KU LEUVEN

(article begins on next page)

Where analog meets digital

Analog-to-Information conversion and beyond

Marian Verhelst and Ahmad Bahai

Energy efficiency, long battery life and low latency are some of the key attributes of many emerging ultra-low power sensing and monitoring systems. Applications such as always-on reactive sensor systems for natural human-device interfaces and IoT for consumer and industrial applications require ultra-low power designs beyond the promises of state of the art data converters.

These devices demand for a new approach to analog-digital system partitioning with the goal of significant overall reduction in energy consumption. Many IoT applications, unlike most multimedia systems, require signal information extraction or signature extraction, rather than full reconstruction of the original sensed waveforms. Under these conditions, Nyquist rate sampling may no longer offer the optimal digitization scheme. Recent work on alternative sensor digitization strategies target drastic sampling rate reduction in the ADC, while preserving the valuable relevant information (knowledge) present in the sensed signal. This paper aims to give an overview of this emerging field of analog-to-information conversion in light of various sub-Nyquist sampling techniques recently appearing in literature, as well as highlight new opportunities, challenges and applications emerging by such converters.

I.   Nyquist rate vs. Information rate:

Over the last several decades, a growing number of signal processing architects have embraced intensive digital signal processing preceded by a standard analog frontend and analog-to-digital converter. This trend has been exacerbated by the exponential rate of miniaturization in silicon, growing complexity of signal processing algorithms, and more systematic digital design and technology porting compared to analog design in deep submicron technology nodes. The interface between analog and digital signals has as such generally been governed by sampling at or above the Nyquist sampling rate of the analog waveforms. The dimensionality of a bandlimited signal $f(t)$ with a physical bandwidth W over a period of T is $2WT$, indicating the number of samples sufficient for perfect digital signal reconstruction. Such sampling at the Nyquist rate of $2W$ ensures the integrity of the signal represented by samples which are Fourier series coefficients in a Fourier series expansion of function $F(w)$ over fundamental interval $[-W \ W]$ [1]. The original signal can subsequently be reconstructed by superimposing a set of orthogonal basis functions (sinc functions) weighted by the samples $f(nT)$. Sampling the incoming signal at this rate hence guarantees that no information about the incoming signal is lost *without* taking into account *any heuristic* or *a priori* side information about the signal or its information content other than the physical bandwidth. While sampling at or above Nyquist rate offers a classic and straightforward approach, it can compromise overall power efficiency.

Many emerging sensing applications, such as reactive user interfaces, sensors for the internet-of-things (IoT), medical monitoring systems, or radar applications, evolve around sensing natural signals, whose physical bandwidth is much higher than their actual information rate (Figure 1). In other words, *a-priori* information on the sensed signal can be exploited to reduce the information rate well below the physical bandwidth. Examples are heartbeat signals, or the reception of reflected pulses in an ultrasound system (see inset 1). This *a-priori* information can take various forms, such as the shape, periodicity, or the sparseness of the sensed signal. Taking this *a-priori* information into account, reduces the effective information rate of the received signal well below the theoretical Nyquist rate. In theory, the signal can now be sampled at this lower rate, while preserving full information conversion into the digital domain. Yet, in practice it is not necessarily straightforward to achieve sampling rate reduction all the way down to the information rate, as pursued by analog-to-information converters.

FIRST INSET: Many emerging sensing applications, such as reactive user interfaces, sensors for the internet-of-things (IoT), medical monitoring systems, or radar applications, evolve around sensing natural signals, whose **physical bandwidth** is much higher than their actual information rate . For example, a pulsed radar signal which consists of sparse pulses in the time domain with a known pulse shape, can be completely defined by only characterizing the amplitude and position of the pulses. Their **information rate** is hence much smaller than their physical Nyquist bandwidth (Figure 1).

In addition, some applications are not even concerned with the complete information rate , but only interested in a selected subset of features that can be extracted from the signals. This can, for example, be the amplitude of the pulses. As a result, the relevant **feature rate** of the signal can again be smaller than its information rate.
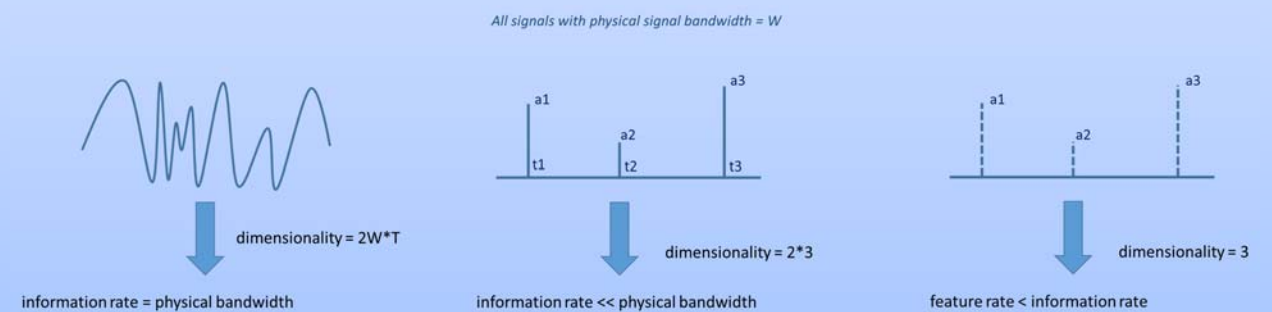
All signals with physical signal bandwidth = W

dimensionality = 2W*T
information rate = physical bandwidth

a1   a2   a3
t1   t2   t3
dimensionality = 2*3
information rate << physical bandwidth

a1   a2   a3
dimensionality = 3
feature rate < information rate

Figure 1: Physical bandwidth vs. information rate vs. feature rate

II.   Alternative sampling techniques
A. Beyond Nyquist through analog-to-information conversion:
The term "analog-to-information" emerged with the introduction of a sub-Nyquist sampling technique called compressed sensing (CS) for sparse signals [2, 3, 4]. Sparsity means that a signal is compressible and can be represented with fewer samples on an appropriate basis [5]. CS exploits the fact that the information rate of a waveform that is sparse in a particular domain (such as e.g. in the time domain, in

the frequency domain, or in a wavelet domain) is significantly smaller than the Nyquist rate [6]. By correlating the signal with waveforms which are not coherent with the sparse basis, the analog bandwidth is narrowed down to near the information rate. Such projections can be implemented in the analog domain through non-uniform sampling, as well as through various modulate-and-integrate schemes. Subsequently, the original waveform can be recovered in the digital domain with signal processing of far fewer samples through finding sparse solutions to an underdetermined linear system [5]. The resulting analog-to-digital convertor architecture, depicted in Figure 3.b, uses a-priori knowledge of the signal, in terms of the basis in which the signal is sparse. This allows the ADC to trade-off sampling rate reduction against additional analog and digital complexity for analog random basis projection and digital signal reconstruction. A large part of CS theory deals with the optimum choice of undersampling rate and choice of incoherent basis functions. In some applications, feasibility of achieving up to 1 order of magnitude sampling rate reductions have been demonstrated in imager, radar, spectrum sensing, and biomedical systems [7, 8, 9, 10]. While full signal reconstruction comes with a very large computational load, interesting emerging work involves the direct extraction of features in the digital domain from the compressed signal without prior full signal reconstruction. This has also successfully been applied to visual object tracking [11], and power spectrum determination of unoccupied bands in efficient spectrum sensors [7]. Yet, it is important to note that CS-based sampling techniques strongly rely on signal sparsity to avoid information loss, and are as such not suitable for arbitrary signals. Moreover, impact of circuit impairments, clock jitter, noise folding and the complexity of the required digital signal processing, diminishes benefits of compressed sensing for signals with large dynamic range and bandwidth [11].

