

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Chip-to-Chip and On-Chip Communications

---

Josef A. Nossek, Peter Russer, Tobias Noll, Amine Mezghani,  
Michel T. Ivrlač, Matthias Korb, Farooq Mukhtar,  
Hristomir Yordanov, Johannes A. Russer

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52981>

---

## 1. Introduction

In high-performance integrated circuits manufactured in CMOS deep sub-micron technology, the speed of global information exchange on the chip has developed into a bottleneck, that limits the effective information processing speed. This is caused by standard on-chip communication based on multi-conductor interconnects, e.g., implemented as parallel interconnect buses. The supported clock frequency of such wired interconnects - at best - remains constant under scaling, but - for global interconnects - reduces by a factor of four, as the structure size is reduced by half. Such multi-conductor interconnects also exhibit some undesirable properties when used for chip-to-chip communication. The much larger distances that have to be bridged, force the clock frequencies for the chip-to-chip interconnects to much lower values than those for on-chip circuitry. In widening up this bottleneck by increasing the number of parallel wires, the separation between the wires has to decrease. This causes increased mutual coupling between neighboring wires, which reduces the supported clock frequency and counters the effect of having more wires in the first place.

The high clock frequencies used in on-chip interconnects and the huge information rate of chip-to-chip communication lets possible solutions belong to the domain of ultra-wideband (UWB) technology. Pursuing suitable solutions, we explore firstly the improvement of the multi-conductor interconnect by signal processing and coding. From information theory, it is known that information can be transmitted through a noisy channel with arbitrary low probability of error as long as the rate is lower than the channel capacity given by the Shannon theorem. Achieving this capacity requires, however, sophisticated digital signal processing and coding. In particular, the DAC (Digital-to-analog converters) and the ADC (analog-to-digital converter) components which are formed by the output or the input of a logic CMOS inverter, respectively, turns to be a limiting factor. In fact, the ADC and DAC components, perform a single-bit conversion between the analog and the digital domain. With such coarse quantization, all state of the art techniques for signal processing fail. We provide information theoretic bounds on the improvements possible by coding the transmission, and propose methods to design suitable codes which allow decoding with low latency.

Thereby, an analytical field-theoretical modeling of multi-conductor interconnects is needed. Moreover, modifications to standard signal processing techniques which make them suitable for medium-low resolution quantization are developed and analyzed, and their performance is studied.

As a promising alternative solution, wireless Ultrawideband (UWB) enables high speed communication at short distances. In fact, it is anticipated that even higher performance is achievable in chip-to-chip and on-chip communication, when multi-conductor interconnects are replaced by wireless ultra-wideband multi-antenna interconnects. Hereby, the signal pulses do not necessarily increasingly disperse as they travel along their way to the receiving end of the interconnect. The propagating nature of the wireless interconnect, the extreme high available bandwidth and the very short distances can offer a much more attractive channel for chip-to-chip and on-chip communications. In addition, applying multiple antennas at the transmitter side as well as the receiver side can drastically improve the data rate and the reliability of UWB systems at the cost of certain computational complexity. This chapter provides theoretical and empirical foundations for the application of ultra-wideband multi-antenna wireless interconnects for chip-to-chip communication. Appropriate structures for integrated ultra-wideband antennas shall be developed, their properties theoretically analyzed and verified against measurements performed on manufactured prototypes. Qualified coding and signal processing techniques, which aim at efficient use of available resources of bandwidth, power, and chip area shall be developed. Since Analog-to-Digital Converters (ADCs) are considered critical components for the UWB, main focus is hereby given to low resolution signal quantization and processing. Therefore, the analysis and the design of UWB systems with low resolution signal quantization (less than 4 bits) is a vital part of this chapter, where optimized receive and transmit strategies are obtained.

On the other hand, detailed cost-models for the digital hardware architecture, which are based on signal flow charts and VLSI implementations of dedicated functional blocks shall be developed, which allow for an informative analysis of elementary trade-offs between computational speed, required chip area, and power consumption. In fact, quantitative optimization in terms of silicon area (manufacturing costs) and even more important in terms of energy dissipation (usage costs) is mandatory already in the standardization and conception phase of digital systems to be highly integrated as System-on-Chips (SoC). This is especially true for digital communication systems where e.g. in the optimization of channel coding traditionally only the transmission power has been considered. In general this leads to highly complex and energy intensive receivers. Actually a proper optimization of such systems requires a joint optimization of the transmitter and receiver cost features, e.g. the minimization of the total energy per transmitted bit. For such a quantitative optimization quite accurate cost models for the components of the transmitter and receiver are required. Instead, if any, only oversimplified cost models are applied today. While quite accurate cost models are available for many communication system components there is a lack of such models for channel decoders like Viterbi, Turbo, and Low-Density-Parity-Check (LDPC) decoders. Out of these, especially the derivation of sufficiently accurate cost models for LDPC decoders is challenging: The realization of the extensive internal exchange of messages between the so-called bit and check nodes in such a decoder results in non-linear dependencies between decoder features and code parameters. For example in high-throughput decoders the data exchange is performed via a complex dedicated interconnect structure. Its realization frequently requires an artificial expansion

of silicon area. In the past various decoder architectures have been proposed to reduce the interconnect impact and trading throughput for silicon area and energy. All that together makes the derivation of LDPC decoder cost models a challenging task.

The Chapter is organized as follows. Radio frequency engineering aspects involved in wired and wireless interconnects are investigated first. There is a multitude of requirements for chip-to-chip communication, which an integrated antenna has to fulfill, like large bandwidth, small geometrical profile, and so on. Therefore, a detailed study of the possible solutions for an integrated on-chip antenna is performed. Novel solutions, which make use of the digital circuit's ground plane as a radiating element, are investigated. In the third section, the signal processing and coding aspects involved are carried out based on the obtained channel models, where both multiconductor interconnects and wireless multiantenna interconnects are interpreted as discrete-time, multi-input-multi-output (MIMO) systems. In the last section of this chapter appropriate silicon area, timing, and energy cost models for high-throughput LDPC decoders, which reproduce accurately the non-linear dependencies and being applicable to bit-parallel as well as to bit-serial decoder architectures are presented. These models allow for a quantitative comparison of different decoder architectures revealing the most area and energy efficient architecture for a given code and throughput specification. Additionally, a new highly area and energy efficient architecture based on a bit-serial interconnect is derived. This architecture is the result of a systematic architecture search and proper optimization based on the cost models.

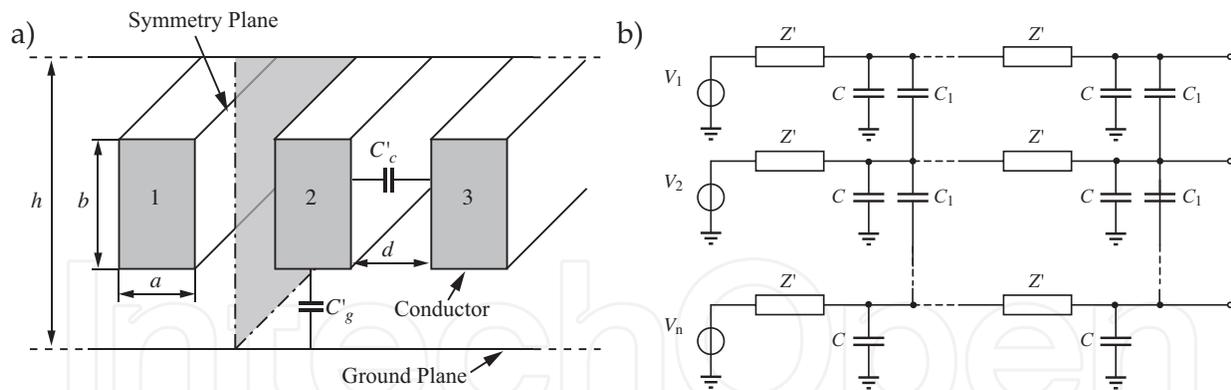
## 2. Multi-conductor interconnects and on-chip antennas

### 2.1. Multi-conductor interconnects

With the increase of the on-chip data transfer rate to several 10 Gbit/s the spatio-temporal intersymbol interference (auto-interference) within the multiwired bus systems becomes a limiting factor for the circuit performance. Due to the limited available space for the bus systems shielding between the wires of the bus should be omitted. This allows for larger wire cross sections and thereby to reduce the signal distortion. An appropriate signal coding and signal processing will compensate for the effects of the coupling between the wires.

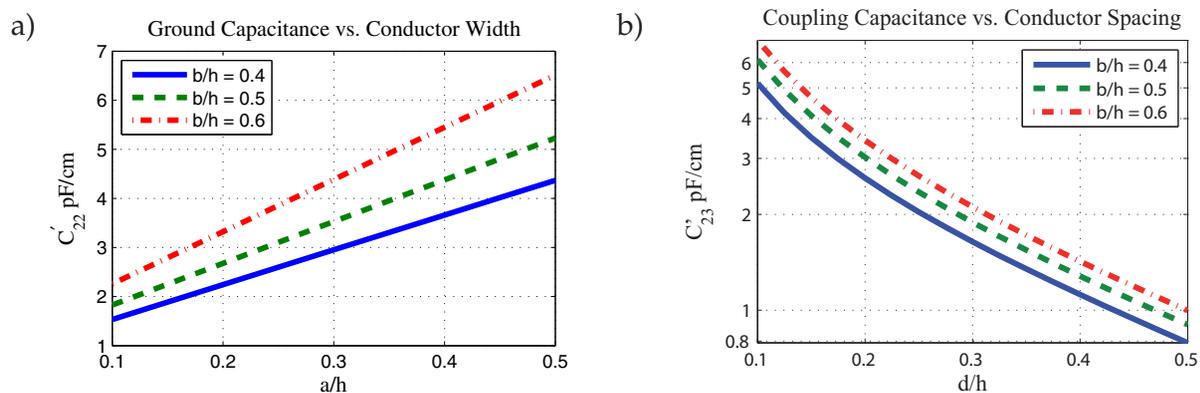
The wiring inside high speed MOS circuits exhibits sub-micron cross-sectional dimensions and conductor width and conductor thickness are of similar size. Within the signal frequency band the cross sectional dimensions are in the order of the skin effect penetration depth. The signal transmission properties of the bus system is detoured by the capacitances per unit of length and the resistance per unit of length.

The TEM modes of a lossless multiconductor transmission line with equidistant conductors of equal cross section and filled with homogenous isotropic dielectric material used for bus have been discussed in [18]. Figure 1 a shows the cross-sectional drawing of the bus. The quasi-electrostatic parameters of the bus embedded in the substrate between two ground planes have been computed. The bus capacitance per unit of length – matrix, describing capacitances with respect to ground and mutual capacitances, has been derived from the conductor geometry using an analytical technique based on even-odd mode analysis [10, 18]. The analytical technique is based on the inversion of the Schwarz-Christoffel conformal mapping [5, pp. 191–201]. The advantages of the proposed method are its accuracy, the lack of geometrical limitations and the algorithm efficiency. The results for the ground and coupling



**Figure 1.** a) A cross section of three-wire digital bus with a coupling and a ground capacitance [10] and b) equivalent lumped element circuit [8].

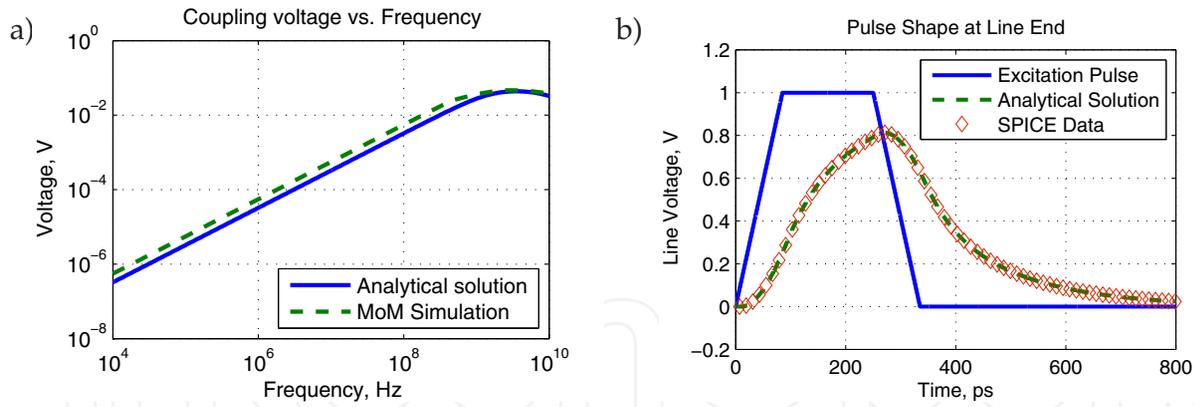
capacitances per unit length for the multi conductor transmission line, filled in with silicon, are presented in Fig. 2. Since the capacitance depends only on the ratio of the line dimensions, all geometrical data are normalized to the distance  $h$  between the ground planes.



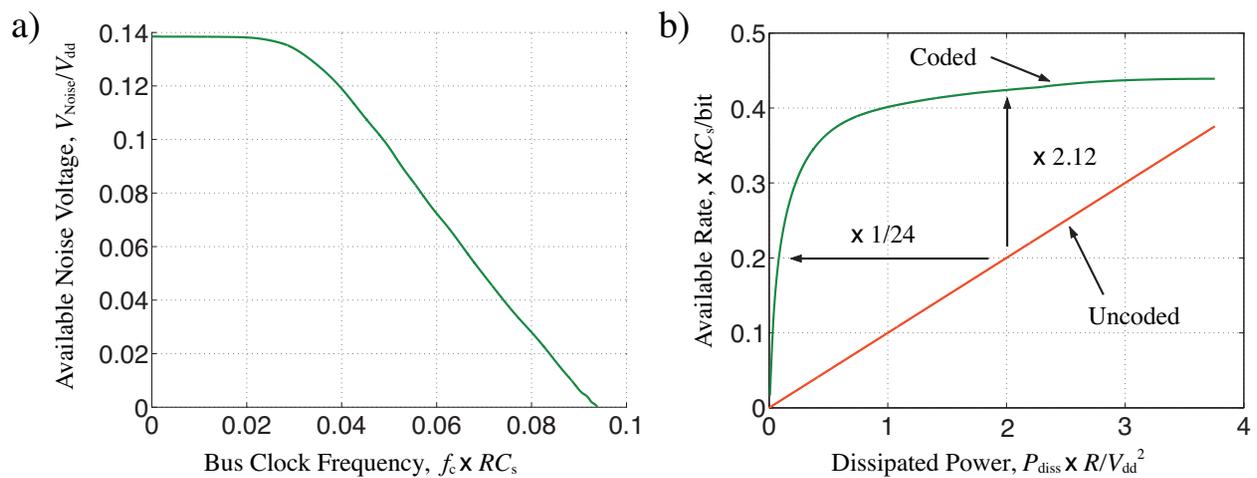
**Figure 2.** a) Ground capacitance vs. geometry for digital transmission line, filled in with Si,  $d/h = 0.125$ , b) Coupling capacitance vs. geometry for digital transmission line, filled in with Si,  $a/h = 0.25$  [10].

