An interference-aware perspective on decoding power

Pulkit Grover, Kristen Ann Woyach, Hari Palaiyanur and Anant Sahai
Wireless Foundations, EECS Department, UC Berkeley, CA-94720, USA
{pulkit, kwoyach, hpalaiya, sahai}@eecs.berkeley.edu

***Abstract*— Traditionally, coding has been seen as a way of saving transmit power: capacity-approaching codes require minimal transmitted energy-per-bit given the bandwidth available. But because transmit power is often smaller than decoding power at short distances, many recent wireless system designs continue to use uncoded transmission!**

We first observe that in wireless systems that both generate and face interference, coding serves another purpose (assuming interference is treated as noise): it allows a system to support a higher density of transmitter-receiver pairs. Bringing decoding power into the picture, we propose an approach to investigate which code/decoder to use and whether to use any coding at all. It turns out that the code's gap to capacity determines how high the maximum supportable link density can be when power is plentiful, whereas the code's decoding complexity governs what link densities can be supported at low power.

I. Introduction

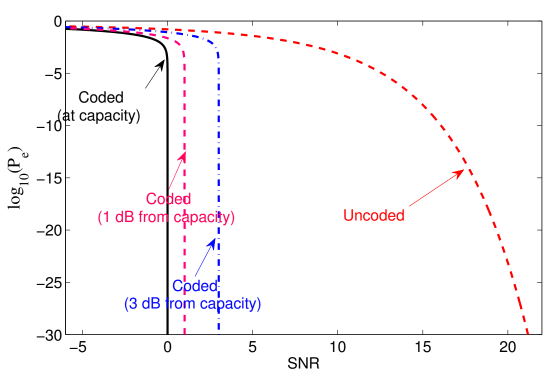

Fig. 1. The traditional Shannon waterfall curve, which provides the minimum required SNR for small bit-error probabilities, predicts a bounded transmit power even as error probability converges to zero. In contrast, uncoded transmission requires power that diverges to infinity.

Shannon theory predicts that when source and channel are perfectly matched, uncoded transmission is optimal. In practice however, the two are almost always mismatched and theory predicts that uncoded transmission requires much higher transmit power. This is illustrated in the traditional Shannon waterfall, shown in Fig 1: for tiny bit-error probabilities, while coded transmissions can have a bounded SNR, the required SNR for uncoded transmission diverges to infinity.

It may therefore seem surprising that for short-distance wireless communication, recent implementations that aim at reducing system power consumption use regular LDPC codes [1] (that are known to operate far from capacity) or even uncoded transmissions (see, for example, [2]). A closer look at these systems exposes another mismatch of sorts — one between the theory and practice of short-distance communication. While coding succeeds in saving transmit power, the power consumed by the decoder itself is often substantially larger than the transmit power. By contrast, the use of uncoded transmission requires more transmit power, but only trivial decoding effort!

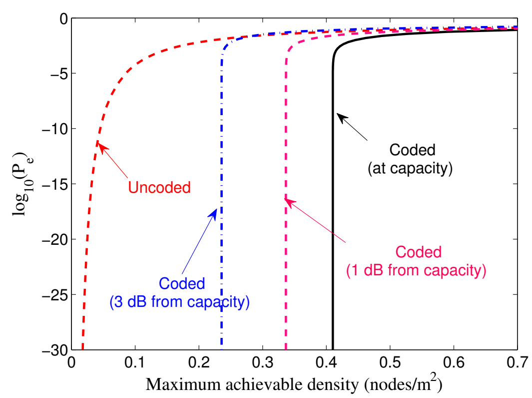

Fig. 2. A plot of achievable link densities with decreasing bit-error probability for a rate of $R = 1$ Gbps, path-loss decay exponent $\alpha = 2.5$, bandwidth $W = 3.5$ GHz, central frequency $f_c = 60$ GHz, distance $r = 1$ meter between the transmitters and their receivers, and angle $\theta = \frac{\pi}{4}$ (see Fig. 3). The plot shows the maximum attainable density with *arbitrarily large* (but equal) transmission powers. The Shannon-waterfall is reflected as another waterfall for coded transmissions, yielding a non-zero node density even as the desired error probability decreases to zero. Similar behavior is demonstrated by codes that operate a few dBs away from capacity. In contrast, because the transmit power for uncoded transmissions must diverge to infinity as $P_e \to 0$, the link density with uncoded transmissions decreases to zero.