B. Finite innovation rate sampling:

A slightly different approach to sub-Nyquist sampling, exploiting a-priori information for improved sampling rate reduction, is 'finite innovation rate sampling' [14, 15, 16]. This technique allows to efficiently sample signals that have a finite number of degrees of freedom per unit time, such as pulse trains (see 1st inset), or piece-wise polynomials using a smoothing kernel. These signals hence are not necessarily bandlimited or sparse in the time or frequency domain. Yet, the selection of an adequate smoothing kernel is typically needed to filter the analog waveform, such that the required analog sampling rate is reduced to the degrees of freedom of the smoothed signal, called the 'rate of innovation' of the signal. This allows reconstruction of a pulse train, for instance, with unknown pulse amplitudes and timing while only sampling at twice the pulse rate, hence orders of magnitude lower than the actual pulse bandwidth (Figure 3.c). Similar to the CS approach, innovation rate sampling [14, 16] attempts to recover a faithful approximation of all information present in the original signal, which is achieved through extensive digital post-processing. This technique is currently being applied towards its first hardware realizations and promises to offer significant benefits in various applications, such as biomedical imaging or radar applications.

Aforementioned analog-to-information converters have recently gained increased attention, and demonstrated applicability in a wide range of application domains where perfect signal reconstruction or complete information retrieval in the digital domain is desired. By exploiting a-priori knowledge of the signal they reduce the information rate below the Nyquist bandwidth without loss of information, yet often at the cost of a considerable increase in digital signal processing complexity [15]. Interesting work

harmonizing these diverse analog-to-information techniques in a general framework, is the Xampling framework [14], based on sampling Union of Subspace models.

SECOND INSET: Analog-to-information sampling techniques pursue ADC sampling rate reduction *without losing any information* present in the analog waveform, hence targeting lossless compression at the analog-to-digital interface to maintain full signal reconstruction capabilities in the digital domain. However, in applications where neither the entire content of the original data, nor its full reconstruction is of interest, the data converter can target the conversion of a specific subset of information extracted from the waveform, called "features". This opportunity for lossy signal compression samples exploits the gap between a signal's information rate and its feature rate.
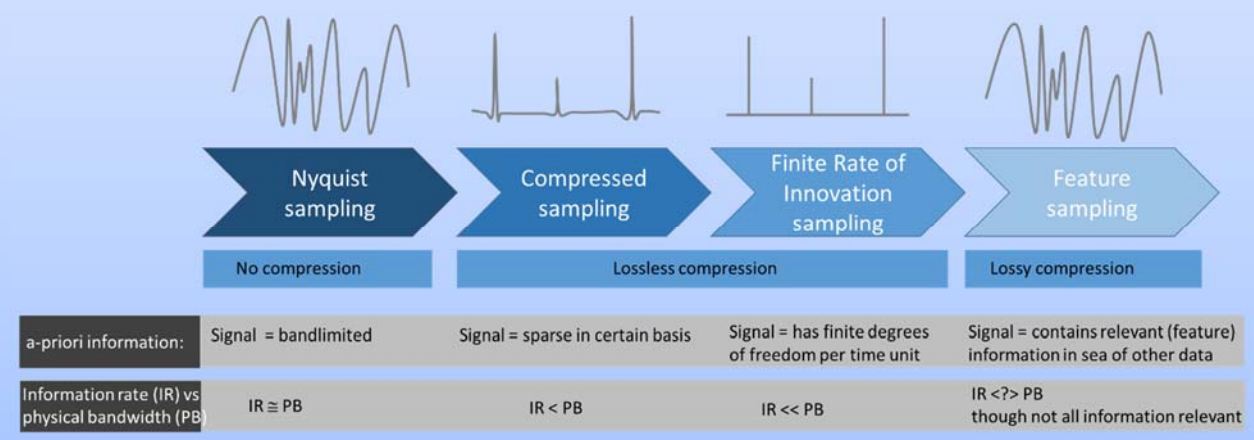
a-priori information:	Signal = bandlimited	Signal = sparse in certain basis	Signal = has finite degrees of freedom per time unit	Signal = contains relevant (feature) information in sea of other data
Information rate (IR) vs physical bandwidth (PB)	$IR \cong PB$	$IR < PB$	$IR << PB$	$IR <?> PB$ though not all information relevant

Nyquist sampling — No compression
Compressed sampling / Finite Rate of Innovation sampling — Lossless compression
Feature sampling — Lossy compression

Figure 2: Comparison of sampling techniques based on their information preservance.

C. Feature extracting ADCs through analog analytics:

Analog to Information techniques as mentioned earlier attempt to reduce the sampling rate without loss of information present in the analog waveform, hence target lossless compression at the analog-to-digital interface to maintain full signal recovery capabilities in the digital domain. However, in many natural signals and sensor applications such as visual, acoustic and medical monitoring systems, the signal is not necessarily sparse nor has finite degrees of freedom, due to corruption from noise, interfering signals or circuit impairments. As such, standard sub-Nyquist sampling techniques are not applicable. Furthermore, in many of these applications neither the entire information content of the original signal, nor its full reconstruction is typically of interest. Instead, many applications require a specific subset of information extracted from the waveform, called "features", such as the maximum signal level over a period of time, the number of zero-crossings, etc. The significance of lossy signal compression is particularly paramount for signal classification and pattern recognition applications. Such signal processing techniques are used extensively in many IoT and wearable devices, such as speech recognition, gesture detection systems, heart rate monitors, etc, where the information rate of the signals considerably exceeds the *relevant* information rate. (See Figure 1 and 2nd inset.)

An emerging class of ADCs, which we will denote by *feature extracting ADC's*, does not convert the complete signal into the digital domain, nor relies on signal sparsity. Instead, they only target to sample the signal at its *relevant Information* rate, termed the *feature rate*. This is achieved through extracting a specific set of features which are embedded in the analog waveforms. By combining analog signal processing and data conversion, the signal is first projected onto a specific feature space, after which conversion at the feature rate takes place (See Figure 3.d). This allows the signal processing to exclusively focus on feature-bearing information, and discard irrelevant information as early in the signal chain as possible. The signal's projection or transformation (linear or nonlinear) into the feature domain is achieved through a feature enhancing filter, boosting the relevant signal features, while suppressing other irrelevant information or distorting interferers. By discarding irrelevant information as early in the signal processing chain we can significantly improve overall system energy efficiency. This of course implies moving the boundaries between analog and digital, requiring more intelligent analog signal processing prior to sampling. This analog processing, denoted as *analog analytics*, should emphasize relevant features and reduces the dimensionality of the waveform through a "feature preserving" transformation with the intention of classifying features instead of reconstructing the original waveform. The ultimate goal is to sample the signal as close as possible to the Nyquist rate of the relevant information present in the incoming waveform, or *the feature rate*, as this would offer the ultimate feature - power efficiency trade-off.