The obtained results have been used to compute the transmission line parameters of the bus [7, 8, 10, 18]. The bus model is based on multiconductor TEM transmission line theory [5, pp. 356–363]. In case of the TEM transmission line the inductance per unit of length matrix follows directly from the capacitance per unit of length matrix and the material [18]. In case of small conductor cross sections the resistance per unit of length becomes such high that the inductance per unit of length matrix can be neglected in comparison with the resistance per unit of length. In this case the impedance per unit of length matrix becomes diagonal [8]. The ohmic losses in the conductors are modeled by resistance per unit length  $R'$ . These parameters determine the lumped element equivalent circuit of the bus shown in Fig. 1 b).

The crosstalk between the conductors of the bus has been investigated in [18]. Figure 3 a shows the response of the crosstalk voltage at the end of the line. The results, obtained by solving the transmission line equations have been compared with numerical Method of Moment (MoM) full-wave simulations. The analytic model exhibits good accuracy up to frequencies beyond 10 GHz. Figure 3 b shows the pulse distortion at the end of the transmission line. Analytical data computed with the transmission line model have been compared with numerical data obtained from SPICE simulation.



**Figure 3.** a) Frequency response of the crosstalk voltage at the end of the line, b) pulse response at the end of the line [18].



**Figure 4.** a) Maximum permissible noise-voltage at the receiver of a four-conductor bus as function of the bus clock frequency, b) achievable information rates of coded and uncoded transmission as function of the dissipated power [7].

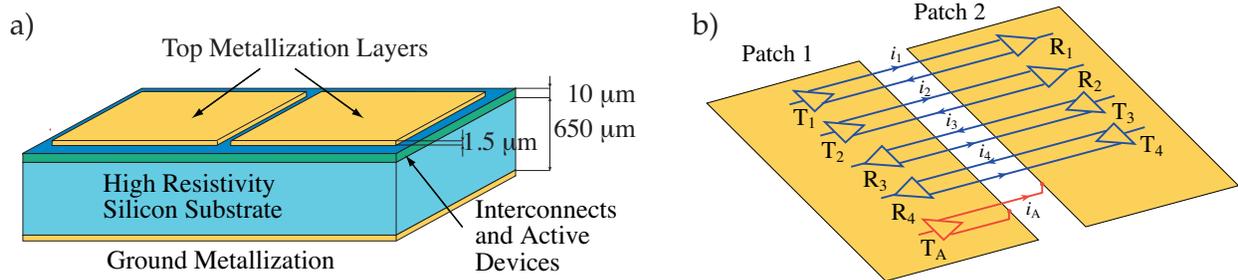
The space-time intersymbol interference present in on-chip interconnection buses is a limiting factor of the performance of digital integrated circuits. This effect has greater influence as the transfer data rate increases and the circuit dimensions decrease. In order to be able to develop coding techniques for reducing the detrimental effects of intersymbol interference, an efficient and precise method for calculating the impulse response of the interconnect is required [31]. In [7] a quasi-analytical method was applied for computing the impulse response of a digital interconnection bus. The fundamental performance limits of bus systems due to information theory have been analyzed. Figure 3 shows the maximum permissible noise-voltage  $V_{\text{Noise}}$  at the receiver of a four-conductor bus as function of the bus clock frequency  $f_c$  and the achievable information rates of coded and uncoded transmission as function of the dissipated power  $P_{\text{diss}}$ . The clock frequency for the coded transmission is set to  $0.11/(RC_s)$ , which is above the cutoff of the uncoded bus and proves to work well with the coded system. Here  $V_{\text{dd}}$ ,  $R$  and  $C_s$  are the magnitude of the signal voltage at the input, the total resistance, and the total average substrate capacitance, respectively. Here,  $C_c = 6C_s$  is assumed, but similar results are obtained for other ratios.

### Conclusion

The developed methods allow to compute the impulse response of the multi-conductor bus, and – building on this ground – to compute information theoretic measures, like mutual information. Those measures allow to quantify the possible gains in performance that can be achieved by employing suitable coding schemes to the multi-conductor interconnection bus. The obtained results reveal a huge potential of coded transmission both in terms of increasing the data rate and in decreasing the dissipated power.

## 2.2. On-chip antennas

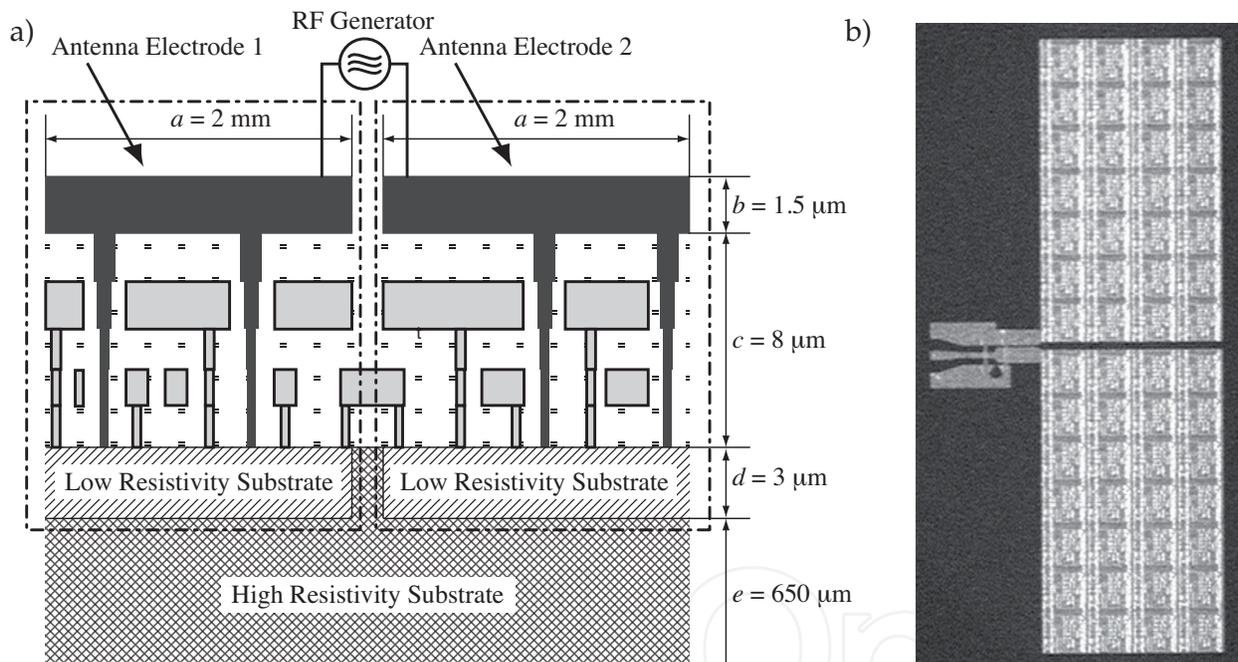
An interesting future possibility for handling Gbit/s data streams on chip and from chip to chip will be wireless intra-chip and inter-chip communication. This section describes investigations of integrated on-chip antennas for broad-band intra-chip and inter-chip communications. At frequencies of 60 GHz and beyond antennas can be made sufficiently small to be integrated on monolithic circuits [1, 19]. However, there are still problems when integrating millimeterwave antennas on CMOS circuits. The integration of millimeterwave antennas on silicon requires a high resistivity substrate in order to achieve low losses, whereas for CMOS circuits the substrate resistivity has to be low in order to provide isolation of the circuit elements. Furthermore, chip surface is a cost factor and should not be wasted for antennas.



**Figure 5.** a) Schematic drawing of a chip with an integrated antenna, b) Differential lines, connecting the digital circuits under the separate antenna patches.

An integrated on-chip antenna for chip-to-chip wireless communication, based on the usage of the digital circuits' ground planes as radiating elements was presented in [12–17]. Figure 5 a shows schematically the realization of this principle in silicon technology. The integrated circuit is fabricated on a high resistivity silicon substrate ( $\geq 1\text{k}\Omega\cdot\text{cm}$ ) with a thickness in the order of of  $650\ \mu\text{m}$ . The substrate is backed by a metallic layer. On top of the substrate a low-resistivity layer ( $\approx 5\Omega\cdot\text{cm}$ ) of few micrometer thickness is grown. A homogeneous low-resistivity layer of  $3\ \mu\text{m}$  to  $5\ \mu\text{m}$  thickness is followed by a top with embedded CMOS circuitry and the interconnects. A low resistivity top layer is required for the circuit insulation. The electromagnetic field of the circuits is mainly confined in this top layer. The antenna field is spreading over the whole thickness of the substrate. Due to the high resistivity of the substrate the antenna losses are low. Since only a small fraction of the antenna near-field energy is stored in the low-resistivity layer, the coupling between the antenna near-field and the circuit field is weak. Furthermore, the interference between the CMOS circuits and the antenna field can be reduced when the main part of the circuit is operating in a frequency band distinct from the frequency band used for the wireless transmission.

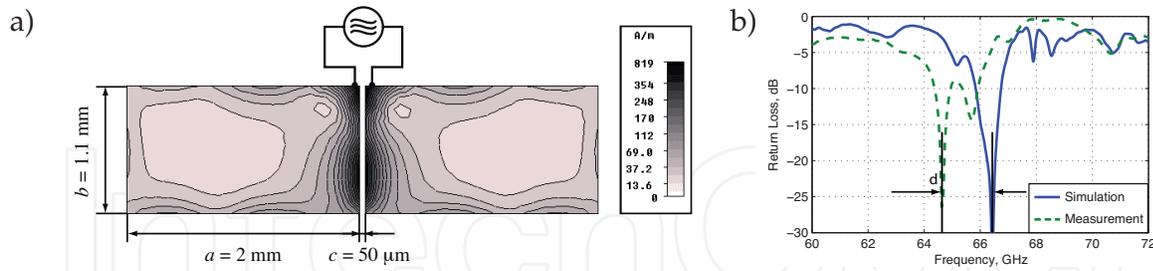
The utilization of the electronic circuit ground planes as radiating elements for the integrated antennas allows for optimal usage of chip area, as the antennas share the chip area with the circuits. It has to be taken care that the interference between the antenna field and the field propagating in the circuit structures stays within tolerable limits. Consider the structure represented schematically in Fig. 5 b. The structure contains two antenna patches 1 and 2. Both antenna patches serve as the ground planes of circuits. These circuits contain line drivers  $T_1 \dots T_4$  driving over symmetrical interconnection lines the line receivers  $R_1 \dots R_4$ . Furthermore there is a driver  $T_A$ , the output of which is connected to both patches, however only one conductor bridges the gap between the patches. The currents  $i_1 \dots i_4$  all are flowing back over the symmetric lines. The sum of the currents  $i_1 \dots i_4$  flowing in both directions through the transmission line modes vanishes and is not exciting the antenna. Different from this, the current  $i_A$  excites an antenna radiation mode. The circuit for this current is closed via the displacement current in the near-field of the antenna. By exciting the interconnection structures in transmission line modes and the antennas in antenna modes the interference between circuit and antennas can be minimized. We need not to use differential lines between the patches. In general a interconnection structure consisting of  $N$  conductors can guide up to  $N - 1$  quasi-TEM transmission line modes and one antenna mode.



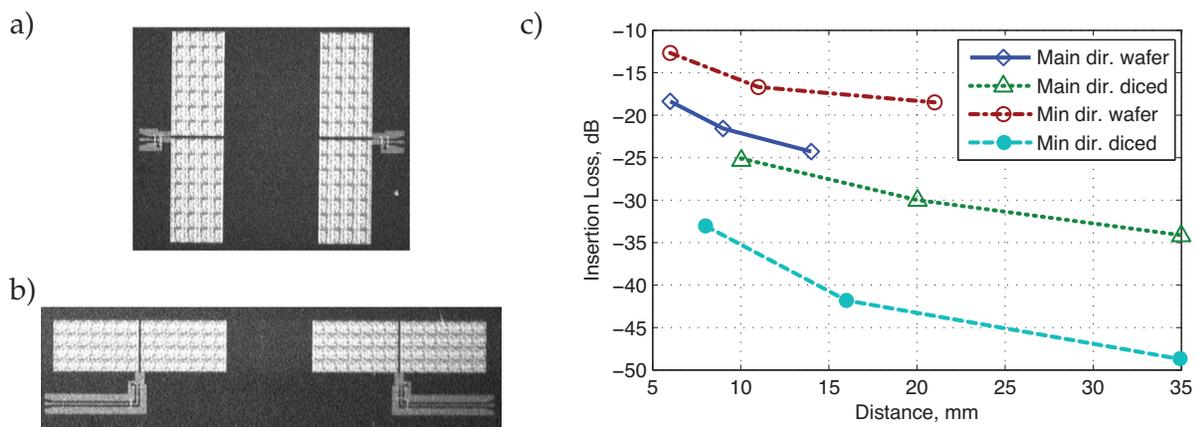
**Figure 6.** Cross-sectional view of integrated on-chip antenna, using the ground planes as antenna electrodes [11, 14].

Figure 6 a shows a cross-sectional view of the integrated on-chip antenna, using the ground planes in layer  $b$  on top of the integrated circuit as the antenna electrodes. Layer  $c$  with a total thickness of  $8 \mu\text{m}$  contains the active and passive circuit elements and the interconnect wiring. The low-resistivity layer  $d$  provides isolation of the circuit elements. The thick substrate layer  $e$  is of high resistivity. Figure 6 b shows a photograph of the fabricated open-circuit slot antenna with CMOS circuits under the antenna electrode [15]. Figure 7 a shows the simulated current distribution of a two-patch V-band antenna [14, 15]. The current distribution in both patches mainly is concentrated in the neighborhood of the slots. The antenna behaves as an open-circuited slot antenna. The guided wavelength is in the range of a millimeter. The

open-circuited slot with a length of about one millimeter is a transmission line resonator with a resonance frequency in the V-band. The standing wave in the slot excites the radiation field.



**Figure 7.** (a) Top view and current distribution of a two-patch dipole antenna, operating at 66 GHz, (b) Measured return loss of the on-chip open slot antenna [14, 15].

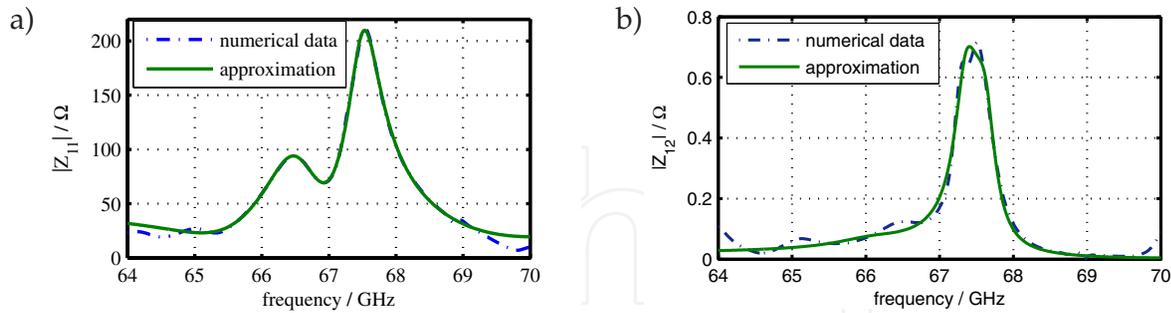


**Figure 8.** Antenna orientation with a) collinear and b) parallel slots, c) Measured insertion loss of a wireless chip-to-chip link depending [14, 15].