Are transmit power and decoding power always exchangeable, as the above argument presupposes[1]? At the face of it, the argument seems reasonable: both are measured in the same unit, Watts, and to minimize the system power drain (to maximize the battery-life), one should minimize the sum of all the powers at the transmitter and the receiver. However, there is a crucial difference between the transmit and decoding powers that shows up in multiterminal situations: transmit power pollutes. Loud transmissions by one link increase the interference faced by other links, thereby reducing their rates. In interference-limited multiterminal systems, increasing the transmit power beyond a certain limit is ineffective: because the transmit power of every user increases, the interference

[1]The question was posed to the authors by Rüdiger Urbanke [3].

increases by the same factor as well — keeping the signal to noise and interference ratio (SINR) almost constant.

The impact of interference on attainable rates has been explored in the medium-access control (MAC) protocol literature using the metric of *area-spectral efficiency*, which is defined as the capacity per unit-area per unit-bandwidth. In a cellular situation, considered in [4], of primary interest is examining how much bandwidth each base-station should use. If the bandwidth used by each user is smaller than the total available bandwidth, then the users can be split into various bands. While this reduces interference, it also requires each link to have a larger SINR to support the same rate. Thus there is an optimal division of the bandwidth to maximize the area-spectral efficiency, and Alouini and Goldsmith characterize this optimal number.

The analysis of Alouini and Goldsmith in [4] focuses on transmit power, which is likely the dominant power in cellular networks. But since decoding power is substantial at short distance, how do we decide which coding scheme should be used? A theoretically guided approach that takes into account decoding power is needed to understand power consumption in the context of wireless ecosystems with interference.

Previous information-theoretic work [5]–[10] has mostly dealt with isolated systems. While the work in [5], [6] uses a "black-box" model for a receiver, our coarse model in [7], [8] uses decoding complexity, measured in the required number of iterations, as a proxy for decoding energy. We assume that each processing element (PE) in a message-passing decoder consumes a fixed amount of energy per iteration. The number of iterations can thus be used to estimate the total decoding energy. Assuming that the decoding throughput is the same as the rate across the channel, this energy immediately yields the decoding power. Extensions of these results to an isolated broadcast network with a single transmitter are obtained in [9]. The PE-centric model is complemented in [10] by an interconnect-centric model that explicitly models the parasitic capacitances in the decoder circuit that consume increasingly large power when the decoding throughput is high.

While the information-theoretic understanding of decoding power developed in [5]–[10] is insightful, experience from classical network information theory problems has shown that it is not easy to extend point-to-point insights to network problems. In particular, the question is to what extent does the core insight of the previous work still hold. To minimize total power consumption, should a code operate at a gap to capacity?

To attack the problem in the simplest possible setting, we consider the case of multiple transmitters sending messages to their respective receivers in which:

- No multihop relaying is allowed (unlike in [11]).
- Nodes do not use cooperative interference management strategies (such as those in [12]) beyond frequency-reuse.
- The aggregate interference is assumed to behave like additive white Gaussian noise.

For this problem, more insightful than the waterfall curve in Fig. 1 is Fig. 2, that plots the maximum *density* of nodes that can be supported for a given error probability. As explained in Section III-A, the waterfall in Fig. 1 translates into a non-zero density of links whose simultaneous operation can be supported using coded transmissions, even in the limit of small bit-error probability. By contrast, the supportable density using uncoded transmissions decreases to zero. This maximum density plot is obtained in the limit of infinite power. Since decoding power, while substantial, causes no interference[2], it has no impact on the maximum link density.

In Section III-B, decoding power becomes a consideration. For simplicity, we assume that the designer has only two PHY-options (beyond partitioning the bandwidth): uncoded QPSK, or a code with a specified decoding architecture such as that implemented in [1]. Within this context, the importance of the twin goals of coding theory show up naturally: a code's *gap from capacity* limits the maximum link density that can be supported, while high *decoding complexity* (and consequently high decoding power) prevents the code from supporting good densities at small total power. Furthermore, even the coarse bounds used here show that when the target link density is not maximal, it is best to operate codes away from capacity to minimize total power consumption.