THIRD INSET: Many novel ADC sampling strategies have emerged over the last decade, targeting a significant reduction in sampling energy consumption compared to traditional Nyquist rate conversion (Figure 2.a), by exploiting a-priori signal information. To this end, compressed sensing (Figure 2.b) and innovation rate sampling (Figure 2.c) try to reduce the sampling bandwidth as close as possible to the signal's information rate. Feature sampling ADCs (Figure 23.d) reduce the dimensionality of the waveform through *analog analytics* in order to retain only application-relevant signal features, with the intention of classifying these features instead of reconstructing the original waveform.
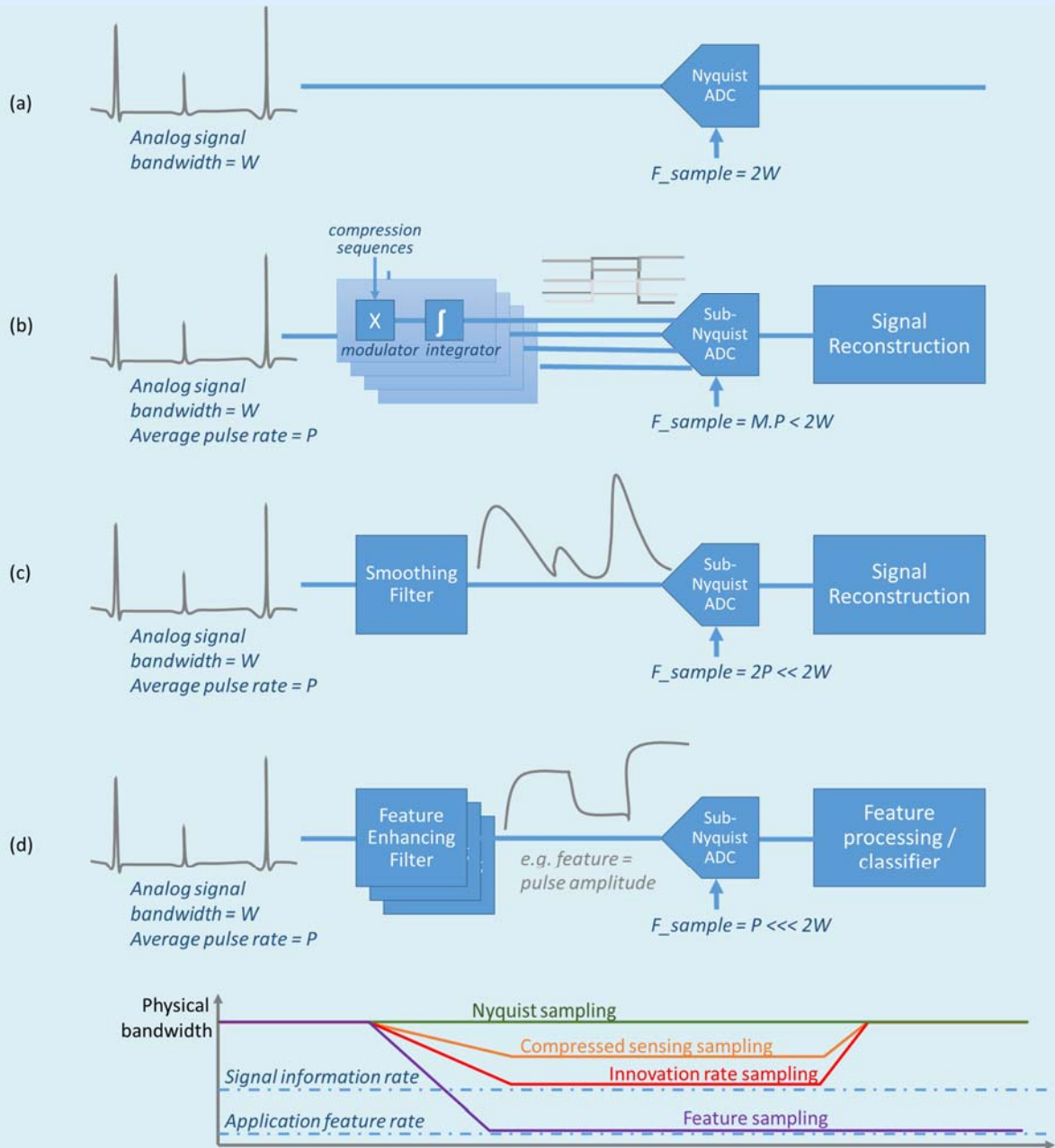

Figure 3. Comparing sampling techniques. (a) Nyquist rate sampling; (b) Compressed sensing sampling (c) Innovation rate sampling; (d) Feature sampling using analog analytics

Sub-Nyquist sampling and analog analytics has implicitly been exploited since long in digital communication systems. Also in such systems, the ultimate goal is the integrity of data (not signal) transmission over communication channels plagued by noise and interference. Sub-Nyquist sampling, in this case, can be tolerated as long as signal distortion does not corrupt the extracted features (communicated data symbols). Projection into the feature space and resulting sampling bandwidth reduction is achieved in the analog domain by boosting relevant signals while suppressing noise and filtering away irrelevant interfering signals. The resulting signal is then sampled at its feature rate. As shown in Figure 4, a typical waveform such as a root raised cosine has a bandwidth larger than 1/2T, yet sampling at the feature rate 1/T is theoretically sufficient to extract the transmit data. Clearly this sampling below Nyquist rate results in signal distortion, as shown in Figure 5, yet the relevant information present at sampling instants with interval k/T is preserved.
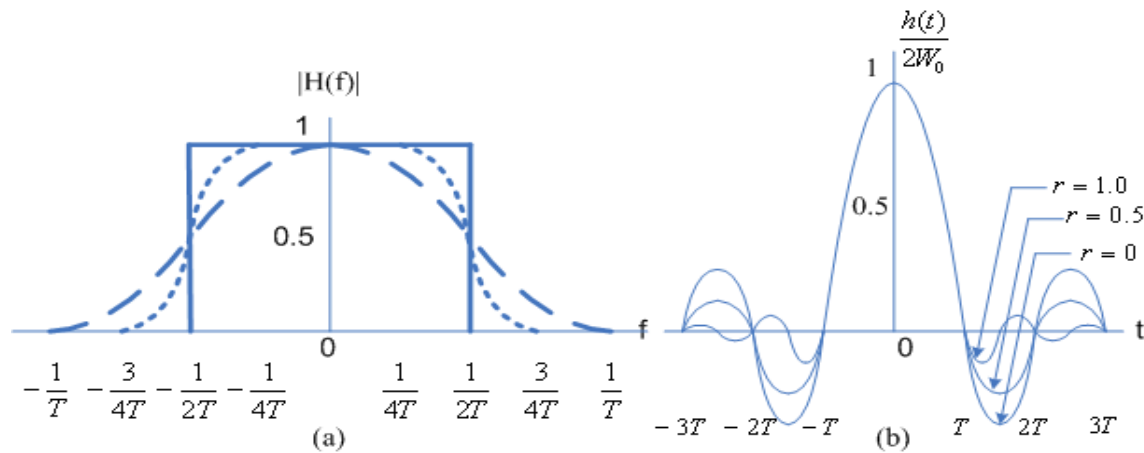

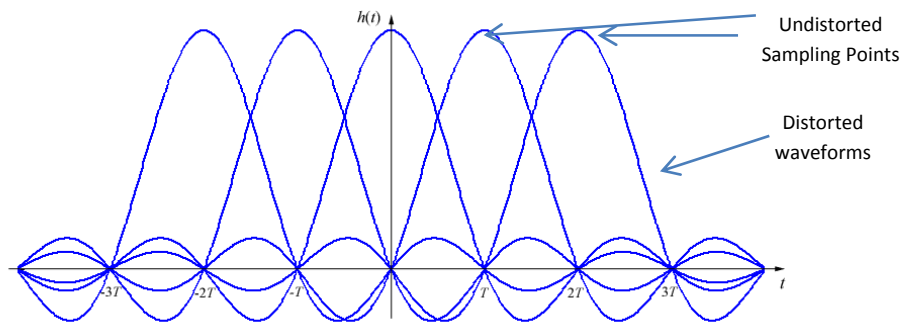Figure 4: Root Raised cosine in frequency and time domains