The antennas have been measured on-wafer and diced. The measured return loss of the diced slot antenna from Fig. 7 a is compared with the simulation results in Fig. 7 b. The insertion loss of a transmission link has been measured for the two antenna alignments shown in Figs. 8 a and b, where the antennas were positioned in each other radiation minima and maxima. Figure 8 s c shows the measured insertion loss of a wireless links formed by two antennas. When the antennas are oriented such that their slots are collinear, they are in each other's direction of minimum radiation. When they are oriented such that their slots are parallel, they are in each other's direction of maximum radiation. Both cases were investigated for on-wafer and for diced chips. The chip-to-chip links with both antennas on different chips exhibit higher insertion loss. The lower insertion loss of links between antennas on the same chip is due to the contribution of surface waves. The worst-case transmission link (gain-chip-to-chip link) in the direction of minimum radiation shows an insertion loss of -47dB, which is sufficient for high-rate data links.

Lumped element circuit models can provide a compact description of wireless transmission links [20–22, 25]. Distributed circuits can be modeled also in a broad frequency band with arbitrary accuracy using lumped element network models. A general way to establish network models is based on modal analysis and similar techniques [2–5]. In the case of wireless transmission links, high insertion losses have to be considered. Therefore methods for the synthesis of lossy multiports have to be applied. In [23, 24] a lumped-element two-port

antenna model is presented where the antenna near-field is modeled by a reactive two-port and the real resistor  $R_r$  terminating the two-port models the energy dissipation in the far-field.



**Figure 9.** Comparison of the numerical data of (a)  $|Z_{11}|$  and (b)  $|Z_{12}|$  obtained from the full-wave simulation of the wireless transmission link with the data computed from lumped element model

Figure 9 shows the comparison of the numerical data of the magnitudes of the two-port impedance parameters  $|Z_{11}|$  and  $|Z_{12}|$  obtained from the full-wave simulation of the wireless transmission link with the data computed from lumped element model. For details of the model see [3]. The numerical full-wave simulations have been performed using CST software. An accurate model of  $Z_{11}$  is achieved for the frequency band from 65 GHz to 69 GHz based on four pairs of poles and two single poles at zero and infinity. The frequency range may be extended by increasing the number of poles.

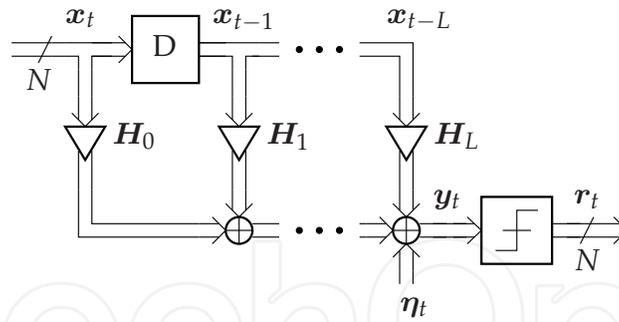
### Conclusion

We have investigated methods for an area-efficient design of on-chip integrated antennas, based on the utilization of the same metallization structures both as a CMOS circuit ground plane and as antenna electrodes. An experimental setup has been designed for validating the computed antenna parameters, as well as the interference between the CMOS interconnects and the antenna. Equivalent circuits have been established to model integrated antennas and wireless intra-chip and inter-chip transmission links.

## 3. Communication theoretical limits, coding and signal processing

Both multiconductor interconnects and wireless multiantenna interconnects can be interpreted as discrete-time, multi-input-multi-output (MIMO) systems. Such systems have been subject to extensive study in the recent past in the field of digital, especially mobile communications. Starting from the analysis of their promising information theoretic capabilities (e.g., [41]), a large amount of signal processing and coding techniques have been developed, that aim at achieving the information theoretic bounds (e.g., [42–44, 46]).

The common approach to handle spatio-temporal interference in MIMO systems, involves either linear or non-linear transmit and receive signal processing, which job is to transform the original MIMO system into a »virtual« MIMO system, where large amounts of spatio-temporal interference have been removed [48]. All state of the art MIMO signal processing techniques have in common that they assume that either the receiver, the transmitter, or both, have access to, or can generate signals with arbitrary precision. This implies, in practice, the existence of ADC and DAC components with a large enough resolution such that the non-linear effects of signal quantization can be neglected. However, in multiconductor, or wireless



**Figure 10.** MIMO channel with ISI and single-bit outputs modeling wired as well as the wireless interconnects.

multiantenna interconnects, used for high-speed on-chip or chip-to-chip communication, such an assumption of having available high-resolution ADC and DAC components, cannot be made.

In case of on-chip multiconductor interconnects, the DAC and the ADC components are formed by the output or the input of a logic CMOS inverter, respectively. Hence, the ADC and DAC components, perform a *single-bit* conversion between the analog and the digital domain. With such coarse quantization, all state of the art techniques for MIMO signal processing fail.

In the case of wireless multiantenna interconnects for chip-to-chip communication the situation can be expected to be slightly better. However, because of the huge bandwidth, the requirements on conversion time are extremely high, such that only moderate resolution (4–5 bits) ADC and DAC components are reasonable. As it turns out, such a moderately high resolution is still too low for reliable operation of state of the art MIMO signal processing.

In this section, we treat the ADC and DAC components as an integral part of the MIMO system. We develop signal processing and coding techniques, which utilize the information theoretic gains of MIMO systems with very-low to moderately-low resolution signal quantization. We first provide suitable design principles for low-latency channel-matched codes applied on general frequency selective MIMO channels, which are based on an information theoretic ground.

### 3.1. Single-bit ADC/DAC: Coding and performance limits

Consider the MIMO channel with inter-symbol interference (ISI) and single-bit output quantization shown in Fig. 10. The channel has a memory of length  $L$  and it is governed by the channel law

$$\mathbf{r}_t = \mathcal{Q}\{\mathbf{y}_t\} = \mathcal{Q}\left\{\sum_{k=0}^L \mathbf{H}_k \mathbf{x}_{t-k} + \boldsymbol{\eta}_t\right\}. \quad (1)$$

Here,  $\mathbf{H}_k \in \mathbb{C}^{N \times N}$  is the  $k$ -th channel matrix.  $\mathbf{x}_k \in \mathcal{X}$ ,  $\boldsymbol{\eta}_k \in \mathbb{C}^N$ ,  $\mathbf{y}_k \in \mathbb{C}^N$  and  $\mathbf{r}_k \in \mathcal{Y} = \{\alpha + j\beta | \alpha, \beta \in \{+1, -1\}\}^N$  denote the channel input vector, the noise vector, the unquantized receive vector and the channel output vector, at the  $k$ -th time instant, respectively. The single-bit quantization operator  $\mathcal{Q}$  returns the sign of the real and imaginary part of each component of the unquantized received signal  $\mathbf{r}_t$ , i.e.,

$$\mathcal{Q}\{\mathbf{y}_t\} = \text{sign}(\text{real}\{\mathbf{y}_t\}) + j \cdot \text{sign}(\text{imag}\{\mathbf{y}_t\}). \quad (2)$$

The conditional probability of the channel output satisfies

$$\Pr(\mathbf{r}_t | \mathbf{x}_{-L}^\infty, \mathbf{r}_1^{t-1}, \mathbf{r}_{t+1}^\infty) = \Pr(\mathbf{r}_t | \mathbf{x}_{t-L}^t), \quad t \geq 0. \quad (3)$$

Here,  $\mathbf{x}_0^\infty$  and  $\mathbf{r}_1^{t-1}$  stand for the sequences  $[\mathbf{x}_{-L}, \mathbf{x}_{-L+1}, \dots]$  and  $[\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{t-1}]$ , respectively. The noise is additive white Gaussian with covariance matrix  $E[\mathbf{n}_t \mathbf{n}_t^H] = \sigma_\eta^2 \mathbf{I}_N$ . The transmit signal energy is normalized to 1, that is  $\|\mathbf{x}_t\|_2 = 1$ . The signal-to-noise ratio is accordingly defined as

$$\text{SNR} = 1/\sigma_\eta^2. \quad (4)$$

The channel transition probabilities can be calculated via

$$\Pr(\mathbf{r}_t | \mathbf{x}_{t-L}^t) = \prod_{c \in \{\mathbb{R}, \mathbb{I}\}} \prod_{i=1}^N \Phi \left( \frac{[\mathbf{y}_t]_{c,i} \cdot [\bar{\mathbf{y}}]_{c,i}}{\sigma_\eta / \sqrt{2}} \right). \quad (5)$$

Here,  $[\mathbf{x}_t]_{\mathbb{R},i}$  ( $[\mathbf{x}_t]_{\mathbb{I},i}$ ) denotes the  $i$ -th real (imaginary) component of the input vector  $\mathbf{x}_t$ ,  $\bar{\mathbf{y}} = \sum_{k=0}^L \mathbf{H}_k \mathbf{x}_{t-k}$  is the noise-free unquantized receive vector and  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$  is the cumulative normal distribution function. The input symbols are modulated using QPSK (or BPSK with 1-bit DAC). Consequently, we consider ISI channels with the input and output cardinality  $|\mathcal{X}| = |\mathcal{R}| = 4^N$ .

### 3.1.1. Code-design

We are looking for codes which maximally increase throughput and that allow fast decoding with good performance. In this way, the code design has to focus on finding low-complexity codes providing good coding gain while having low overhead both with respect to circuit complexity and power dissipation.

Even though linear block and convolution coding schemes are favorable candidates for error correction, they are not able to decrease power consumption and eliminate the residual error floor caused by the crosstalk even in the noiseless case [31]. This is due to their structural properties (linear codes) and the coarse quantization of the channel.

Therefore, the coding schemes which are needed are non-linear and – for having good performance and, at the same time, a low complexity – for instance have memory of order one. In [51–53], an information-theoretic framework was developed as a practical design guideline for novel codes. To this end, the following optimization has been considered in [52]

$$\begin{aligned} C_{M_s}^{\text{uniform}} = \max_{P_{ij}; (i,j) \in \mathcal{T}} \mathcal{I}(\mathbf{x}; \mathbf{y}) \quad \text{s.t.} \quad & P_{ij} = \Pr(S_k = j | S_{k-1} = i) \in \{0, 2^{-K}\}, \\ & S_k = \mathbf{x}_{k-M_s+1}^k \in \mathcal{X}^{M_s}, \\ & M_s \geq L, \end{aligned} \quad (6)$$

where  $K$  defines the rate  $R = K/(2N)$  of the code (for QPSK modulation) and  $C_{M_s}^{\text{uniform}}$  is the maximum channel capacity that can be attained with a homogeneous Markov source of order  $M_s$ . In general, the capacity of an unconstrained Markov source of order  $M_s$  [30] is higher than this *uniform capacity* ( $P_{ij}$  takes 0 or  $2^{-K}$ ), i.e.,  $C_{M_s}^{\text{uniform}} \leq C_{M_s}$ .

This coding approach incorporates the following four ideas:

1. In order to avoid the complexity of maximizing an arbitrary Markov source, we restrict the optimization to homogeneous Markov sources.

2. We choose the memory length  $M_s$  of the source to be roughly as long as the number of channel taps  $L$  but not shorter. The reason is that the information rate of a Markov source of order  $M_s = L$  is noticeably larger compared to the i.u.d. capacity, but memory lengths above  $L$  yield only a small additional gain in information rate.
3. As we want to avoid the use of distribution shapers, the number of transmit symbols is fixed at  $2^K$  (irrespective of the current state). Thus, the encoder can be realized as a look-up table and we obtain the data encoding rule

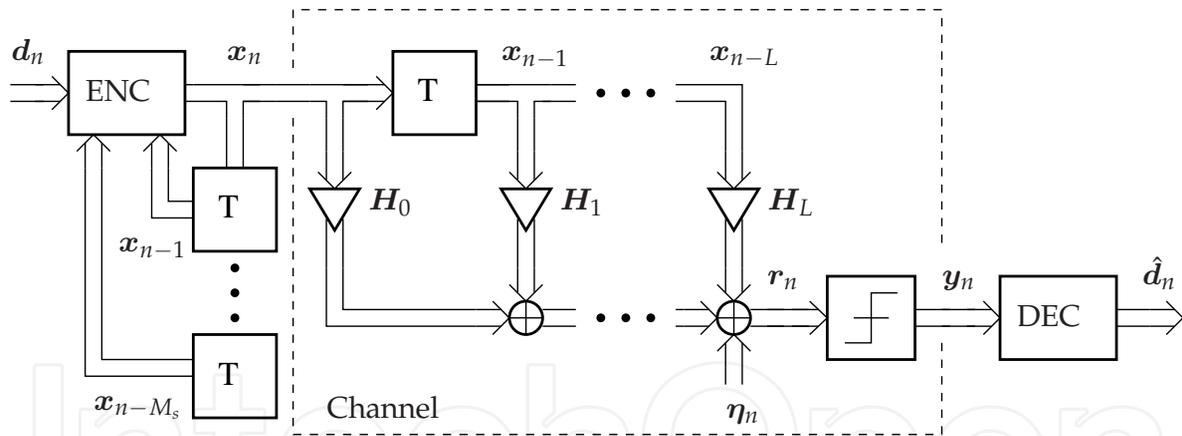
$$\mathbf{x}_n = \text{ENC}(\mathbf{d}_n, [\mathbf{x}_{n-1}, \dots, \mathbf{x}_{n-M_s}]),$$

where  $\mathbf{d}_n \in \{0, 1\}^K$  is the data vector of the source. Using QPSK modulation the realizable code rates  $R$  are

$$\frac{K}{2N} \in \left\{ \frac{1}{2N}, \frac{2}{2N}, \dots, 1 \right\}.$$

4. The optimized transition probabilities are uniformly distributed and at the same time they approximate the capacity-achieving input distribution. Hence, the optimized transition probability matrix  $P_{ij}$  serves as an inner code that can be concatenated with an outer Turbo-like code, e.g. *low density parity check code* (c.f. section 4), in order to reach information rates (well) above the i.u.d. capacity [53].

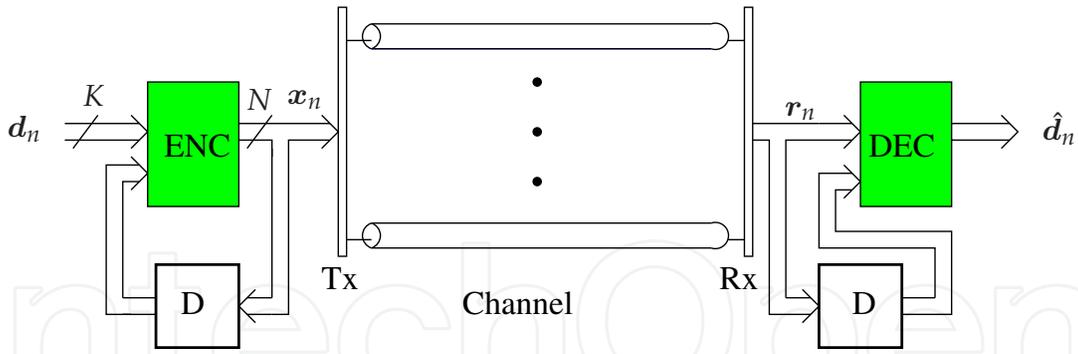
Fig. 11 illustrates the channel model together with the encoder and decoder. The optimization



**Figure 11.** Encoding and Decoding for MIMO Channels with ISI and Single-Bit Output Quantization.

(6) is non-convex but can be solved by an efficient greedy algorithm [52] that delivers an optimized transition probability matrix  $\mathbf{P} = [P_{i,j}]$  that maximizes the mutual information between the input and the output.