II. System model and notation

The system is assumed to be a collection of point-to-point links with non-cooperating nodes. Each node transmits at the same power P_T to its receiver located at distance r from the transmitter. This distance is assumed to be fixed, and *does not scale*[3] with the density of links. The communication rate R (bits per second) and the desired bit-error probability are assumed to be fixed, and equal for all links. For simplicity, the transmitters are assumed to lie on a triangular grid, as shown in Fig. 3. Nearest nodes are separated by a distance d.

We assume a power-law propagation model and so the received power at distance x meters is given by

$$P_{rec}(P_T, x) = \min\left\{P_T, \frac{P_T \lambda^\alpha}{x^\alpha}\right\} \quad (1)$$

where $\lambda = \frac{c}{f_c}$ is the wavelength of transmission, $c = 3 \times 10^8$ meters per second is the speed of light, f_c is the center frequency in Hertz, and α is the path-loss coefficient, which is larger than 2 in practical situations[4]. More critically, we assume that the link designer treats aggregate interference as Gaussian noise. Although smarter interference management techniques can be used, we do not consider them in this paper. The thermal noise power observed by a receiver operating in bandwidth W_{used} is given by $\frac{kTW_{used}}{2}$ in each dimension (real and imaginary), where T is the temperature. A given distance d between transmitting nodes immediately translates into the

[2]In the 60-GHz band, because the operating wavelength is small (less than a centimeter), global interconnects in a decoding chip could act as radiating antennas. Though the impact of the resulting interference may be worthy of consideration, it is ignored in this paper.

[3]This means that we allow for links to "cross" in the sense that there can be potentially interfering transmitters in the middle of a given transmitter-receiver pair. To see how this could arise in a practical context, consider a smart phone communicating with a farther laptop while a nearby hands-free headset is communicating to its own smart phone.

[4]This also rules out that unrealistic possibility of infinite aggregate interference at finite link transmit powers, which is a mathematical consequence of using $\alpha = 2$ in large networks.

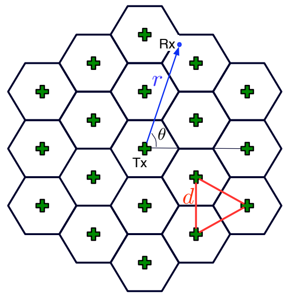

Fig. 3. We consider spatial networks where the node placement is deterministic, and the points lie on a triangular grid (efficient packing). The node density is calculated as follows: the area of an equilateral triangle is $\frac{\sqrt{3}}{4}d^2$, and each triangle contains 0.5 nodes (3 nodes, each shared with 6 other triangles). This gives a density of $\frac{2}{\sqrt{3}d^2}$.

following node densities (as explained in Fig. 3)

$$\rho_{tri} = \frac{2}{\sqrt{3}d^2}. \tag{2}$$

Let $x(i) = id - r\cos\theta$, and $y(j) = j\frac{\sqrt{3}}{2}d - r\sin\theta$ for integers i and j, where θ is as shown in Fig. 3. Then the total aggregate interference is given by[5]

$$I(P_T, \lambda, r, d) = \sum_{\substack{i,j=-\infty,\\ (i,j)\neq(0,0)\\ j \text{ even}}}^{\infty} P_{rec}\left(P_T, \sqrt{x(i)^2 + y(j)^2}\right) + \sum_{\substack{i,j=-\infty,\\ (i,j)\neq(0,0)\\ j \text{ odd}}}^{\infty} P_{rec}\left(P_T, \sqrt{\left(\frac{d}{2} + x(i)\right)^2 + y(j)^2}\right). \tag{3}$$

Following the work of Alouini and Goldsmith [4], we allow both coded and uncoded transmissions to split the band into multiple sub-bands (of equal bandwidth) in order to reduce co-channel interference while keeping overall density high. Each user uses one of the sub-bands for transmission. Rather than dealing with the potentially irregular spatial patterns that can arise from frequency reuse, we simplify and pretend that links in the multiple bands are noninteracting worlds that are assumed to have the same grid structure. The distance d is redefined to be the distance between the nearest transmitters transmitting *in the same band*.