Figure 5: Zero ISI condition

The more significant and enormous bandwidth saving opportunities for sensor applications facilitated by feature sampling can be clearly illustrated with an example from the speech processing domain: Speech detection and speech analysis are often done using features representing the averaged energy-content of

exponentially-spaced bandpass frequency bands (called mel-scaled frequency bands) as a coarse form of spectrum analysis. As the speech signal, typically corrupted by background noise, is not sparse, nor has finite degrees of freedom, standard sub-Nyquist sampling techniques are not appropriate. Yet, significant sampling rate reduction can be achieved by introducing a feature sampling ADC. For example, voice activity detection can be implemented by extracting features in the analog domain representing the energy profile of mel-scaled frequency bands, averaged across 20msec frames [16]. Good speech detection performance has been demonstrated using 8 of such mel-frequency features, coarsely computed in the analog domain. The resulting feature rate hence corresponds to 8/20msec = 400Hz, or more than an order of magnitude below the Nyquist rate of the audio signal. Hence, the speech signal information bands are enhanced through analog analytics, while distorting background noise is discarded (if out-of-band) or suppressed (if in-band).

Clearly, feature extracting ADCs enable drastic sampling rate reduction, beyond what is possible in traditional lossless analog-to-information converters. It is important to note that the feature extracting ADCs exploiting analog analytics, are not suggested as a replacement for classical converters. In many applications where reconstruction of a signal is required, such as multimedia applications, a standard Nyquist approach of Figure 3.a is still required. Feature extracting ADCs are most appropriate for applications which do not involve reconstruction of original signals, or can function as a smart wake-up front-end to such systems. Yet, these new opportunities also come with many remaining challenges, both at the systems and circuit level, which will be discussed in Section III and IV.

III.   System challenges and opportunities of feature extracting ADCs:

A typical system architecture for feature extracting data converters is represented in Figure 6. One or more feature channels project the analog input onto a feature enhancing basis. Subsequently, (a subset) of all feature channels is scanned and sampled into the digital domain for further processing and potential signal classification. This new paradigm comes with a set of new system challenges. First of all, the choice and design of the analog feature enhancing filters is not straightforward, as yet generic and flexible analog feature extraction has not been realized. Moreover, unlike most standard data conversion approaches where each block (ADC, DSP,…) is independently evolving, this new approach does require a system optimization to realize all the benefits of performance and power efficiency, and hence is mostly application specific. Nevertheless, several system-level techniques and architectural opportunities can be exploited across many applications. We will discuss them in detail, and illustrate each aspect with aforementioned voice-activity detection example.
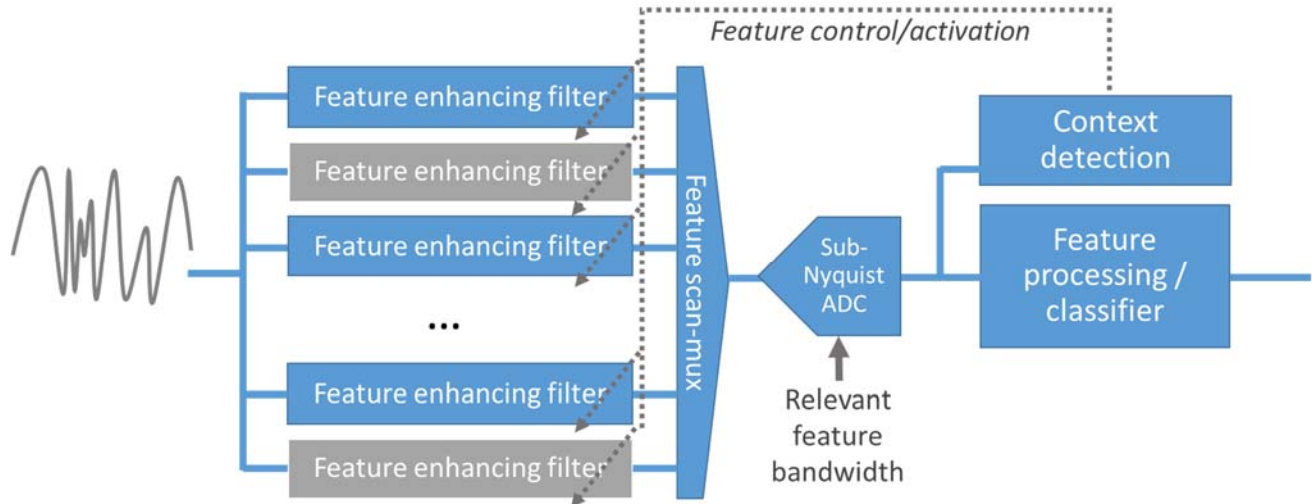
Figure 6: Feature extracting ADC architecture

A. *Choice of feature enhancing filters:*

A crucial parameter determining the accuracy of the classifying sensor interface under noise and distortion, as well as its power efficiency, is the choice of the feature enhancing filters. The optimal feature enhancing filter set maximally spreads information bearing data, while suppressing irrelevant distortion data. At the same time, this set consists of the minimal filter set necessary to achieve the desired performance, such that the power consumption footprint is minimized. Inspiration can be drawn from techniques commonly applied in the domain of machine learning, yet care must be taken regarding their implementation through analog analytics.

Feature learning is a well-studied task in the machine learning community. The goal is to automatically discover a good sub-space representation of the data to be analyzed. In contrast to heuristic, manual feature design, where domain-specific expert knowledge is exploited to handcraft features, feature learning targets the optimization of an objective function that captures the goodness of the features [17]. Techniques such as principal component analysis (PCA) [18] and deep learning [19], automatically reveal the most informative portions of the incoming waveforms, resulting in demonstrated improved classification accuracies relative to standard features. However, all these approaches do not take the analog implementation challenges of the generated features into account, and typically result in very complex feature generation networks which cannot be mapped on power efficient analog analytics.

A second criteria of utmost importance in feature extraction is the ease of implementation in the analog domain, and sensitivity to circuit impairments. A widely used type of feature fitting these requirements are statistical metrics directly extracted from raw sensor data, in a frame-by-frame, or sliding window approach [20] [21]. Statistical metrics frequently applied in sensor classification include mean, standard deviation, energy, zero crossings and correlation coefficients in the time domain, bandpass energy (or coarse Fourier coefficients) in the frequency domain or wavelet transform. For each of these features, an analog implementation, configurable along several parameters, such as feature precision, window length,

etc. can readily be derived. A study in [20] related to activity detection illustrates the information quality of these feature types.

A third important aspect to the selection of the feature enhancing filters is the complexity and variety of features that can be computed. The optimal feature enhancing front-end, is programmable, and can extract a broad variety of complex features, rendering the front-end reusable across diverse sensing applications. Yet, this is still an open challenge. The difficulty of introducing programmability and challenges of implementing efficient memory elements in the analog domain currently limit feature enhancing ADCs to a subset of relatively simple features, such as the statistical metrics described above. Yet, due to their low power cost, many of such features can be combined, rendering high classification accuracy at low power-cost.

Starting from such an extensive set of implementable features, machine learning techniques for dimensionality reduction and feature selection using mutual information criteria [22] can be exploited to select the minimal subset of features achieving the targeted detection quality (Figure 7). A front-end with the derived set of (configurable) features can subsequently be implemented. This design is, at the moment, application specific with limited versatility and reuseability. An Interesting future challenge for analog analytics is the design of programmable feature extracting front-ends capable of extracting a generic set of features, rendering them more widely reusable and configurable across many applications or various operating environments, as discussed in Section III.B.