In Figure 12, a coded bus system employing a memory-based code with a code rate of  $K/N$  is shown. A fixed bus access time  $T_{\text{CU}}^{\text{cod}}$  is chosen such that two channel temporal taps are significant ( $L = 1$ ). The encoding scheme is time-invariant and has the property that the data vector,  $\mathbf{d}_n = [d_1[n], \dots, d_K[n]]^T \in \{0, 1\}^{K \times 1}$ , is encoded and decoded instantaneously (without latency). The actual code vector  $\mathbf{x}_n$  depends on the input data vector  $\mathbf{d}_n$  and the previous transmitted vector  $\mathbf{x}_{n-1}$ .



**Figure 12.** Noisy bus communication with a memory-based Code.

$x_{n-1} \setminus \text{Data}$	000	001	010	011	100	101	110	111
0000	0000	0001	0011	0111	1000	1100	1110	1111
0001	0000	0001	0111	1000	1001	1100	1101	1111
0010	0000	0010	0011	0100	1010	1011	1110	1111
0011	0000	0010	0011	0100	1010	1011	1110	1111
0100	0000	0001	0100	0101	0111	1100	1101	1111
0101	0000	0001	0100	0101	0111	1100	1101	1111
0110	0000	0011	0100	0110	0111	1100	1110	1111
0111	0000	0001	0100	0110	0111	1100	1110	1111
1000	0000	0001	0011	1000	1001	1011	1110	1111
1001	0000	0001	0011	1000	1001	1011	1100	1111
1010	0000	0010	0011	0100	1010	1011	1110	1111
1011	0000	0010	0011	0100	1010	1011	1110	1111
1100	0000	0001	0100	0101	0111	1100	1101	1111
1101	0000	0001	0100	0101	0111	1100	1101	1111
1110	0000	0010	0011	0110	0111	1000	1110	1111
1111	0000	0001	0011	0111	1000	1100	1110	1111

**Table 1.** The mapping function of a 3/4 optimized code:  $x_n = \text{ENC}(d_n, x_{n-1})$ .

At the receiver side, the decoder uses the value of the current and the previous channel outputs to reconstruct an estimate of the data vector  $\hat{d}_n$  as

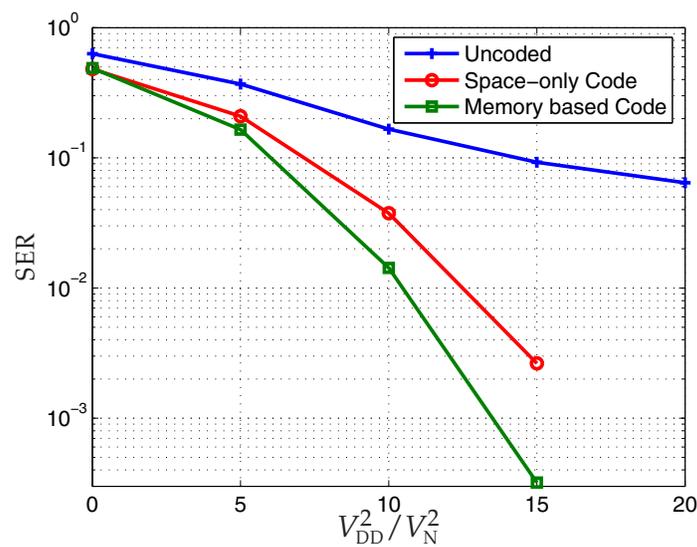
$$\hat{d}_n = g(\mathbf{y}_n, \mathbf{y}_{n-1}) = \arg \max_d \Pr(d_n = d \mid \mathbf{y}_n, \mathbf{y}_{n-1}). \quad (7)$$

Obviously, the mapping done by this function performs a maximum-likelihood estimation of  $d_n$  based on  $\mathbf{y}_n$  and  $\mathbf{y}_{n-1}$ <sup>1</sup>.

Although this approach seems quite heuristic, its usefulness can be demonstrated by simulation. Table 1 lists the mapping function of a code designed for a bus with  $N = 4$  mutually coupled, tapped  $RC$  lines as shown in Fig. 1 used at  $T_{CU} = 9RC_s = 1.5RC_c$  symbol time, where  $R$  is the serial resistance,  $C_s$  is the ground capacitance and  $C_c$  is the coupling capacitance (c.f. section 2.1).

<sup>1</sup> The reliability of the decoding can be improved if earlier or later outputs are considered using the BCJR algorithm (forward-backward algorithm), at the cost of some latency.

Its performance in terms of symbol error rate (SER) when applied to a noisy bus system, compared to uncoded transmission, is shown in Figure 13. The uncoded transmission reveals an error floor (a residual SER at vanishing noise variance) due to signaling along the  $RC$ -specific time. However, as we see in Figure 13, the optimized code does not see any error floor. Besides, it turns out that the achievable power savings of this code (in terms of energy per transmitted information bit) is 40%, without taking into account the power overhead of the codec circuit. The SER curve of a space-only code, which has been optimized by exhaustive search, is also plotted. Due to its simplicity, this code performs inherently worse than the discussed memory-based code. Although several coding schemes can be found in the literature [29, 32], such a unified framework that jointly address power, rate, and reliability aspects, simultaneously is new.



**Figure 13.** Symbol error rate (SER) as function of  $V_{DD}^2/V_N^2$  for an interconnect with four lines, employing a memory-based code of rate 3/4, compared to the uncoded case. The performance of a space-only code is also plotted.

We note that, for large buses, it is impractical to encode all bits at once because of the large complexity in the design and the implementation of the codec circuit. Therefore, partial coding can be employed in which the bus is partitioned into sub-buses of smaller width, which are encoded separately. The partitioning requires some additional wires since a shielding wire has to be placed between every two adjacent sub-buses.

### 3.2. Low-resolution ADC: Linear signal-processing

In the following, we concentrated on receive signal processing and our aim is to study the applicability of standard equalization techniques for our application, where the receiver is equipped with a low to moderate ADC for each antenna or port. A modified version of the standard linear receiver designs is presented in the context of MIMO communication with quantized output, taking into account the presence of the quantizer. An essential aspect of our analysis is that no assumption of uncorrelated white quantization errors is made. The performance of the modified receiver designs as well as the effects of quantization are

studied theoretically and experimentally. Thereby, perfect channel state information (CSI) at the receiver is assumed, which can be obtained even with coarse quantization as discussed in Section 3.3.

In [33], the joint optimization of the linear receiver and the quantizer in a MIMO system is addressed. The figure of merit that has been used for the design of the optimum quantizer and receiver is the *mean square error* (MSE). Based on this MSE approach, the communication performance (in terms of channel capacity) of the quantized MIMO channel is studied. Our work [34] generalizes this modified MMSE filter to frequency selective channels. Motivated by the same approach, the authors of [36] optimized the Decision Feedback Equalizer (DFE) for the flat MIMO channel with quantized outputs.

In this and the following Section, we provide a summary of these works. Throughout these sections,  $r_{\alpha\beta}$  denotes  $E[\alpha\beta^*]$ . The operators  $(\bullet)^T$ ,  $(\bullet)^H$ ,  $(\bullet)^*$ ,  $\text{tr}[\bullet]$  stand for transpose, Hermitian transpose, complex conjugate, and trace of a matrix, respectively.

### 3.2.1. System model

Let us now consider a point to point MIMO Gaussian channel, where the transmitter operates  $M$  antennas and the receiver employs  $N$  antennas. Figure 14 shows the general form of a quantized MIMO system, where  $\mathbf{H} \in \mathbb{C}^{N \times M}$  is the channel matrix. For simplicity, inter-symbol interference (ISI) is ignored, even though considering it would be straightforward. The vector  $\mathbf{x} \in \mathbb{C}^M$  comprises the  $M$  transmitted symbols with zero-mean and covariance  $\mathbf{R}_{xx} = E[\mathbf{x}\mathbf{x}^H]$ . The vector  $\boldsymbol{\eta}$  refers to zero-mean complex circularly symmetric Gaussian noise with a covariance matrix  $\mathbf{R}_{\eta\eta} = E[\boldsymbol{\eta}\boldsymbol{\eta}^H]$ , while  $\mathbf{y} \in \mathbb{C}^N$  is the unquantized channel output:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta}. \quad (8)$$

In our system, the real parts  $y_{i,R}$  and the imaginary parts  $y_{i,I}$  of the receive signals  $y_i$ ,  $1 \leq i \leq N$ , are each quantized by a  $b$ -bit resolution uniform/non-uniform scalar quantizer. Thus, the resulting quantized signals are given by

$$r_{i,l} = Q(y_{i,l}) = y_{i,l} + q_{i,l}, \quad l \in \{R, I\}, \quad 1 \leq i \leq N, \quad (9)$$

where  $Q(\cdot)$  denotes the quantization operation and  $q_{i,l}$  is the resulting quantization error.

The matrix  $\mathbf{G} \in \mathbb{C}^{M \times N}$  represents the receive filter, which delivers the estimate  $\hat{\mathbf{x}}$

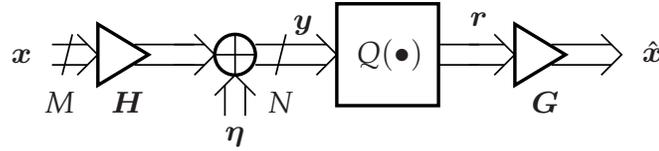
$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{r}. \quad (10)$$

Our aim is to choose the quantizer and the receive matrix  $\mathbf{G}$  minimizing the MSE =  $E[\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2]$ , taking into account the quantization effect. Since the ADC can drastically affect the performance of the system, it should be also designed carefully.

### 3.2.2. Quantizer characterization

Each quantization process can be given a distortion factor  $\rho_q^{(i,l)}$  to indicate the relative amount of quantization noise generated, which is defined as follows

$$\rho_q^{(i,l)} = \frac{E[q_{i,l}^2]}{r_{y_{i,l}y_{i,l}}}, \quad (11)$$



**Figure 14.** Quantized MIMO System.

where  $r_{y_{i,l}y_{i,l}} = E[y_{i,l}^2]$  is the variance of  $y_{i,l}$  and the distortion factor  $\rho_q^{(i,l)}$  depends on the number of quantization bits  $b$ , the quantizer type (uniform or non-uniform) and the probability density function of  $y_{i,l}$ . Note that the signal-to-quantization noise ratio (SQNR) has an inverse relationship with regard to the distortion factor. The uniform/non-uniform quantizer design is based on minimizing the *mean square error* (distortion) between the input  $y_{i,l}$  and the output  $r_{i,l}$  of each quantizer. In other words, the SQNR values are maximized. With this optimal design of the scalar finite resolution quantizer, whether uniform or not, the following equations hold for all  $0 \leq i \leq N, l \in \{R, I\}$  [35, 37, 38]

$$E[r_{i,l}q_{i,l}] = 0 \quad (12)$$

$$E[y_{i,l}q_{i,l}] = -\rho_q^{(i,l)}r_{y_{i,l}y_{i,l}}. \quad (13)$$

Obviously, (13) follows from (11) and (12). Under multipath propagation conditions and for large number of antennas, the quantizer input signals  $y_{i,l}$  will be approximately Gaussian distributed and thus, they undergo nearly the same distortion factor  $\rho_q$ , i.e.,  $\rho_q^{(i,l)} = \rho_q \forall i \forall l$ . Furthermore, the optimal parameters of the uniform as well as the non-uniform quantizer and the resulting distortion factor  $\rho_q$  for Gaussian distributed signal are tabulated in [35] for different resolutions  $b$ .

Now, let  $q_i = q_{i,R} + jq_{i,I}$  be the complex quantization error. Under the assumption of uncorrelated real and imaginary part of  $y_i$ , the following relations are obtained

$$r_{q_i q_i} = E[q_i q_i^*] = \rho_q r_{y_i y_i}, \text{ and } r_{y_i q_i} = E[y_i q_i^*] = -\rho_q r_{y_i y_i}. \quad (14)$$

This particular choice of the (non-)uniform scalar quantizer minimizing the distortion between  $r$  and  $y$ , combined with the receiver developed in the next Section, is also optimal with respect to the total MSE between the transmitted symbol vector  $x$  and the estimated symbol vector  $\hat{x}$ , as we will see later.

### 3.2.3. Nearly optimal linear receiver

The linear receiver  $G$  that minimizes the MSE,  $E[\|\varepsilon\|_2^2] = E[\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2] = E[\|\mathbf{x} - G\mathbf{r}\|_2^2]$ , can be written as:

$$G = \mathbf{R}_{xr} \mathbf{R}_{rr}^{-1}, \quad (15)$$

and the resulting MSE equals

$$\text{MSE} = \text{tr}(\mathbf{R}_{\varepsilon\varepsilon}) = \text{tr}(\mathbf{R}_{xx} - \mathbf{R}_{xr} \mathbf{R}_{rr}^{-1} \mathbf{R}_{xr}^H), \quad (16)$$

where  $\mathbf{R}_{xr}$  equals

$$\mathbf{R}_{xr} = E[\mathbf{x}\mathbf{r}^H] = E[\mathbf{x}(\mathbf{y} + \mathbf{q})^H] = \mathbf{R}_{xy} + \mathbf{R}_{xq}, \quad (17)$$

and  $\mathbf{R}_{rr}$  can be expressed as

$$\mathbf{R}_{rr} = \mathbb{E}[(\mathbf{y} + \mathbf{q})(\mathbf{y} + \mathbf{q})^H] = \mathbf{R}_{yy} + \mathbf{R}_{yq} + \mathbf{R}_{yq}^H + \mathbf{R}_{qq}. \quad (18)$$

We have to determine the linear filter  $\mathbf{G}$  as a function of the channel parameters and the quantization distortion factor  $\rho_q$ . To this end, we derive all needed covariance matrices by using the fact that the quantization error  $q_i$ , conditioned on  $y_i$ , is statistically independent from all other random variables of the system. First we calculate  $r_{y_i q_j} = \mathbb{E}[y_i q_j^*]$  for  $i \neq j$

$$\begin{aligned} \mathbb{E}[y_i q_j^*] &= \mathbb{E}_{y_j} [\mathbb{E}[y_i q_j^* | y_j]] \\ &= \mathbb{E}_{y_j} [\mathbb{E}[y_i | y_j] \mathbb{E}[q_j^* | y_j]] \\ &\approx \mathbb{E}_{y_j} [r_{y_i y_j} r_{y_j y_j}^{-1} y_j \mathbb{E}[q_j^* | y_j]] \end{aligned} \quad (19)$$

$$\begin{aligned} &= r_{y_i y_j} r_{y_j y_j}^{-1} \mathbb{E}[y_j q_j^*] \\ &= -\rho_q r_{y_i y_j}. \end{aligned} \quad (20)$$