For computing decoding power, we assume a streaming application where the decoding throughput (measured in bits per second) is equal to the rate across the channel. To obtain decoding power from decoding throughput, we assume that the designers of the decoder supply us with decoding power as a function of decoding throughput, R, P_e, and SINR.

For numerical evaluation, we assume $f_c = 60$ GHz, $W = 3.5$ GHz, the rate $R = 1$ Gbps, $P_e = 10^{-12}$, $r = 1$ meter, $\theta = \frac{\pi}{4}$ and $\alpha = 2.5$.

[5] Aggregate interference does not have a closed-form expression here, unlike in [13], because we cap the maximum received power in (1) and consider fixed non-zero distances between a transmitter and its receiver.

III. The approach

Attained density for uncoded transmission

For uncoded transmission, we assume that the transmission is using quadrature phase-shift keying (QPSK). The bandwidth occupied by the transmissions is assumed to be equal to the data-rate (assuming ideal sinc pulse-shapes).

If the total available bandwidth is larger than the data rate, the users will obviously split the entire band amongst themselves to reduce interference. The number of sub-bands therefore equals $2W/R$. Allowing for this frequency reuse, we obtain the following expression for error probability

$$P_e = \mathbb{Q}\left(\sqrt{\frac{P_{rec}(P_T, r)}{kTR + I(P_T, \lambda, r, d)}}\right), \tag{4}$$

where $I(P_T, \lambda, r, d)$ is the interference function given by (3). For a fixed P_e, one can now calculate the required distance d between nodes and the resulting maximum link density ρ that can still support rate R.

Attained density for coded transmissions

We assume that the entire band is split into B equal-sized sub-bands. The maximum allowed interference is now implicitly given by the inequality

$$\frac{W}{B}\log_2\left(1 + \frac{P_{rec}(P_T, r)}{kT\frac{W}{B} + I(P_T, \lambda, r, d)}\right) \geq R(1 - h_b(P_e)), \tag{5}$$

where $I(P_T, \lambda, r, d)$ is the interference function given by (3). For fixed P_e and R, the distance d and density ρ can again be calculated. Notice that there is freedom in the choice of B since the code could target arbitrary rates. This freedom does not exist for uncoded transmission, where the bandwidth per link and the rate are intimately tied.

A. Maximum attainable density at infinite power

We first find the asymptotic density in the limit of infinite power. Because decoding power does not pollute, it can be ignored. Fig. 2 reflects the waterfall curve to compare the maximum achievable link density using coded and uncoded transmissions for a fixed rate. Unlike the behavior for uncoded transmissions, the supportable density does not vanish to zero for coded transmissions as error probability decreases to zero. This means expending decoding power is required to support higher link densities.

B. Attainable density at finite total power

In practice, the available total power per link is finite. Coding thus needs to be penalized for using decoding power. In particular, at the low densities that are achievable using uncoded transmissions, there is the possibility that uncoded transmissions could use less total power despite needing more transmit power. Even at densities that require coding, should we be using capacity achieving codes?

Answering such questions requires a model for decoding power. Consider an actual decoder implementation like the one in [1]. This decoder is designed for decoding a regular

LDPC code of rate 0.8125 and block-length 2048 at extremely high throughputs of $6-50$ Gbps attaining an error probability of 10^{-12} at an SNR of 5.5 dB. Since our interest is not in this particular code/decoder but rather in codes/decoders "like this," we need to extrapolate.

In particular, to get the power-consumption at our desired rate of 1 Gbps, it is natural to extrapolate the decoding power curves in [1, Fig. 15]. A quadratic extrapolation using the five points nearest to our desired rate yields the curve shown in Fig. 4. Notice that the curve does not go to zero even as the throughput approaches zero. This means that the decoder has some "static power" (approximately 43 milliwatts) that is present regardless of the decoding throughput.