Example: Voice activity detection has proven benefits of statistical feature extraction through analog analytics. Voice activity detectors based on mel-scaled mean energy features [23], Fourier coefficients [24], and zero-crossing frequency [25] have already been implemented. These voice activity detectors demonstrate good detection accuracy with at an energy cost which is orders of magnitude lower than their classical digital counterparts. Such a feature extracting voice activity detector can hence serve as the perfect low-power always-on wake-up sensor for a more complex digital processing system. These systems are typically tuned during their training phase to have very low miss detection probabilities, at the expense of an increased false detection rate, to avoid impacting the system's sensitivity.
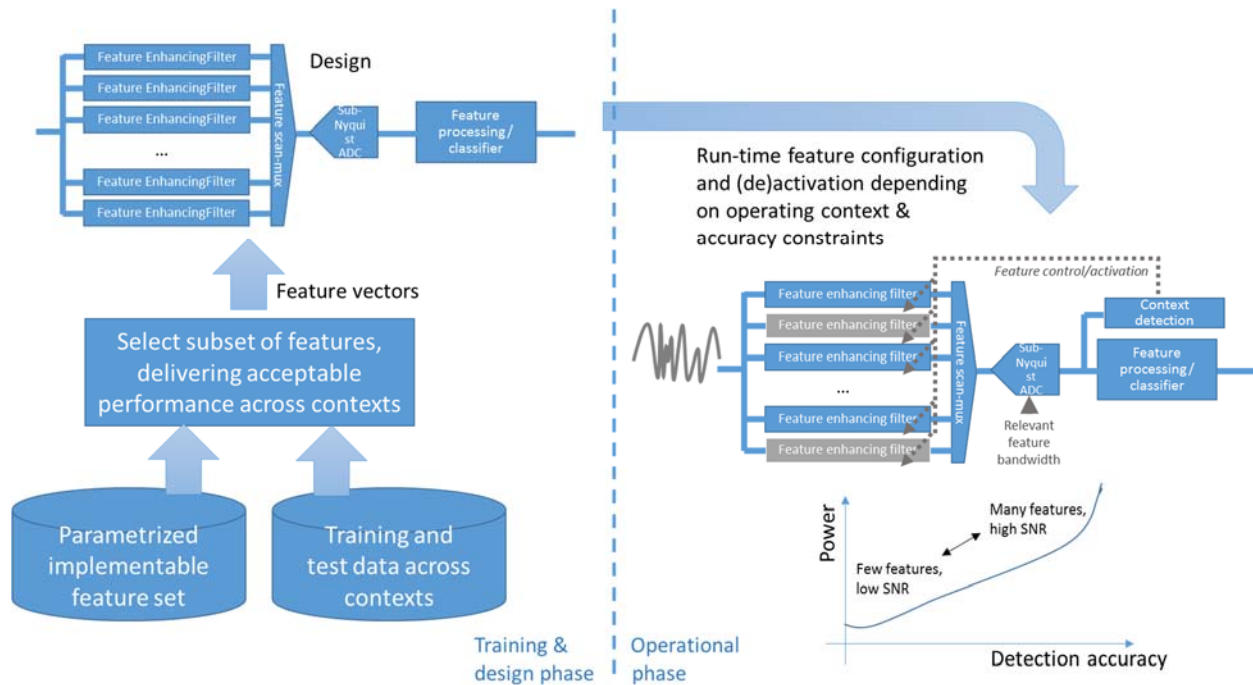
Figure 7: Feature ADC training principle and operational principle

B. *Dynamically (de)activating feature channels:*

An important challenge is the fact that in most application domains, there is neither a single feature, nor a limited feature set with acceptable performance across all operating circumstances. For instance, in order to achieve a good voice activity detection accuracy under various types of background noises (street noise, babble noise, subway noise, etc.), many of the analog features have to be observed in parallel [26], resulting in a large power consumption footprint. Also the study in [20], related to activity detection, points out varying optimal window lengths and feature types across operating contexts. A static implementation of the super-set of all relevant features extractors on the chip which are sampled continuously, would significantly diminish the power consumption benefits of the feature extracting ADC. Yet, the current limited analog programmability also prevents the implementation of a single feature extraction filter, which can be completely reprogrammed on the fly. An interesting and proven alternative is to implement a diverse set of parallel feature extractors with limited programmability in the analog domain, yet only activate the subset of features most beneficial under specific operating conditions at any given moment in time.

As illustrated in Figure 7, this approach relies on a set of configurable feature filters implemented on the chip. These filters can be configured at run-time along several parameters, such as window length, gain, bandwidth, etc. At run-time, a context detection block determines the current operating context, and hence the optimal analog feature set. A run-time configuration manager subsequently only activates and configures the relevant feature set in the analog domain. As such, the implemented feature sampling system truly achieves the goal to only spend resources on information bearing data, and discard all other irrelevant data as early as possible.

Example: This principle has been implemented in a voice activity detector in [23]. The voice activity detector can extract the energy content in 16 different mel-scaled frequency bands between 50 and 4000Hz, with configurable gain and window length. Across various background noises (street noise, babble noise, etc.) different frequency bands are most relevant regarding speech vs. no speech discrimination (Figure 8). This stems from the different frequency profiles of the distorting background noise. The feature selection strategy introduced in [26] selects features based on information content, as well as power consumption footprint, to maximize detection accuracy under a power constraint (or vice versa). The resulting run-time feature (de)-activation saves up to one order of magnitude in power consumption by only activating the most distinctive frequency bands, as illustrated in Figure 8. This front-end hence realizes efficient and programmable feature extracting, its programmability is limited to a particular type of feature, being energy per frequency subband.
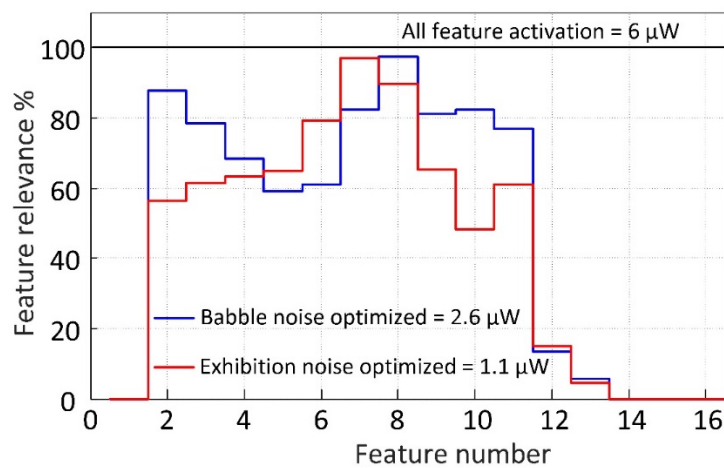

Figure 8: Feature relevance varies in function of operating context. Resulting power saving opportunities from feature deactivation depicted.

C.   *Power-performance scalability though flexible analog analytics*:

The configurable feature extraction implementation through limited programmablity "*analog analytics*" enables more than just maintaining detection performance across various operating contexts at low power consumption. Due to the analog-centric implementation, the configurability can be exploited further towards efficient run-time power – accuracy scalability. As studied extensively by Vittoz [27], Sarpeshkar [28] and others [29], analog power consumption shows a much more pronounced dependency on the required signal to noise ratio (SNR) compared to digital power consumption, which is only logarithmically dependent on SNR requirements. As a result, dynamic accuracy scalability, which is not very effective in the digital domain, does offer opportunities in analog analytics (See 4th INSET).