Note that, in (19), we approximate the Bayesian estimator  $\mathbb{E}[y_i | y_j]$  with the linear estimator  $r_{y_i y_j} r_{y_j y_j}^{-1} y_j$ , which holds with equality if the vector  $\mathbf{y}$  is jointly Gaussian distributed. Eq. (20) follows from (14). Summarizing the results of (14) and (20), we obtain

$$\mathbf{R}_{yq} \approx -\rho_q \mathbf{R}_{yy}. \quad (21)$$

Similarly, we evaluate  $r_{q_i q_j}$  for  $i \neq j$  using (21), and with (14) we arrive at

$$\mathbf{R}_{qq} \approx \rho_q \mathbf{R}_{yy} - (1 - \rho_q) \rho_q \text{nondiag}(\mathbf{R}_{yy}), \quad (22)$$

where  $\text{nondiag}(\mathbf{A})$  obtained from a matrix  $\mathbf{A}$  by setting its diagonal elements to zero. Inserting the expressions (21) and (22) into (18), we obtain

$$\mathbf{R}_{rr} \approx (1 - \rho_q)(\mathbf{R}_{yy} - \rho_q \text{nondiag}(\mathbf{R}_{yy})). \quad (23)$$

Also in a similar way, we get  $\mathbf{R}_{xq} = \mathbb{E}[\mathbf{x} \mathbf{q}^H] \approx -\rho_q \mathbf{R}_{xy}$ , and (17) becomes

$$\mathbf{R}_{xr} \approx (1 - \rho_q) \mathbf{R}_{xy}. \quad (24)$$

In summary, we get from (23) and (24) the following expression for the Wiener filter from (15) operating on quantized data

$$\mathbf{G}_{\text{WFQ}} \approx \mathbf{R}_{xy} (\mathbf{R}_{yy} - \rho_q \text{nondiag}(\mathbf{R}_{yy}))^{-1}, \quad (25)$$

and for the resulting MSE, we obtain using (16)

$$\text{MSE}_{\text{WFQ}} \approx \text{tr} \left[ \mathbf{R}_{xx} - (1 - \rho_q) \mathbf{R}_{xy} (\mathbf{R}_{yy} - \rho_q \text{nondiag}(\mathbf{R}_{yy}))^{-1} \mathbf{R}_{xy}^H \right]. \quad (26)$$

We obtain  $\mathbf{R}_{yy}$  and  $\mathbf{R}_{xy}$  easily from our system model

$$\mathbf{R}_{yy} = \mathbf{R}_{\eta\eta} + \mathbf{H} \mathbf{R}_{xx} \mathbf{H}^H, \quad (27)$$

$$\mathbf{R}_{xy} = \mathbf{R}_{xx} \mathbf{H}^H. \quad (28)$$

Let us examine the first derivative of the  $\text{MSE}_{\text{WFQ}}$  in (26) with respect to  $\rho_q$

$$\frac{\partial \text{MSE}_{\text{WFQ}}}{\partial \rho_q} = \text{tr} \left[ \mathbf{G}_{\text{WFQ}} \text{diag}(\mathbf{R}_{yy}) \mathbf{G}_{\text{WFQ}}^H \right] > 0, \quad (29)$$

where  $\mathbf{G}_{\text{WFQ}}$  is given in (25). Therefore the  $\text{MSE}_{\text{WFQ}}$  is monotonically increasing in  $\rho_q$ . Since we choose the quantizer to minimize the distortion factor  $\rho_q$ , our receiver and quantizer designs are jointly optimum with respect to the total MSE.

### 3.2.4. Lower bound on the mutual information and the capacity

In this section, we develop a lower bound on the mutual information rate between the input sequence  $\mathbf{x}$  and the quantized output sequence  $\mathbf{r}$  of the system in Figure 14, based on our MSE approach. Generally, the mutual information of this channel can be expressed as [26]

$$I(\mathbf{x}, \mathbf{r}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{r}). \quad (30)$$

Given  $\mathbf{R}_{xx}$  under a power constraint  $\text{tr}(\mathbf{R}_{xx}) \leq P_{\text{Tr}}$ , we choose  $\mathbf{x}$  to be Gaussian, which is not necessarily the capacity achieving distribution for our quantized system. Then, we can obtain a lower bound for  $I(\mathbf{x}, \mathbf{r})$  (in bit/transmission) as

$$\begin{aligned} I(\mathbf{x}, \mathbf{r}) &= \log_2 |\mathbf{R}_{xx}| - h(\mathbf{x}|\mathbf{r}) \\ &= \log_2 |\mathbf{R}_{xx}| - h(\mathbf{x} - \hat{\mathbf{x}}|\mathbf{r}) \\ &\geq \log_2 |\mathbf{R}_{xx}| - h(\mathbf{x} - \hat{\mathbf{x}}) \end{aligned} \quad (31)$$

$$\geq \log_2 \frac{|\mathbf{R}_{xx}|}{|\mathbf{R}_{\varepsilon\varepsilon}|}. \quad (32)$$

Since conditioning reduces entropy, we obtain inequality (31). On the other hand, the second term in (31) is upper bounded by the entropy of a Gaussian random variable whose covariance is equal to the error covariance matrix  $\mathbf{R}_{\varepsilon\varepsilon}$  of the linear MMSE estimate of  $\mathbf{x}$ . Finally, we get using (26) and (28)

$$I(\mathbf{x}, \mathbf{r}) \gtrsim -\log_2 \left| \mathbf{I} - (1 - \rho_q) \mathbf{R}_{xy} (\mathbf{R}_{yy} - \rho_q \text{nondiag}(\mathbf{R}_{yy}))^{-1} \mathbf{H} \right|. \quad (33)$$

Considering the case of low SNR values, we get easily with  $\mathbf{R}_{yy} \approx \mathbf{R}_{\eta\eta}$ , (33) and (28), the following first order approximation of the mutual information<sup>2,3</sup>

$$I(\mathbf{x}, \mathbf{r}) \gtrsim (1 - \rho_q) \text{tr}[\mathbf{R}_{xx} \mathbf{H}^H \mathbf{R}_{\eta\eta}^{-1} \mathbf{H}] / \log(2). \quad (34)$$

Compared with the mutual information  $I(\mathbf{x}, \mathbf{y})$  for the unquantized case, also at low SNR [39], the mutual information for the quantized channel degrades only by the factor  $(1 - \rho_q)$ . For the spacial case  $b = 1$ , we have  $\rho_q|_{b=1} = 1 - \frac{2}{\pi}$  (see [35]) and the degradation of the mutual

<sup>2</sup> We assume also that  $\rho_q \ll 1$  (or  $\mathbf{R}_{\eta\eta}$  is diagonal).

<sup>3</sup> Note that  $\log |\mathbf{I} + \Delta \mathbf{X}| \approx \text{tr}(\Delta \mathbf{X})$ .

information becomes

$$\lim_{\text{SNR} \rightarrow 0} \frac{I(\mathbf{x}, \mathbf{r})}{I(\mathbf{x}, \mathbf{y})} \Big|_{b=1} \approx \frac{2}{\pi}. \quad (35)$$

In other words, the power penalty due to the 1-bit quantization is approximately equal  $\frac{\pi}{2}$  (1.96 dB) at low SNR. This shows that mono-bit ADCs may be used to save system power without an excessive degradation in performance, and confirms the significant potential of the coarsely quantized UWB MIMO channel. Using a different approach, [40] presented a similar result, and showed that the above approximation is asymptotically exact.

### 3.2.5. Simulation results

The performance of the modified Wiener filter for a 4- and 5-bit quantized output MIMO system (WFQ), in terms of BER averaged over 1000 channel realizations, is shown in Figure 15 for a  $10 \times 10$  MIMO system (QPSK), compared with the conventional Wiener filter (WF) and Zero-forcing filter (ZF). The symbols and the noise samples are assumed to be uncorrelated, that is,  $\mathbf{R}_{xx} = \sigma_x^2 \mathbf{I}$  and  $\mathbf{R}_{\eta\eta} = \sigma_\eta^2 \mathbf{I}$ . Hereby, the SNR (in dB) is defined as

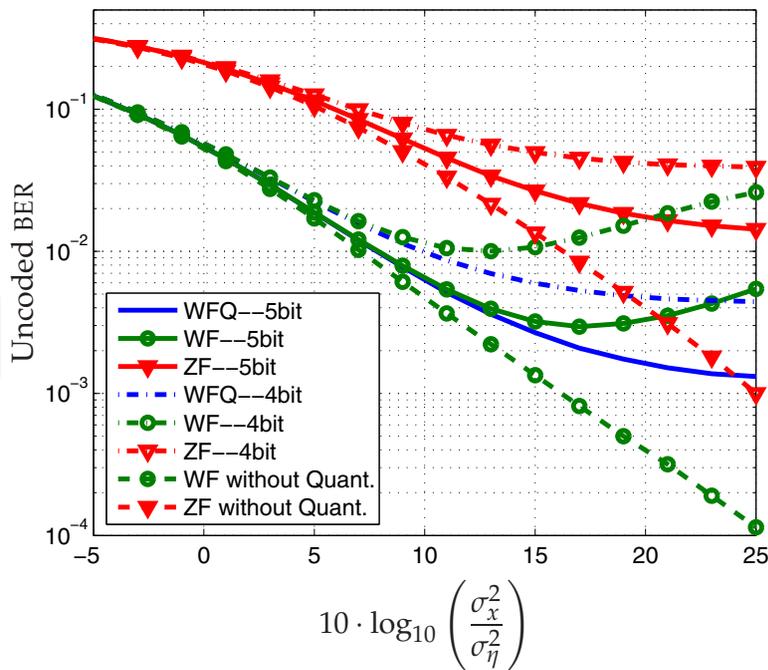
$$\text{SNR} = 10 \cdot \log_{10} \left( \frac{\sigma_x^2}{\sigma_\eta^2} \right). \quad (36)$$

Furthermore, we used a generic channel model, where the entries of  $\mathbf{H}$  are complex-valued realizations of independent zero-mean Gaussian random variables with unit variance. Clearly, the modified Wiener filter outperforms the conventional Wiener filter at high SNR. This is because the effect of the quantization error is more pronounced at higher SNR values when compared to the additive Gaussian noise variance. Since the conventional Wiener filter converges to the ZF-filter at high SNR values and loses its regularized structure, its performance degrades asymptotically to the performance of the ZF-filter, when operating on quantized data. For comparison, we also plot the BER curves for the WF and ZF filter, for the case when no quantization is applied.

## 3.3. Channel estimation

Because in general, the MIMO channel cannot be assumed known a-priori, a channel estimation has to be performed. In practice, it is highly desirable that the channel is estimated directly by the communication device – in our case by on-chip digital circuitry. However, this implies that the channel estimator is restricted to use received signal samples of a pilot sequence after single-bit quantization in the extreme case. This motivates investigation of channel estimation with coarse quantization. This problem was first addressed by [27], where a maximum likelihood (ML) channel estimation with quantized observation is presented. In general, the solution cannot be given in closed form, but requires an iterative numerical approach, which hampers the analysis of performance.

In [28], it has been shown that – in contrast to unquantized channel estimation – different orthogonal pilot sequences (with same average total transmit power and same length) yield different performances. Especially, orthogonality in the time-domain (time-multiplexed pilots) can be preferable to orthogonality in space. With orthogonal pilots



**Figure 15.** The WFQ vs. the conventional WF and ZF receivers, QPSK modulation with  $M = 10$ ,  $N = 10$ , 4- ( $\rho_q = 0.01154$ ) and 5- ( $\rho_q = 0.00349$ ) bit uniform quantizer.

that are multiplexed in time, the problem can be reduced from the MIMO to the SIMO (single-input-multiple-output) case, because each line of the multiconductor interconnect is excited separately for time-multiplexed pilots. Finally, the problem can be reduced to the SISO (single-input-single-output) case, when the channel estimation is performed separately in parallel at each receiving end of the multiconductor interconnect. For this case, in [28], a closed-form solution can be found for the maximum likelihood channel estimation problem, which makes performance analysis possible in an analytical fashion.

In [50], a more general setting for parameter estimation based on quantized observations was studied, which covers many processing tasks, e.g. channel estimation, synchronization, delay estimation, Direction Of Arrival (DOA) estimation, etc. An Expectation Maximization (EM) based algorithm is proposed to solve the Maximum a Posteriori Probability (MAP) estimation problem. Besides, the Cramér-Rao Bound (CRB) has been derived to analyze the estimation performance and its behavior with respect to the signal-to-noise ratio (SNR). The presented results treat both cases: pilot aided and non-pilot aided estimation. The paper extensively dealt with the extreme case of single bit quantization (comparator) which simplifies the sampling hardware considerably. It also focused on MIMO channel estimation and delay estimation as application area of the presented approach. Among others, a  $2 \times 2$  channel estimation using 1-bit ADC is considered, which shows that reliable estimation may still be possible even when the quantization is very coarse, with any desired accuracy, provided the pilot sequence is long enough. Since in on-chip and chip-to-chip communications, the channel almost does not change in time, it is possible to use very long pilot sequences, and run the channel estimation only once, or once in a while.

### 4. Efficient digital hardware architecture

Sole optimization of transmitting power in the standardization and conception phase of communication channels results in highly complex and energy-intensive receivers with a complex channel decoder as one of its key components. Neglecting the energy dissipation of the integrated decoder in this early phase results in suboptimal and, thus, costly communication systems in terms of manufacturing and usage costs. In the previous part of this chapter approaches to reduce the ADC complexity and, thus, the complexity of the subsequent digital components by using single-bit or medium-low resolution quantizations have been discussed. A quantitative comparison of these new approaches to standard receivers requires accurate cost models of the digital components. Quite accurate cost models are available for most of the communication system components except for channel decoders. While such cost models can be easily derived for Viterbi, Reed-Solomon, and Turbo Decoders, an estimation of the silicon area and the energy dissipation of LDPC decoders is challenging due to the high internal communication effort between the basic components.

Although LDPC codes have already been introduced by Gallager in 1962 [55], up to now they are known to achieve the best decoding performance [56] and are adopted in various communication standards (e.g. [57],[58],[59]) and other applications such as hard-disk drives [60]. They belong to the class of block codes and, thus, can be defined by a parity-check matrix  $H$  with  $m$  rows and  $n$  columns or by the corresponding Tanner Graph. Both are shown in Figure 16 for a very simplified LDPC code. Each row of the parity-check matrix represents one parity check wherein a '1'-entry in column  $i$  and row  $j$  indicates, that the received symbol  $i$  takes part in parity check number  $j$ . In the Tanner Graph such a parity check is represented by one so called check node and each column by one bit node. Furthermore, the number of one entries per row  $d_C$  (column  $d_V$ ) defines the number of connected bit (check) nodes per check (bit) node.