With this plausible understanding of power consumption for this decoder at our desired rates, we plot the performance of this code/decoder pair in Fig. 5. As expected, the supportable density is higher than uncoded transmission at large total power, even though the code operates at a gap from capacity. This is not an artifact of the low error probability used: even at bit-error probabilities of 10^{-3}, this code performs better than uncoded transmissions. After all, below 43 milliwatts, uncoded transmission is the only game in town since that is the static power consumption of this decoder.

So how much could we gain by changing the code? We could certainly improve on the maximum attainable density by building codes that approach capacity. The challenge is that decoding power depends on both the code *and* the decoding architecture. To overcome this challenge, we take a semi-empirical approach: theoretical bounds from [7], [8] are used to lower-bound the number of decoding iterations required for *any* code, but these iterations are translated into power consumption by extrapolating from the empirical performance of the decoder in [1]. This extrapolation is based on the fact that the decoder in [1] runs for 8 iterations. If a hypothetical decoder only had to run for 4 iterations, it could run at half the clock-speed, and we assume this will consume power corresponding to a full 8 iterations at half the throughput[6].

In an outer bound, it seems unfair to bound the performance of the decoder with static power. Hence we assume that the static power is zero, giving the lower dashed curve in Fig. 4. However, we do assume that the decoder must run for at least one iteration if this is to be anything other than uncoded transmission. As you can see from the plot shown in Fig. 5, this implies a minimum decoder power-consumption of about 1 milliwatt, which is about 1 picojoule per decoded bit since the rate here is 1 Gbps. Below this power consumption, uncoded transmission is the only choice.

A more important observation, illustrated in Fig. 6 is that we do not operate the codes at capacity to optimize the link density given a total power constraint. To keep the decoding power low, the code needs to operate at a finite gap from capacity. This is a consequence of the bounds in [7] that show that the decoding complexity must diverge to infinity as the gap from capacity decreases to zero.

[6]So implicitly, we are assuming the same blocklength of 2048, but we will not dwell on this fact. After all, at finite block-lengths, the finite probability of error implies a certain gap from the asymptotic capacity due to the "dispersion" term [14]. However, our lower bounds on iterations from [7], [8] give the decoder the benefit of assuming that the block length is infinite.

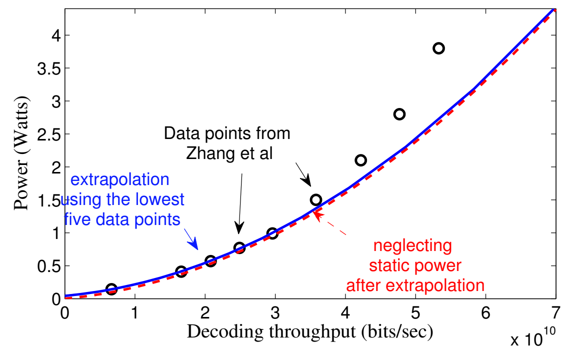

Fig. 4. Extrapolation of the required decoding power curve from [1] to lower throughput values. Since we focus on lower throughputs (no more than 7 Gbps), we perform curve-fitting based on the 5 smallest values of throughput. The achievable density in Fig. 5 by an actual decoder is estimated using this curve. To obtain a realistic outer bound, we use the power values from this plot, force the static power to zero (red dashed curve), and combine them with our lower bounds on complexity from [7], [8].

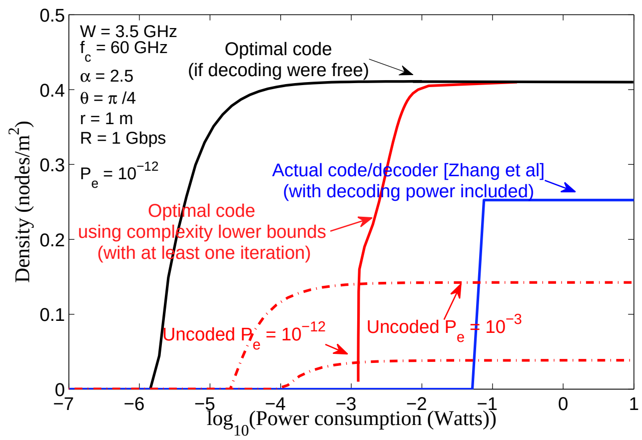

Fig. 5. Comparison of achievable link density versus total power for a rate of 1 Gbps. Because the required SINR in the code-decoder of [1] is rather high (5.5 dB for a rate of 0.8125 bits/channel use), there is a substantial gap from the optimal density even when power is plentiful. Also plotted is a semi-empirical upper bound that is based on the complexity bounds of [7], [8] on message-passing decoding.

