/Taesang/, 2006.
- [6] D. O'Neill, A. Goldsmith, and S. Boyd, "Optimizing adaptive modulation in wireless networks via utility maximization," in *Communications. IEEE International Conference on*, pp. 3372–3377, May 2008.
- [7] V. Rodriguez, "Robust modeling and analysis for wireless data resource management," in *IEEE Wireless Communications and Networking Conf.*, vol. 2, pp. 717–722, 2003.
- [8] V. Rodriguez, "An analytical foundation for resource management in wireless communication," in *Global Telecommunications Conf., IEEE*, vol. 2, pp. 898–902 Vol.2, Dec. 2003.
- [9] B. Sklar, "A structured overview of digital communications-a tutorial review-part i," *Communications Magazine, IEEE*, vol. 21, pp. 4–17, Aug 1983.
- [10] J. Mitola, III, "Software radios: Survey, critical evaluation and future directions," *IEEE Aerospace and Electronic Systems Magazine*, vol. 8, pp. 25–36, Apr 1993.
- [11] K. Cho and D. Yoon, "On the general BER expression of one- and twodimensional amplitude modulations," *Communications, IEEE Transactions on*, vol. 50, pp. 1074–1080, Jul 2002.