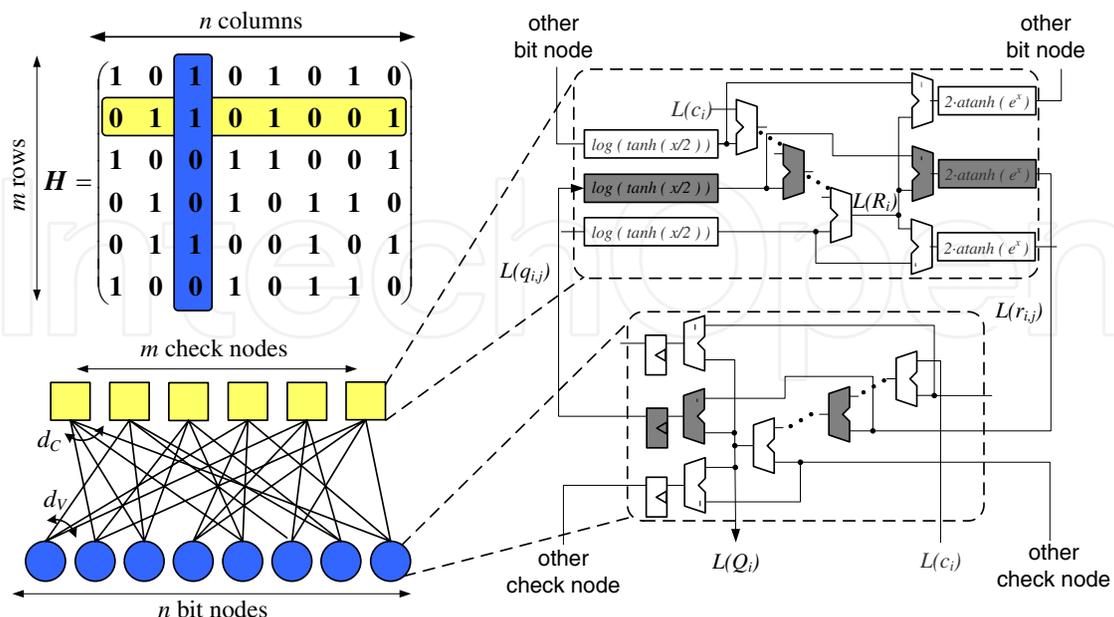


Figure 16. Parity-check matrix, Tanner-Graph and non-linear recursive decoder loop

Figure 16 also illustrates one of the  $n \cdot d_V$  non-linear recursive loops of the resulting decoder. In each iteration the extrinsic information  $L(q_{i,j})$  on the received symbol  $i$  is sent to check node  $j$ . Here, new A-posteriori information  $L(r_{i,j})$  is derived. The sign of  $L(r_{i,j})$  is chosen in such a way, that the confidence in the received symbol  $i$  indicated by the magnitude of  $L(q_{i,j})$  increases. For the sake of clarity only the magnitude calculation is illustrated in Figure 16. Considering the original Sum-Product decoding algorithm [55] the check node consists of transcendent functions and a multi-operand adder with subsequent subtractor stages. Here, the basic idea is, whenever all participating symbols in that parity check feature a high confidence in their current estimation, the magnitude of the A-posteriori information is high. The A-posteriori information  $L(r_{i,j})$  is then sent back to the bit node. Here, all information of symbol  $i$ , namely the  $d_V$  A-posteriori information and the received information  $L(c_i)$ , are combined using a multi-operand adder resulting in a new estimation  $L(Q_i)$  of symbol  $i$ . To avoid decoding-performance-demoting cycles, in the next decoding iteration only the extrinsic information  $L(q_{i,j}) = L(Q_i) - L(r_{i,j})$  is used instead of  $L(Q_i)$ . For more information on the decoding algorithm and possible fix-point realizations refer to [61].

A metric for the code's and, thus, the decoder's complexity is the number of '1'-entries in the matrix  $n \cdot d_V = m \cdot d_C$ . Each '1'-entry can be assigned to a part of the bit- and check-node logic as highlighted in gray in Figure 16. Thereby, each '1'-entry leads to four two-operand adders/subtractors, a block for the calculation of  $\log(\tanh(x/2))$ , a block for the calculation of  $2 \cdot \operatorname{atanh}(e^x)$  and a register stage at the output of the bit node. Additionally,  $n \cdot d_V$  is a measure for the communication between the nodes as  $2 \cdot w \cdot n \cdot d_V$  bits are exchanged between the bit and the check nodes in each decoding iteration with  $w$  being the word length of the exchanged messages.

In high-throughput applications with a time-invariant parity-check matrix all bit and check nodes are typically instantiated in parallel as in the first integrated LDPC decoder [66]. Here, typically the  $m$  check nodes are realized in the center of the decoder floorplan surrounded by the  $n$  bit-node instances. The communication between the nodes is then realized by  $2 \cdot w \cdot n \cdot d_V$  dedicated interconnect lines. In [66] the logic area, which is the accumulated silicon area of all logic gates, is approximately  $25 \text{ mm}^2$ . However, the total of 26,624 interconnect lines can not be realized on this area. The silicon area needs to be artificially expanded until a successful routing of all interconnect lines could be established. The resulting global interconnect has a length of 80 m on a macro size of  $52.5 \text{ mm}^2$ . Thus, only 50% of the active silicon area is utilized in the final decoder. The impact of the complex global interconnect complicates the derivation of accurate area, timing and energy cost models which might be the reason why no cost models are available in literature so far. However, such models are necessary to avoid costly wrong decisions in early design phases, for example when choosing a certain LDPC code in the system-conception phase. Also in later design phases those models are indispensable, for example for a quantitative exploration of the architecture design space.

#### 4.1. Accurate area, timing, and energy cost models

In general the silicon area of a high-throughput LDPC decoders can be estimated using

$$A_{DEC\_P} = \max(A_L, A_R), \quad (37)$$

with  $A_L$  being the logic area and  $A_R$  the required area to realize the global interconnect. To reduce the logic area typically the approximative Min-Sum algorithm [70] is used which

estimates the magnitude of  $L(r_{i,j})$  using the minimal and second minimal magnitude of  $L(q_{i,j})$  (e.g. [62],[63],[64],[65]). The derivation of  $A_L$  for this decoding algorithm as the accumulated silicon area of all logic gates has been presented in [67]. The resulting total logic area can be estimated using

$$A_L = l_L^2 = 1000 \cdot n \cdot d_V \cdot (11.5 \cdot w + 2 \cdot ld(d_V)) \cdot \lambda^2. \quad (38)$$

This equation reveals a linear dependency between the code complexity  $n \cdot d_V$  and the accumulated gate area.

The major challenge in deriving an accurate routing-area model is the adaptability to different LDPC codes. It is possible to divide the problem into two parts: an estimation of the available and the required Manhattan length. Considering a certain logic area, the available Manhattan length is a measure for the routing resources above the decoder's node logic. Considering that the node layouts require  $M_L$  of the total  $M$  metal layers in the CMOS stack for the local interconnect,  $M_R = M - M_L$  metal layers are available for the realization of the global bit- and check-node communication. The required routing area  $A_R$  can then be determined by equalizing the available and the required Manhattan lengths. This means, that the available Manhattan length allows the realization of the required Manhattan length. Thereby, the available Manhattan length can be derived as

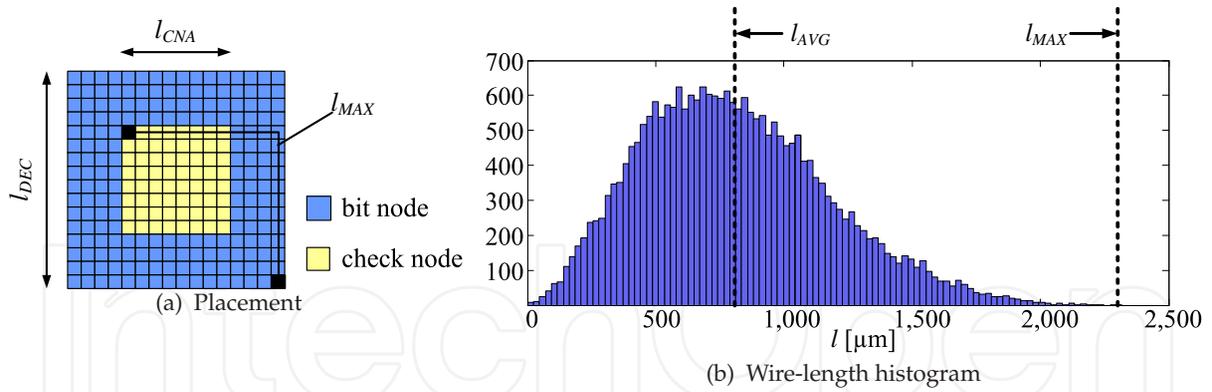
$$l_{AVAIL} = \frac{u}{p} \cdot \left( l_L^2 \cdot M_R + \min \left( l_{DEC}^2 - l_L^2, 0 \right) \cdot M \right), \quad (39)$$

with the routing pitch  $p$ , an utilization factor  $u$  for each metal layer, and a decoder macro side length of  $l_{DEC}$ . Considering that no artificial increase of the decoder is required ( $l_{DEC} \leq l_L$ ) the second term is zero and the available routing resources are on top of the node logic. If the decoder needs to be expanded, the whole metal stack is available between the node instances for the realization of the global interconnect. Therefore, this part is weighted with  $M$ .

The estimation of the required Manhattan length is more challenging as it depends on code characteristics as for example the number of interconnect lines and the average length of one interconnect line. An upper bound estimation of the required Manhattan length could be derived by using the maximum possible length  $l_{MAX}$  of one bit- and check-node connection which is shown in Figure 17(a). In a typical placement with the bit nodes surrounding the check-node array the longest possible connection runs from one corner of the decoder macro to the opposing corner of the check-node array. An analysis of the logic model [67] shows that the check nodes occupy about 60% of the complete decoder macro leading to a maximum Manhattan length of

$$l_{MAX} = 2 \cdot \left( l_{CNA} + \frac{l_{DEC} - l_{CNA}}{2} \right) = l_{DEC} + l_{CNA} = 1.77 \cdot l_{DEC}. \quad (40)$$

When looking at the wire-length histogram for an exemplary code (see Figure 17(b)) the average Manhattan length is significantly smaller than the maximum length leading to an overestimation of the required Manhattan length and, thus, of the required routing area. An analysis of various LDPC codes showed, that the shape of the wire-length histogram is always similar. Especially, the ratio between the average and the maximum Manhattan length was found to be almost constant as can be seen from Table 2. For the derivation of the average Manhattan lengths all placements have been optimized using a custom simulated annealing process [62]. While code no. 11 is the code adopted in [57], the other codes are taken from



**Figure 17.** Bit- and check-node architecture

Code nr.	$n$	$m$	$d_V$	$d_C$	$n \cdot d_V$	$l_{AVG} / l_{MAX}$	$\rho_{AVR} / \rho_{MAX}$ vertical	$\rho_{AVR} / \rho_{MAX}$ horizontal
1	96	48	3	6	288	0.33	0.58	0.59
2	408	204	3	6	1224	0.31	0.56	0.55
3	408	204	3	6	1224	0.30	0.55	0.56
4	408	204	3	6	1224	0.31	0.54	0.50
5	816	408	3	6	2448	0.31	0.52	0.57
6	816	408	5	10	4080	0.34	0.52	0.58
7	816	408	5	10	4080	0.34	0.52	0.57
8	816	408	5	10	4080	0.34	0.50	0.55
9	999	111	3	27	2997	0.37	0.36	0.33
10	1008	504	3	6	3024	0.32	0.54	0.46
11	2048	384	6	32	12288	0.37	0.42	0.40
12	4000	2000	3	6	12000	0.34	0.57	0.54
13	4000	2000	4	8	16000	0.35	0.55	0.55
14	8000	4000	3	6	24000	0.35	0.55	0.45

**Table 2.** Interconnect properties of various LDPC codes

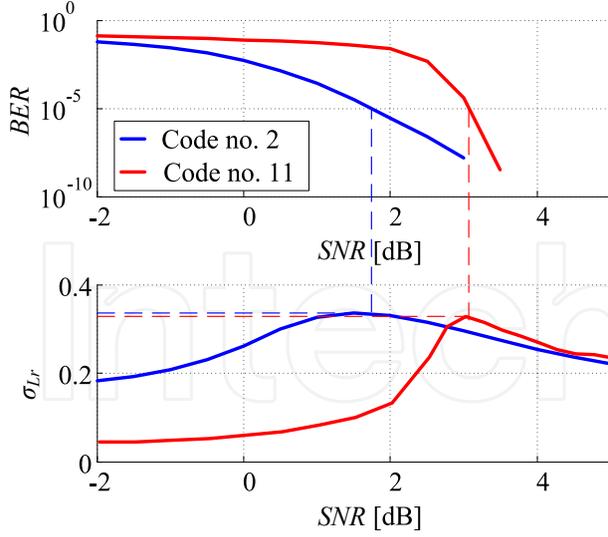
[68]. For a wide range of LDPC codes with code complexities  $n \cdot d_V$  between 300 and 24,000 the ratio varies only between 0.30 and 0.37. Approximating the ratio of the average to the maximum Manhattan length with 0.35 and using (40), the required Manhattan length can be estimated based on the decoder side length as

$$l_{REQ} = 1.2 \cdot n \cdot d_V \cdot w \cdot l_{DEC}. \quad (41)$$

Additionally, an estimation of the achievable utilization is possible based on the comparison of the average routing density  $\rho_{AVG}$  and the maximum routing density  $\rho_{MAX}$ . The ratios of these values for vertical and horizontal interconnect lines are also given in Table 2. Although there are exceptions (e.g. code no. 9) the utilization  $u = \rho_{AVG} / \rho_{MAX}$  is almost constant and will be chosen to  $u = 0.5$  in the following.