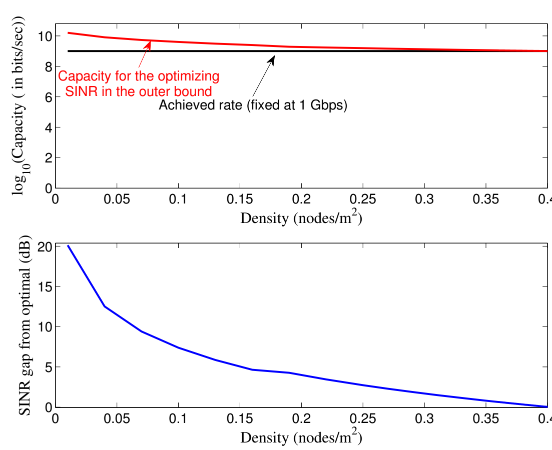

Fig. 6. A plot of capacity corresponding to optimizing SINR in the optimal code performance bound using complexity lower bounds in Fig. 5. A finite gap from capacity is required in order to tame the decoding power.

IV. Discussions and conclusions

The challenge is pictorially captured in Fig. 5. While improvements in codes to make them capacity approaching will bring the high-power performance of links closer to optimal, low-complexity designs are needed to beat uncoded transmission at low power. As is well understood (for message-passing in general [7], [8], [15], as well as for special code families [16], [17]), there are tradeoffs between the two, and obtaining improved bounds on this tradeoff is a challenge for information and coding theorists.

Towards obtaining such bounds, in [7], [8], [10], we allowed unlimited computational ability for each processing element (PE), and allowed the passing of arbitrarily long messages. It appears that capturing the effects of message-quantization is hard in general. To see why, notice that at zero rate (repetition coding with message-passing decoding), the problem is a distributed decision-making problem. While a good deal is known about optimal decoders in this setting [18], evaluation of bounds to the error probability is nontrivial.

For example, consider 1-bit messages being passed in a decoder where the PEs are connected to α other PEs. After L iterations, the PEs reached in the last iteration (assuming no cycles in the decoding graph) can influence the decision only through a bottleneck of α total bits. Further, each of these α bits can only contain information from one of α subgroups of PEs as shown in Fig. 7. If $\alpha = 2$, the size of each subgroup is $n = 500$ and the cross-over probability in observing each bit is $p = 0.1$, the approximately optimal threshold for both groups of PEs is 210 (*not* the symmetric threshold of 250). Surprisingly, the error probability for this asymmetric rule is about 35% better in the error-exponent sense than a strategy where each subgroup makes a (symmetric) majority-logic decision on the bit. Such asymmetries in optimal strategies make finding general lower bounds somewhat perilous (see [19, Ex. 1 and correction] for an example of a distributed detection problem that was thought to have a symmetric optimal solution, but in retrospect is not clear).

Finally, while coding offers gains over uncoded transmission in the achievable density of transmitters, signal-processing techniques, like beam-forming, can offer similar gains. In particular, interference-alignment techniques [12] can effectively provide interference-free transmission in about half the bandwidth for each receiver at the cost of some power attenuation. This offers, at least theoretically, the prospect of increasing the link density as the total power increases. Coding over such signal-processing techniques should further increase the gains in density, and that impact needs to be investigated further.

Acknowledgments

We thank Elad Alon, François Baccelli, Bora Nikolić, Jan Rabaey and Matthew Weiner for helpful discussions, and Rüdiger Urbanke for asking a question that prompted this investigation. Support of NSF grants CCF-0729122, CCF-0917212 and CNS- 0932410 is gratefully acknowledged.