FOURTH INSET: Vittoz, and later Sarpeshkar, analyzed the power consumption footprint of analog and digital processing systems [27] [28]. A clear cross-over point between the analog and digital power consumption was identified, with analog implementation benefits for systems with low-to-medium range SNR requirements (up to about 10 bits ENOB). Since this study, digital power consumption has benefited more profoundly from silicon technology scaling. However, in the context of feature extracting ADCs, the power consumption of the "*analog analytics*" benefits in a similar way from technology scaling due to improved digital enhancement techniques (see Section IV) and dynamic feature selection. As a result, the cross-over point between analog and digital does not seem to be shifted drastically in advanced CMOS technologies (Figure 9). Low-to-medium range SNR processing, as required in typical classifying sensor interfaces, still benefits from analog analytics.

A second important observation, is that the analog power consumption scales much better (proportional) with the precision requirements, than its digital counterpart (logarithmic). This illustrate an additional opportunity for analog analytics: dynamic accuracy-power scalability.
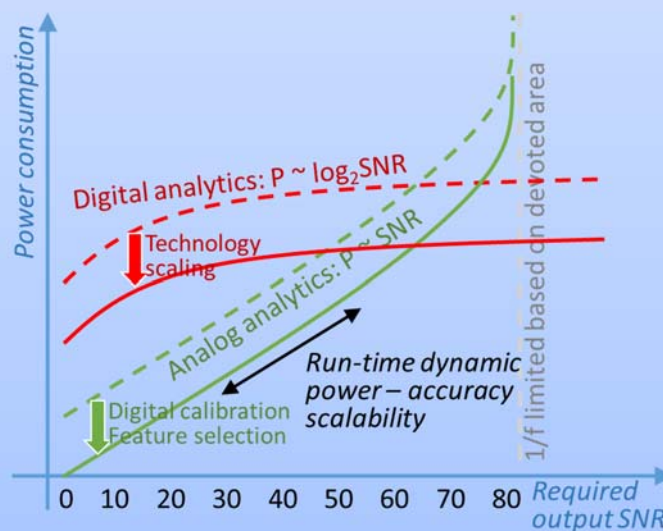

Figure 9: Analog and digital power consumption trends in function of required SNR

A feature extraction ADC can exploit power-vs-accuracy scalability along two axis. On one hand, the system can dynamically activate and deactivate features to increase the feature rate at the expense of additional power consumption, as discussed earlier. In parallel, it can modify the accuracy settings of every analog feature processing block, which also results in power-vs-accuracy scalability. Note that again feature rate is here not defined by the amount of information present in the incoming signal, but instead as the amount of *relevant* information (features) transferred into the digital domain. The result is a highly-linear, very-wide dynamic range achievable power scalability, which can by tuned towards the desired detection accuracy. Again, a smart, power-aware feature selection scheme, such as in [26] is required to select the most appropriate feature set at run-time.

In summary, feature extracting ADCs can exploit various new system level opportunities compared to digital-centric implementations, such as low-cost extraction of a wide range of simple statistical features, improved power-accuracy scalability and power savings through dynamic feature (de)activation. While the realization of a common front-end for generic analog feature extraction is still out-of-reach, large gains have already been achieved with application specific solutions [31]

IV- Implementation challenges and opportunities of feature extracting ADCs:
Feature extracting ADCs exploiting analog analytics promise significant benefits at system level yet introduce new implementation challenges compared to digital-centric solutions. This section will address some of the key design challenges of analog analytics and highlight opportunities for further investigation.

A. *Implementation challenges of analog analytics*

Feature extraction in analog analytics involves additional analog signal processing prior to digitization. Analog circuits are constrained by noise and accuracy requirements which do not necessarily benefit from voltage scaling and in most cases suffer from lower supply voltages [34]. The key parameters for a robust analog design are in broad categories of design parameters such as transistor geometries, process manufacturing parameters, and operational parameters such as temperature [35]. In a typical high performance analog design, traditionally a combination of meticulous layout and floor planning, careful circuit topologies such as fully differential architectures and accurate device modelling are critical to ensure robustness against device mismatch as well as operating condition and process variations. This becomes more challenging in finer geometry process nodes, which increasingly suffer from reduced matching quality for minimum feature size transistors and shrinking of voltage headroom. This problem worsenss when trying to introduce more flexibility of programmability into the analog analytics blocks, requiring more transistor stacking and operational robustness across multiple circuit configurations. Combatting these impairments in the traditional way could seriously compromise power efficiency and/or die size, or otherwise result in additional distortion to the computed features.

B. *Digital enhancement techniques in feature extracting ADCs*

In order to mitigate some undesirable attributes of deep submicron technology nodes for robust analog design, active and passive components matching using digitally assist techniques have increasingly been used in various analog circuits. Data converters, in particular, have extensively benefited from background and foreground digital calibration and compensation [35] for various reasons: Affordability in cost, die area, and power, as well as availability of digital gates for mixed signal design in deep submicron technologies which have promoted application of digital compensation in data converters. Most common approaches to digital calibration, specifically in data converters, as described in [34]: tightly coupled closed loop digital calibration which in most cases utilizes the embedded DAC, as well as redundancy techniques which involve post-processing to correct for imperfections at block or system level. Digital assist calibration circuits such as compensating DACs, typically run much slower than Nyquist rate, hence not imposing significant power penalties.