Considering that the decoder area needs to be expanded and assuming a uniform stretch (39) and (41) still hold. Then, the minimal required decoder area to realize the global interconnect  $A_R$  can be calculated by equating (39) and (41) and solving for  $l_R$  as

$$A_R = l_R^2 = \left( 1.2 \cdot \frac{n \cdot d_V \cdot w}{M} \cdot p + \sqrt{\left( 1.2 \cdot \frac{n \cdot d_V \cdot w}{M} \cdot p \right)^2 - l_L^2 \cdot \frac{M_R - M}{M}} \right)^2. \quad (42)$$



(a) SNR-dependent switching activity

Code nr.	SNR [dB]	$\sigma_{L(q)}$	$\sigma_{L(r)}$
1	2.6	0.14	0.30
2	1.7	0.17	0.34
3	1.7	0.17	0.33
4	2.9	0.14	0.30
5	1.5	0.18	0.34
6	0.5	0.16	0.33
7	0.5	0.16	0.33
8	0.5	0.16	0.33
9	9.8	0.41	0.29
10	1.4	0.19	0.34
11	3.0	0.20	0.33
12	1.2	0.19	0.34
13	0.1	0.20	0.35
14	1.2	0.19	0.34

(b) Table of switching activities ( $BER = 10^{-5}$ )**Figure 18.** Switching activity

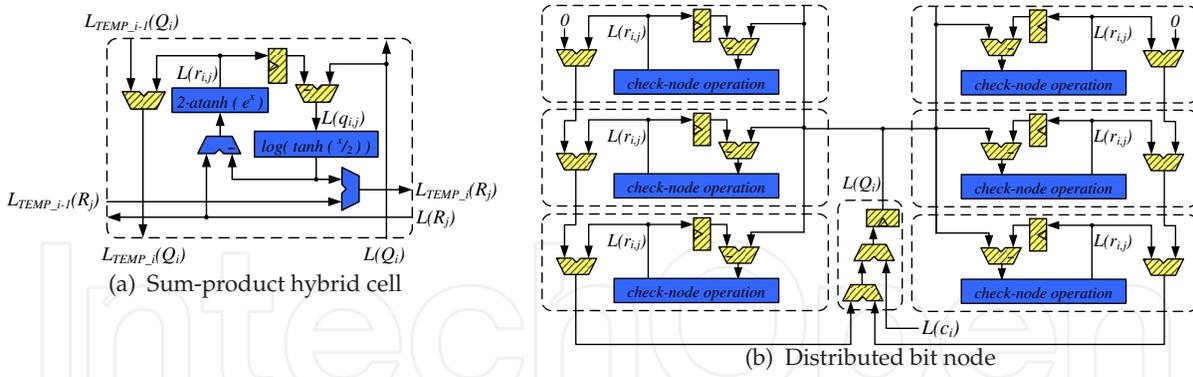
In contrast to the logic area, the routing area shows a quadratic dependence on code complexity. By comparing (38) and (42) it can be shown that the bit-parallel decoder is routing dominated as soon as

$$n \cdot d_V \geq 500 \cdot \frac{M_R^2}{w}. \quad (43)$$

The required artificial increase of the silicon area also impacts the other two decoder features: the energy per iteration  $E_{IT}$  and the iteration period, which is the required time for one decoding iteration and the inverse of the block throughput [69]. Here, only the interconnect fraction of the decoder energy will be discussed in detail. For more information on the derivation of the iteration period and the total decoder energy refer to [67]. The dynamic energy dissipation of the global interconnect can be estimated using

$$E_{INT} = \frac{1}{2} \cdot (\sigma_{Lq}(BER) + \sigma_{Lr}(BER)) \cdot \alpha \cdot C' \cdot \frac{l_{REQ}}{2} \cdot V_{DD}^2, \quad (44)$$

with  $V_{DD}$  being the supply voltage,  $C'$  the capacitive load per unit length of a minimum-spaced interconnect line and  $\alpha$  a fitting factor to cover the fact, that on average the global interconnect lines are not minimum spaced [67]. Furthermore, the switching activities on the interconnect lines from bit to check nodes ( $\sigma_{Lq}$ ) and vice versa ( $\sigma_{Lr}$ ) need to be considered. In Figure 18(a) the BER and the switching activity  $\sigma_{Lr}$  for two codes from Table 2 and different signal-to-noise ratios are illustrated. The switching activity highly depends on the considered SNR and is especially high in the so-called waterfall region when the BER starts to get significantly smaller. Furthermore, the two codes strongly differ when it comes to comparing the switching activities for a given SNR (e.g. 1dB). But, considering a specific BER (indicated by the dashed lines) an almost equal switching activity for the two codes can be observed (approx. 0.33 for a BER of  $10^{-5}$ ). The comparison of the switching activities  $\sigma_{Lq}$  and  $\sigma_{Lr}$  for all codes listed in Figure 18(b) shows, that this behavior is common for almost all other codes, as well. Therefore, a quite accurate estimation of the decoder energy based on the code parameters  $n$  and  $d_V$  is possible without knowledge of the actual LDPC code.



**Figure 19.** Hybrid-cell decoder architecture

#### 4.1.1. Hybrid-cell decoder architecture

The main routing problem of the bit- and check-node architecture arises from the high routing density at the border of the check-node array as it can be seen in the interconnect-density chart in Figure 20(a) for an exemplary code [57]. To overcome this drawback it is possible to break up the bit- and check-node clustering of the logic and rearrange it. The new idea is based on the observation, that each '1'-entry in the parity-check matrix can be assigned to certain parts of the decoder loop. Then, the decoder consists of  $n \cdot d_V$  small, equal basic components. A combination of the logic for one '1'-entry (see grey blocks in Figure 16) leads to the block diagram of one hybrid cell, as it is shown in Figure 19(a). This hybrid cell gets the accumulated information  $L_{TEMP_{i-1}}(Q_i)$  of the received A-priori information  $L(c_i)$  and of all A-posteriori information of the previous hybrid cells and adds the A-posteriori information  $L(r_{i,j})$  of check node  $j$ . The resulting information  $L_{TEMP_i}(Q_i)$  is forwarded to the next hybrid cell. The last hybrid cell in that column calculates  $L(Q_i)$  and sends this value back to all participating hybrid cells. A similar structure is used in the check-node part of the hybrid-cell where the calculation of  $L(R_j)$  is distributed over  $d_C$  hybrid cells. Although, here, the hybrid-cell approach considers a Sum-Product algorithm, it is also applicable to a Min-Sum based decoder. Therefore, the  $\Phi$  function and the multi-operand adder have to be replaced with basic compare-and-swap cells.

In contrast to the bit- and check-node architecture, in which the  $(d_V + 1)$ -operand adder in the bit node and the  $d_C$ -operand adder in the check node would be realized using a tree topology, the hybrid-cell architecture is based on an adder chain topology. However, it is possible to introduce tree-stages for the bit-node operation as illustrated in Figure 19(b). The  $L(r_{i,j})$  values are accumulated in two branches and the intermediate results are added to the channel information  $L(c_i)$  in an additional IO cell. A similar topology is possible for the check-node operation. The global interconnect of the hybrid-cell architecture has been realized in a 90-nm CMOS technology using five metal layers for the same code as used for the bit- and check-node architecture in Figure 20(a). In a first step, the placement of the nodes has been optimized using a custom simulated annealing process. Thereby, a placement scheme as depicted in Figure 20(b) with the hybrid cells surrounded by the io cells has been assumed. The advantage of the hybrid-cell architecture becomes obvious when comparing the two interconnect densities. The routing density of the hybrid-cell architecture is distributed more uniformly especially without high density peaks at the border of the bit- and check-node array. Thus, the average routing density of the hybrid-cell interconnect is higher than of the bit- and check-node architecture, promising a smaller silicon area.

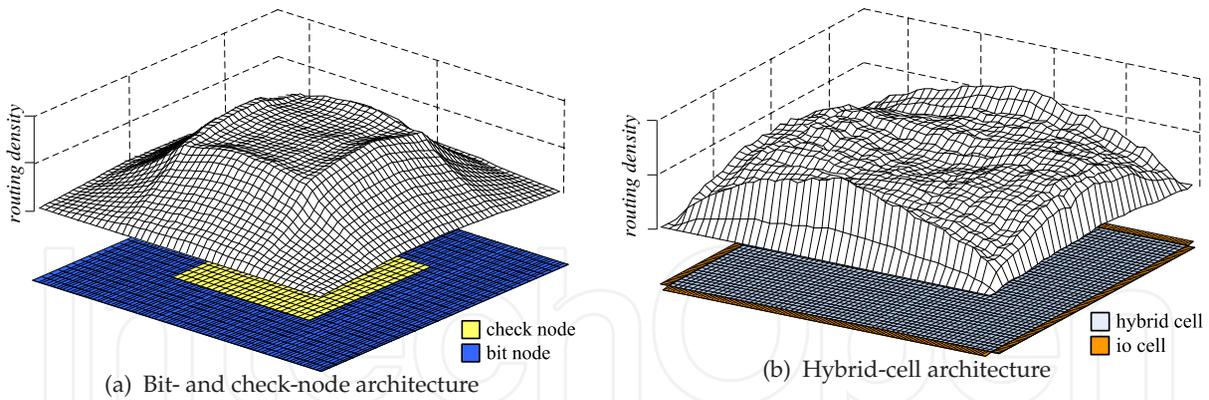


Figure 20. Routing density

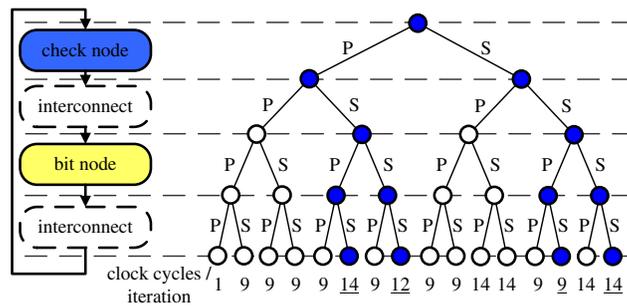
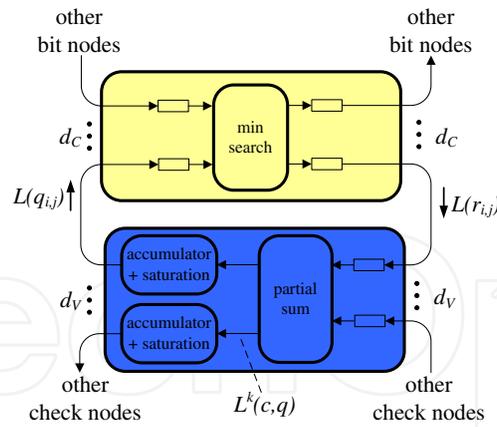


Figure 21. Bit- and check-node architecture design space.

#### 4.1.2. Hardware-efficient partially bit-serial decoder architecture

Another promising approach to reduce the decoder’s silicon area is the introduction of a bit-serial interconnect as proposed in [65]. The number of interconnect lines can be reduced by a factor of  $w$  resulting in a significant reduction in decoder area because of the quadratic dependency in (42). While the realized minimum search in the check node requires a most-significant-bit-first data flow in the check node the multi-operand adder in the bit node has to be realized using a least-significant-bit-first data flow. Therefore, the order of the bits needs to be flipped twice per iteration resulting in a high number of clock cycles. Although the clock frequency of the decoder is higher due to the bit-serial node logic, the high number of clock cycles per iteration limit the achievable decoder throughput and block latency. However, it is possible to introduce a bit-serial data flow in a more fine-grained way. A systematic architecture analysis is possible by breaking the decoder loop into four parts as shown in Figure 21, namely the bit and check node and the communication between the nodes in both directions. Now, possible architectures can be distinguished by assuming either a bit-serial or a bit-parallel approach in each of the four parts. Obviously, also a digit-serial approach is possible as discussed in [69]. Considering only a bit-serial or bit-parallel data flow, in total 16 different architectures are possible. As a first order metric of the decoder throughput, the number of clock cycles per iteration considering a message word length of  $w = 6$  is given. To avoid extensive routing-induced extensions of the silicon area, especially the highlighted architectures with a bit-serial communication in both directions should be taken into account. When comparing the number of clock cycles per iteration for these four architectures, the

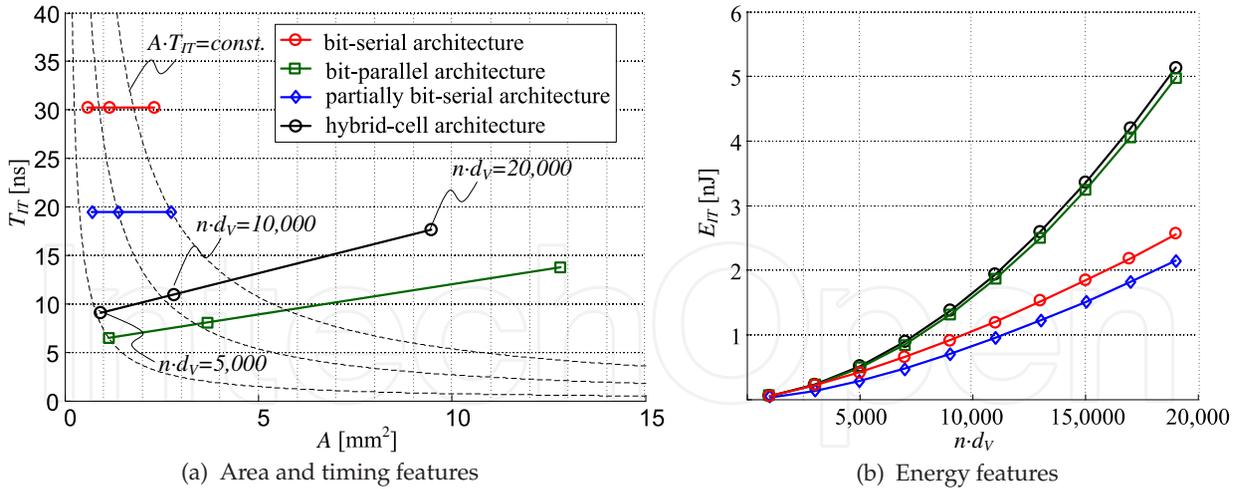


**Figure 22.** Partially bit-serial architecture

architecture with a bit-parallel bit node allows for the smallest number of clock cycles per iteration and, thus, promises the highest decoder throughput. As the bit-parallel realization of the bit node would result in a large silicon area and a long critical path, further optimizations on arithmetic level have to be done. Here, it is possible to gain from the bit-serial input data stream by realizing the multi-operand adder in the bit node bit-serially using an MSB-first data flow. Within each clock cycle a partial sum  $L^k(c, q)$  for the received bit-weight is generated which is accumulated subsequently to derive the new estimation  $L(Q_i)$  as is shown in the decoder loop in Figure 22. The long ripple path in the accumulator unit running over the complete word length can be reduced using a carry-select principle. For further details of the realization on arithmetic and circuit level refer to [62].

#### 4.1.3. Quantitative architecture comparison

The cost models have been adapted to the new architecture concepts to allow for a quantitative evaluation of the architecture design space. Figure 23(a) illustrates the resulting silicon area  $A$  and iteration period  $T_{IT}$  of the fully bit-parallel, fully bit-serial, hybrid-cell and partially bit-serial decoder architecture for three different code complexities  $n \cdot d_V = 5,000, 10,000$  and  $15,000$ . For all code complexities the new architecture concepts are Pareto optimal as they allow for a trade-off between silicon area and iteration period in comparison to the bit-parallel and bit-serial architectures. Considering small code complexities the decoder architectures with a bit-parallel interconnect show the smallest area-time (AT) product and, therefore, are most AT-efficient. Considering a specified decoder throughput, the hybrid-cell architecture is promising whenever the timing constraints cannot be met by using bit-serial approaches, as it reduces the silicon area significantly in comparison to the bit-parallel bit- and check-node architecture. The new partially bit-serial architecture features the smallest area-time product for all code complexities larger than 9,000. In comparison to the bit-serial architecture a significantly smaller iteration period with only a slightly increased area is achieved. The architectures with a bit-parallel interconnect are located further and further away from the curve representing the smallest achievable area-time product. Here, the timing advantage of the bit-parallel architectures vanishes for large code complexities. Figure 23(b) depicts the energy per decoding iteration  $E_{IT}$  of the four decoder architectures for different code complexities. The advantage with respect to energy of the decoder architectures with a



**Figure 23.** Comparison of area, timing and energy features ( $d_V = 6, d_C = 32, M_R = 4, w = 6, \lambda = 40nm$ )

bit-serial interconnect becomes apparent. Considering code complexities larger than 10,000, the energy per iteration of the bit-parallel decoder becomes more than twice as high as the partially bit-serial architecture. The latter allows for the smallest energy per iteration in the complete code complexity range. This emphasizes the efficiency of the new partially bit-serial architecture which allows for the smallest area-time product in a wide range of code complexities and the smallest decoding energy, simultaneously. This work has been supported by the German Research Foundation (DFG) under the priority program UKoLoS (SPP1202).