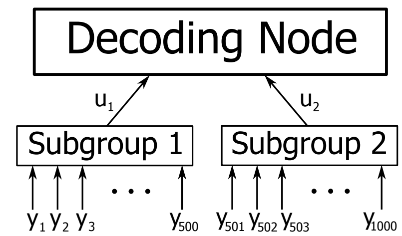

Fig. 7. A bit to be decoded, B, is observed through noise multiple times $Y_i = B \oplus Z_i$ where Z_i are IID Bernoulli random variables with parameter $p = 0.1$. Within each subgroup, 500 of these observations are available, but only one bit (u_1 and u_2) can be passed to the decoding node from each subgroup. The decoding node wishes to minimize the probability of error under equally likely probabilities for the bit B.

References

[1] Z. Zhang, V. Anantharam, M. Wainwright, and B. Nikolic, “An efficient 10GBASE-T Ethernet LDPC decoder design with low error floors,” *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 4, pp. 843 –855, Apr. 2010.
[2] C. Marcu, D. Chowdhury, C. Thakkar, J.-D. Park, L.-K. Kong, M. Tabesh, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. Zamani, E. Alon, and A. Niknejad, “A 90 nm cmos low-power 60 GHz transceiver with integrated baseband circuitry,” *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 12, pp. 3434 –3447, dec. 2009.
[3] R. Urbanke, personal communication, Jan. 2009.
[4] M-S Alouini and AJ Goldsmith, “Area spectral efficiency of cellular mobile radio systems,” *IEEE Transactions On Vehicular Technology*, vol. 48, no. 4, 1999.
[5] P. Massaad, M. Medard, and L. Zheng, “Impact of Processing Energy on the Capacity of Wireless Channels,” in *International Symposium on Information Theory and its Applications (ISITA)*, 2004.
[6] ——, “On the capacity of multiple-access channels with processing power,” in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2004.
[7] A. Sahai and P. Grover, “The price of certainty : “waterslide curves” and the gap to capacity,” Dec. 2007. [Online]. Available: http://arXiv.org/abs/0801.0352v1
[8] P. Grover and A. Sahai, “Green codes: Energy-efficient short-range communication,” in *Proceedings of the 2008 IEEE Symposium on Information Theory*, Toronto, Canada, Jul. 2008.
[9] ——, “Time-division multiplexing for green broadcasting,” in *Proceedings of the 2009 IEEE Symposium on Information Theory*, Seoul, Korea, Jul. 2009.
[10] P. Grover, H. Palaiyanur, and A. Sahai, “Information-theoretic tradeoffs on throughput and chip power consumption for decoding error-correcting codes,” in *Proceedings of the 2010 IEEE Symposium on Information Theory*, Austin, TX, USA, Jun. 2010.
[11] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 388 –404, mar 2000.
[12] V. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the K-user interference channel,” *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3425–3441, aug 2008.
[13] M. Haenggi and R. K. Ganti, *Interference in Large Wireless Networks*. Hanover, MA: Foundations and Trends in Communication and Information theory, NOW Publishing, 2008, vol. 3, no. 8.
[14] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Dispersion of Gaussian channels,” in *IEEE International Symposium on Information Theory*, Seoul, Korea, 2009.
[15] P. Grover, “Bounds on the tradeoff between rate and complexity for sparse-graph codes,” in *2007 IEEE Information Theory Workshop (ITW)*, Lake Tahoe, CA, 2007.
[16] A. Khandekar and R. McEliece, “On the complexity of reliable communication on the erasure channel,” in *IEEE International Symposium on Information Theory*, 2001.
[17] I. Sason and G. Wiechman, “Bounds on the number of iterations for turbo-like ensembles over the binary erasure channel,” *IEEE Trans. Inf. Theor.*, vol. 55, no. 6, pp. 2602–2617, 2009.
[18] J. N. Tsitsiklis, “Decentralized detection,” *Advances in Statistical Signal Processing*, pp. 297–344, 1993.
[19] ——, “Decentralized detection with a large number of sensors,” *Mathematics of Control, Signals, and Systems*, pp. 167–182, 1988.