C. *Impairment mitigation techniques unique to feature extracting ADCs*

In a typical data converter and most analog-to-information converters described in Section II, analog circuit design performance needs to stay within tolerance limits of intended system parameters, such as linearity, offsets, etc. This is however not the case in feature sampling ADCs, whose design differs at two fundamental ways:
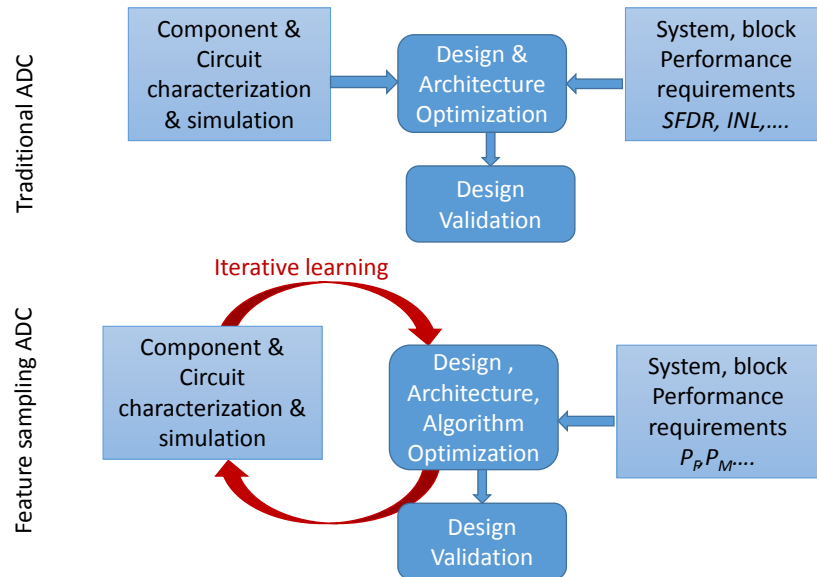
The first key difference is the performance metrics which matter in applications where feature sampling ADCs are deployed. The target system performance parameters are no longer the amount of noise or distortion added to the sampled signal and bit integrity, but rather the system-level probability of false alarms (P_F) and probability of miss detections (P_M) in the system's classification application (Figure 10). As such, the main task of the analog analytics pre-processing is not to extract the features in undistorted form, but to significantly reduce the bandwidth and enhance features for maximal distinction between classes of interest, as illustrated in Figure 11. Therefore, some distortion can be tolerated as long as it does not impede the system's classification performance. This can be exploited by driving the digital enhancement techniques from the classifier's output. In Figure 12.a a typical background digital calibration of a feature sampling ADC is shown. The error term, unlike regular data converters, is derived from classification metrics. These classification error terms can either be obtained based on training sequence inputs which are time interleaved with the incoming analog signal, or without a training signal, based on parallel sporadic accurate conversion and classification of the signal of interest.
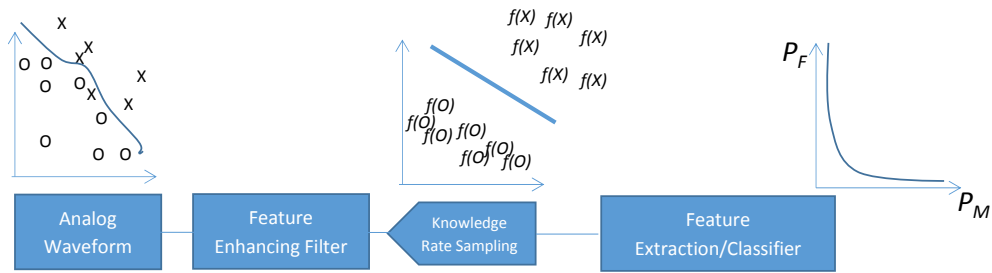
Figure 11: Analog analytics in feature sampling ADCs targets to enhance features towards optimal classification
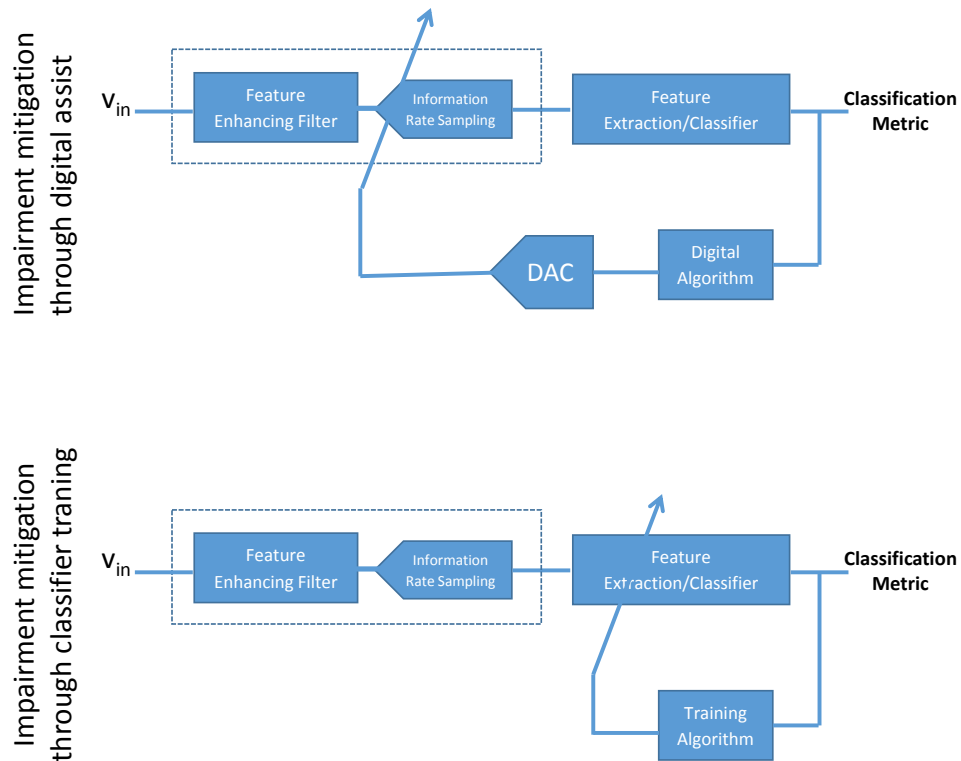

Figure 12: Run time background impairment mitigation in analog analytics

The second key difference, is the iterative adaptive learning in the classification system following the feature extracting ADC. Since classification algorithms such as neural networks of decision trees relay on iterative learning and adaptive threshold setting, they can further enhance tolerance to imperfections in the analog frontend. As the classifier learns the classification features, the neural weights, or the decision thresholds with the analog front-end in the loop, attempts to take into account the imperfections during the training phase, Figure 12.b. Non-linearity, offsets, frequency shifts, and other distortions will as such be absorbed in the trained classifier, and have limited impact on the system performance. This can be illustrated by the design of the voice activity detector in [23], which is plagued by serious shifts in the central frequency and bandwidth of the filter banks, and offsets and non-linearity in the envelop detector circuitry. These shifts are not only process dependent, but also vary slightly due operational condition

variations. Nevertheless, they do not impact system classification performance, as the adaptive learning capability of the classifier learns thresholds, and feature values with these shifts incorporated. This assumes that adaptive learning runs continuously enough to track out variations.

It is important to note that feature extracting systems do not require redundancy, in contrast to many digitally assisted analog techniques, which typically utilize data (bandwidth) redundancies such as fault tolerant encoding schemes or circuit redundancies such as auxiliary comparators. Instead, feature extracting systems inherently exploit the post processing classifier training phase to digitally assist the analog frontend without any analog redundancies.

These alternative impairment mitigation schemes, unique to the feature sampling ADCs, result in very different performance requirements compared to traditional data converters. Impairments can be tolerated without significant performance degradation, to save additional power consumption and area as discussed in the 5th INSET, making feature extracting ADCs very compelling for many applications.

FIFTH INSET:

As analog imperfections such as non-linearities, gain errors, offset, etc. can be absorbed in the classifier commonly following a feature extracting ADC, increased non-idealities can be tolerated, opening up a whole new world of design strategies to further reduce the power and area implications of the introduced analog analytics. We will here briefly introduce a, non-exhaustive, list of upcoming opportunities:

Subthreshold design: One way to exploit the higher tolerance to circuit imperfection of analog analytic to achieve ultra-low power at or below microwatts is the usage of circuit architectures utilizing the weak inversion (subthreshold) mode of transistors. These topologies need very low power supply voltage which are appropriate for transistors with finer geometry. Furthermore, it's exponential behavior in weak inversion (similar to Bipolar) can be exploited for non-linear analog preprocessing required for mapping the signal to lower bandwidth feature spaces [36]. The matching and modelling errors typical in subthreshold circuits are absorbed in the adaptive classifier.

Die size: Another potential tradeoff in analog preprocessor is the die size. Variations in threshold voltage V_T and β which are the main sources of mismatch in MOSFET devices are inversely proportional to the gate area [37]. Therefore, the intuitive approach of increasing the device aspect ratios to mitigate matching issues in analog design has been a common practice. Also here, the higher tolerance of analog analytics systems to mismatch imperfection helps to reduce device size.