## 5. Conclusion

This chapter presented results, accomplished within the frame of the DFG priority program »Ultrabreitband Funktechniken für Kommunikation, Lokalisierung und Sensorik«. Focus was put primarily on the analysis and optimization of on-chip and chip-to-chip multi-conductor/multi-antenna interconnects. While we could show that special techniques of physical optimization, coding and signal processing can improve interconnect performance to a remarkable degree, it is expected that even higher performance is achievable in chip-to-chip communication, when multi-conductor interconnects are replaced by wireless ultra-wideband multi-antenna interconnects. Hereby, the signal pulses do not necessarily increasingly disperse as they travel along their way to the receiving end of the interconnect. The propagating nature of the wireless interconnect can make for a much more attractive channel for chip-to-chip communications. The primary goal has been the development of both theoretical and empirical foundations for the application of ultra-wideband multi-antenna wireless interconnects for chip-to-chip communication. Suitable structures for integrated ultra-wideband antennas have been developed, their properties theoretically analyzed and verified against measurements performed on manufactured prototypes. Qualified coding and signal processing techniques, which aim at efficient use of available resources of bandwidth, power, and chip area has been proposed. In addition, attention was given to the implementation of iterative decoding structure for LDPC codes. Detailed cost-models, which are based on signal flow charts and VLSI implementations of dedicated functional blocks

have been developed, which allow for an informative analysis of elementary trade-offs between throughput, required chip area, and power consumption. This work has been supported by the German Research Foundation (DFG) under the priority program UKoLoS (SPP1202).

## Author details

Nossek Josef A., Mezghani Amine and Michel T. Ivrlač  
*Institute for Circuit Theory and Signal Processing, Technische Universität München, Germany*

Russer Peter, Mukhtar Farooq, Russer Johannes A. and Yordanov Hristomir  
*Institute for Nanoelectronics, Technische Universität München, Germany*

Noll Tobias and Korb Matthias  
*Chair of Electrical Engineering and Computer Systems, RWTH Aachen University, Germany*

## 6. References

- [1] Russer P (1998) Si and SiGe millimeter-wave integrated circuits, Si and SiGe based monolithic integrated antennas for electromagnetic sensors and for wireless communications. *IEEE Transactions on Microwave Theory and Techniques* 1998: 590-603.
- [2] Johannes A. R, Kuznetsov Y, Russer P, Russer J.A, Kuznetsov K, Russer P (2010) Discrete-time network and state equation methods applied to computational electromagnetics. *Mikrotalasma revija*: 2-14.
- [3] Johannes A.R, Baev A, Kuznetsov Y, Mukhtar F, Yordanov H, Russer P (2011) Combined lumped element network and transmission line model for wireless transmission links. *German Microwave Conference (GeMIC)*: 1-4.
- [4] Felsen L.B, Mongiardo M, Russer P (2009) *Electromagnetic Field Computation by Network Methods*. Heidelberg: Springer-Verlag.
- [5] Russer P (2006) *Electromagnetics, Microwave Circuit and Antenna Design for Communications Engineering*. Boston: Artech House.
- [6] Russer J.A, Mukhtar F, Wane S, Bajon D, Russer P (2012) Broad-Band Modeling of Bond Wire Antenna Structures. *German Microwave Conference (GeMIC)*: 1-4.
- [7] Yordanov H, Ivrlač M.T, Mezghani A, Nossek J.A, Russer P (2008) Computation of the impulse response and coding gain of a digital interconnection bus. *23rd Annual Review of Progress in Applied Computational Electromagnetics*: 494-499.
- [8] Yordanov H, Russer P (2008) Computing the Transmission Line Parameters of an On-chip Multiconductor Digital Bus. *Time Domain Methods in Electrodynamics*: 69-78.
- [9] Yordanov H, Ivrlač M.T, Nossek J.A, Russer P (2007) Field modelling of a multiconductor digital bus. *Microwave integrated circuit conference, 2007*: 579-582.
- [10] Yordanov H, Russer P (2008) Computation of the Electrostatic Parameters of a Multiconductor Digital Bus. *International Conference on Electromagnetics in Advanced Applications 2007*: 856-859.
- [11] Yordanov H, Russer P (2010) Antennas embedded in CMOS integrated circuits. *Proceedings of the 10th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems: Electronics and Energetics 2010*: 53-56.

- [12] Yordanov H, Russer P (2009) Wireless Inter-Chip and Intra-Chip Communication. Proceedings of the 39th European Microwave Conference: 145-148.
- [13] Yordanov H, Russer P (2009) On-Chip Integrated Antennas for Wireless Interconnects. Semiconductor Conference Dresden (SCD) 2009.
- [14] Yordanov H, Russer P (2010) Integrated On-Chip Antennas Using CMOS Ground Planes. Proceedings of the 10th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems: 53-56.
- [15] Yordanov H, Russer P (2010) Area-Efficient Integrated Antennas for Inter-Chip Communication. Proceedings of the 10th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems: 401-404.
- [16] Yordanov H, Russer P (2008) Chip-to-chip interconnects using integrated antennas. Proceedings of the 38th European Microwave Conference, EuMC 2008: 777-780.
- [17] Yordanov H, Russer P (2008) Integrated on-chip antennas for chip-to-chip communication. IEEE Antennas and Propagation Society International Symposium Digest 2008: 1-4.
- [18] Yordanov H, Ivrlač M.T, Nossek J.A, Russer P (2007) Field modelling of a multiconductor digital bus. Microwave Conference, 2007. European. 37th European: 1377-1380.
- [19] Russer P, Fichtner N, Lugli P, Porod W, Russer A.J, Yordanov H (2010) Nanoelectronics Based Monolithic Integrated Antennas for Electromagnetic Sensors and for Wireless Communications. IEEE Microwave Magazine: 58-71.
- [20] Mukhtar F, Yordanov H, Russer P (2010) Network Model of On-Chip Antennas. URSI Conference 2010.
- [21] Mukhtar F, Yordanov H, Russer P (2011) Network model of on-chip antennas. Advances in Radio Science: 237-239.
- [22] Mukhtar F, Kuznetsov Y, Russer P (2011) Network Modelling with Brune's Synthesis. Advances in Radio Science: 91-94.
- [23] Russer J.A, Dončov N, Mukhtar F, Stošić B, Asenov T, Milovanović B, Russer P (2011) Equivalent lumped element network synthesis for distributed passive microwave circuits. Mikrotalasna revija: 23-28.
- [24] Russer J.A, Gorbunova A, Mukhtar F, Yordanov H, Baev A, Kuznetsov Y, Russer P (2011) Equivalent circuit models for linear reciprocal lossy distributed microwave two-ports. Microwave Symposium Digest: 1-4.
- [25] Russer J.A, Dončov N, Mukhtar F, Stošić B, Milovanović B, Russer P (2011) Compact equivalent network synthesis for double-symmetric four-ports. Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS), 2011 10th International Conference on: 383-386.
- [26] Cover T.M, Thomas J.A (1991) Elements of Information Theory. New York: John Wiley and Son.
- [27] Lok T.M, Wei V.K.W (1998) Channel Estimation with Quantized Observations. IEEE International Symposium on Information Theory 1998: 333.
- [28] Ivrlač M.T, Nossek J.A (2007) On MIMO Channel Estimation with Single-Bit Signal-Quantization. ITG Smart Antenna Workshop 2007.
- [29] Sotiriadis P.P, Wang A, Chandrakasan A (2000) Transition Pattern Coding: An Approach to Reduce Energy in Interconnect. ESSCIRC 2000: 320-323.

- [30] Kavčić A (2001) On the capacity of markov sources over noisy channels. IEEE Global Communications Conference 2001: 2997-3001.
- [31] Ivrlac M.T, Nossek J.A (2006) Chalanges in Coding for Quantized MIMO Systems. IEEE International Symposium on Information Theory 2006: 2114-2118.
- [32] Lyuh G, Kim T (2002) Low Power Bus Encoding with Crosstalk Delay Elimination. 15th Annual IEEE International ASIC/SOC Conference 2002: 389-393.
- [33] Mezghani A, Khoufi M.S, Nossek J.A (2007) A Modified MMSE Receiver for Quantized MIMO. ITG/IEEE WSA 2007.
- [34] Mezghani A, Nossek J.A (2007) Wiener Filtering for Frequency Selective Channels with Quantized Outputs. IEEE SSD 2007.
- [35] Max J, Quantizing for Minimum Distortion (1960) IEEE Trans. Inf. Theory 1960: 7-12.
- [36] Mezghani A, Khoufi M.S, Nossek J.A (2008) Spatial MIMO Decision Feedback equalizer Operating on Quantized Data. IEEE ICASSP 2008.
- [37] Jayant N.S, Noll P (1984) Digital Coding of Waveforms. New Jersey: Prentice-Hall 1984.
- [38] Proakis J.G (1995) Digital Communications. New York: McGraw Hill 1995.
- [39] Verdú S (2002) Spectral Efficiency in the Wideband Regime - IEEE Trans. Inform. Theory 2002: 1319-1343.
- [40] Mezghani A, Nossek J.A (2007) On Ultra-Wideband MIMO Systems with 1-bit Quantized Outputs: Performance Analysis and Input Optimization - IEEE International Symposium on Information Theory 2007.
- [41] Teletar E (1999) Capacity of Multi-Antenna Gaussian Channels. European Transactions on Telecommunications 1999: 585-595.
- [42] Schubert M, Boche H (2005) Iterative multiuser uplink and downlink beamforming under SINR constraints. IEEE Transactions on Signal Processing: 2324-2334.
- [43] Mezghani A, Joham M, Huger R, Utschick W (2006) Transceiver Design for Multi-user MIMO Systems. Proc. ITG Workshop on Smart Antennas 2006.
- [44] Joham M, Kusume K, Gzara M.H, Utschick W, Nossek J.A (2002) Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems. ISSSTA 2002: 9-13.
- [45] Ivrlac M.T, Utschick W, Nossek J.A (2002) On Time-Switched Space-Time Transmit Diversity. Proceedings of the 56th IEEE Vehicular Technology Conference 2002: 710-714.
- [46] Ivrlac M.T, Choi R.L.U, Murch R, Nossek J.A (2003) Effective use of Long-Term Transmit Channel State Information in Multi-user MIMO Communication Systems. Proceedings of the IEEE Vehicular Technology Conference 2003.
- [47] Widrow B, Kollár I (2008) Quantization Noise. Cambridge University Press.
- [48] Joham M, Utschick W, Nossek J.A (2003) Linear Transmit Processing in MIMO Communication Systems. IEEE Transactions on Signal Processing 2003.
- [49] Mezghani A, Khoufi M.S, Nossek J.A (2008) Maximum Likelihood Detection For Quantized MIMO Systems. In Proc. ITG/IEEE WSA 2008.
- [50] Mezghani A, Antreich F, Nossek J.A (2010) Multiple Parameter Estimation With Quantized Channel Output. In Proc. ITG/IEEE WSA 2010.
- [51] Lang T, Mezghani A, Nossek J.A (2010) Channel adaptive coding for coarsely quantized MIMO systems. In Proc. ITG/IEEE WSA 2010.
- [52] Lang T, Mezghani A, Nossek J.A (2010) Channel-matched trellis codes for finite-state intersymbol-interference channels. In Proc. SPAWC 2010.

- [53] Mezghani A, Ivrlac M, Nossek J.A (2008) On Ultra-Wideband MIMO Systems with 1-bit Quantized Outputs: Performance Analysis and Input Optimization. IEEE International Symposium on Information Theory and its Applications 2008.
- [54] Gray R.M (1990) Quantization noise spectra. IEEE Trans. Inform. Theory 1990: 1220-1244.
- [55] Gallager R. (1962) Low-density parity-check codes. Information Theory, IRE Transactions on, vol. 8, pp. 21-28.
- [56] Sae-Young C., Forney, Jr. G. D. , et al. (2001) On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit. Communications Letters, IEEE, vol. 5, pp. 58-60.
- [57] IEEE (2006) IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements Part 3: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications. IEEE Std 802.3an-2006 (Amendment to IEEE Std 802.3-2005), pp. 1-167.
- [58] IEEE (2005) 802.16e. Air Interface for Fixed and Mobile Broadband Wireless Access Systems. IEEE P802.16e/D12 Draft, oct 2005.
- [59] IEEE (2006) 802.11n. Wireless LAN Medium Access Control and Physical Layer specifications: Enhancements for Higher Throughput. IEEE P802.16n/D1.0, Mar 2006.
- [60] Galbraith R. L., Oenning T., et al. (2010) Architecture and Implementation of a First-Generation Iterative Detection Read Channel. Magnetics, IEEE Transactions on, vol. 46, pp. 837-843.
- [61] Korb M., Noll T. G. (2012) A Quantitative Analysis of Fixed-Point LDPC-Decoder Implementations using Hardware-Accelerated HDL Emulations. to appear in Embedded Computer Systems (SAMOS) International Conference.
- [62] Korb M. Noll T. G. (2011) Area- and energy-efficient high-throughput LDPC decoders with low block latency. ESSCIRC, Proceedings of the, 2011, pp. 75-78.
- [63] Mohsenin T., Baas B. M. (2006) Split-Row: A Reduced Complexity, High Throughput LDPC Decoder Architecture. Computer Design, International Conference on, 2006, pp. 320-325.
- [64] Zhang Z., Anantharam V., et al. (2009) A 47 Gb/s LDPC Decoder With Improved Low Error Rate Performance. VLSI Circuits, Symposium on, 2009, pp. 286-287.
- [65] Darabiha A., Carusone A. C., et al. (2007) A 3.3-Gbps Bit-Serial Block-Interlaced Min-Sum LDPC Decoder in 0.13- $\mu$ m CMOS. Custom Integrated Circuits Conference, 2007. CICC '07. IEEE, 2007, pp. 459-462.
- [66] Blanksby A. J., Howland C. J. (2002) A 690-mW 1-Gb/s 1024-b, Rate-1/2 Low-Density Parity-Check Code Decoder. Solid-State Circuits, IEEE Journal of, vol. 37, pp. 404-412.
- [67] Korb M. Noll T. G. (2010) LDPC decoder area, timing, and energy models for early quantitative hardware cost estimates. System on Chip (SoC), 2010 International Symposium on, 2010, pp. 169-172.
- [68] MacKay D. <http://www.inference.phy.cam.ac.uk/mackay/codes/data.html>
- [69] Korb M., Noll T. G. (2009) Area and Latency Optimized High-Throughput Min-Sum Based LDPC Decoder Architectures. ESSCIRC, Proceedings of the, 2009, pp. 408-411.

- [70] Fossorier M. P. C., Mihaljevic M., et al. (1999) Reduced complexity iterative decoding of low-density parity check codes based on belief propagation. *Communications, IEEE Transactions on*, vol. 47, pp. 673-680, 1999.

IntechOpen

IntechOpen