Programmability: As discussed previously, feature extracting ADCs could benefit tremendously from having a programmable analog analytics front-end. This enables the architecture to address multiple applications with programmable features. While this is straightforward in digital circuits, where e.g. digital filters and equalizers are tunable to different center frequency or bandwidth by changing clock frequency and parameter adjustments, this is less evident in analog designs. Typically analog circuit blocks require an involved redesign for e.g. tuning to different frequencies. However, architectures such as switched capacitor filters [38] have addressed programmability of analog blocks effectively. The large threshold frequency (F_t) of advance CMOS process technologies and relatively low signal bandwidth of many event driven applications can accommodate novel ultra-low power and programmable signal processing circuit design techniques such as analog adaptive filtering using digital calibration and adaptive switched capacitor biquad bank of filters as in TI's ultra-low power voice signature detection [41]. While there is still a long way to go towards the fully programmable analog feature extractors, switched-cap or N-path architectures might lead the way, facilitated by the high impairment tolerance of the feature extracting ADC approach. Also here, impairments such as charge injection, clock feedthrough, etc. can be absorbed by the classifier.

These examples illustrate a new world of opportunities opening up due to the increased impairment tolerance in feature extracting ADCs integrated in a classifier application. This makes them optimally prepared to make a difference in deep sub-micron implementations.

Conclusions and outlook

A growing new class of low power applications does not require perfect signal reconstruction or full information retrieval in the digital domain, rather targets specific feature extraction from an observed signal. This is especially apparent in the emerging application domain of natural human-device interfaces, IoT, and ubiquitous sensing. The ultra-low-power, always-on requirements for these applications prompts revisiting traditional system partitioning to achieve a significant reduction in system energy consumption. This paper gave an overview of the emerging field of analog-to-information processing in light of various sub-Nyquist sampling techniques recently appearing in literature. Special attention is given to feature extracting ADCs, which, extracts a specific subset of features from the waveform by combining analog analytics with low rate sampling. It is important to note that feature extracting ADCs are not suggested as a replacement for classic converters. In many applications where reconstruction of signal is required, such as multimedia applications, a standard approach is still required. Instead, it forms a complementary approach for applications which do not involve reconstruction of original signals, or can form a smarter wake-up front-end to higher complexity systems.

Both at system level, as well as at circuit level, these feature extracting ADCs allow new design opportunities towards run-time energy scalability and power savings. Yet, new challenges also arise. Unlike most standard approaches where each block (ADC, DSP,…) is independently evolving, the new approach requires a system optimization to realize all the benefits of performance and power efficiency, hence is currently mostly application-specific. While several application-specific implementations have proven the power consumption benefit of this approach, at the moment, it is not known how to expand this to efficiently sampling generic features, in sharp contrast with programmable digital-centric solutions. This paper hopes to stimulate this discussion, which will require an interesting interaction between information theory and circuit design.

Bibliography

[1]   C. Shannon, "Communication in the Presence of Noise," *Proc. Inst. Radio Eng.,* vol. 37, no. 2, pp. 10-21, 1949.

[2]   D. H. a. D. J. Brady, "Compression at the Physical Interface," *IEEE Signal Processing Mag.,* vol. 25, no. 2, pp. 67-71, 2008.

[3]   S. K. J. L. M. W. M. D. D. B. T. R. Y. M. a. R. Baraniuk, "Analog-to-information conversion via random demodulation," in *IEEE Dallas/CAS Workshop on Design, Applications, Integration and Software*, 2006.

[4] E. J. Candes and M. B. Wakin, "An Introduction To Compressed Sampling," *IEEE Signal Process. Mag.,* vol. 25, no. 2, pp. 21-30, 2008.

[5] J. a. T. T. E.J. Candes, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf Theory,* vol. 52, no. 2, pp. 489-509, 2006.

[6] R. Baraniuk, "Compressive sensing," *IEEE signal processing magazine,* vol. 24, no. 4, 2007.

[7] M. F. M. A. D. D. T. J. N. L. T. S. K. E. K. a. R. G. B. Duarte, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine,* vol. 25, no. 2, 2008.

[8] F. A. P. C. a. V. S. Chen, "A signal-agnostic compressed sensing acquisition system for wireless and implantable sensors," in *IEEE Custom Integrated Circuits Conference (CICC)*, 2010.

[9] Y. a. A. E. G. Oike, "A 256× 256 CMOS image sensor with ΔΣ-based single-shot compressed sensing," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2012.

[10] A. M. E. G. A. D. G. a. D. J. A. Dixon, "Compressed sensing system considerations for ECG and EMG wireless biosensors," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 6, no. 2, pp. 156-166, 2012.

[11] O. F. L. F. C. a. V. S. Abari, "Why analog-to-information converters suffer in high-bandwidth sparse signal applications," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 60, no. 9, pp. 2273-2284, 2013.

[12] M. V. a. P. Mariliano, "Sampling signals with finite rate of innovation,," *IEEE Trans. Signal Process. ,* vol. 59, no. 4, pp. 1417-1428, 2002.

[13] T. P.-L. D. M. V. P. M. a. L. C. Blu, "Sparse sampling of signal innovations," *IEEE Signal Processing Magazine,* vol. 25, no. 2, pp. 31-40, 2008.

[14] M. M. Y. C. E. a. A. J. Elron, "Xampling: Signal acquisition and processing in union of subspaces," *IEEE Transactions onSignal Processing,* vol. 59, no. 10, pp. 4719-4734, 2011.

[15] F. Ren and D. Markovic, "A configurable 12-to-237KS/s 12.8mW sparse-approximation engine for mobile ExG data aggregation," in *IEEE International Solid- State Circuits Conference*, 2015.

[16] J. N. S. K. a. W. S. Sohn, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters,* vol. 6, no. 1, pp. 1-3, 1999.

[17] T. N. Y. H. a. P. O. Plötz, "Feature learning for activity recognition in ubiquitous computing," in *In IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2011.

[18] I. Jolliffe, Principal component analysis, John Wiley & Sons, Ltd, 2002.

[19] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends in Machine Learning 2,* vol. 1, 2009.

[20] T. a. B. S. Huynh, "Analyzing features for activity recognition," in *In Proceedings of the ACM 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*, 2005.

[21] J. Schürmann, Pattern classification: a unified view of statistical and neural approaches, New York: Wiley, 1996.

[22] H. F. L. a. C. D. Peng, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1226-1238, 2005.

[23] K. S. L. W. M. a. M. V. Badami, "Context-aware hierarchical information-sensing in a 6μW 90nm CMOS voice activity detector," in *In IEEE International Solid-State Circuits Conference-(ISSCC)*, 2015.

[24] A. C. T. W. B. M. D. J. W. T. a. V. D. Raychowdhury, "A 2.3 nJ/frame voice activity detector-based audio front-end for context-aware system-on-chip applications in 32-nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 48, no. 8, pp. 1963-1969, 2013.

[25] H. T. T. M. Y. a. H. K. Noguchi, "An ultra-low-power VAD hardware implementation for intelligent ubiquitous sensor networks," in *In IEEE Workshop on Signal Processing Systems*, 2009.

[26] S. K. B. W. M. a. M. V. Lauwereins, "Context-and cost-aware feature selection in ultra-low-power sensor interfaces," in *In European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2014.

[27] E. Vittoz, "Future of analog in the VLSI environment," in *IEEE International Symposium on Circuits and System*, 1990.

[28] R. Sarpeshkar, "Analog versus digital: extrapolating from electronics to neurobiology," *Neural computation,* vol. 10, no. 7, pp. 1601-1638, 1998.

[29] B. Murmann, Limits on ADC power dissipation, Springer Netherlands, 2006.

[30] L. a. J. A. S. Gu, "Radio-triggered wake-up for wireless sensor networks," *Real-Time Systems,* vol. 29, no. 2-3, pp. 157-182, 2005.

[31] R. a. R. L. Jafari, "A low power wake-up circuitry based on dynamic time warping for body sensor networks," in *In IEEE International Conference on Body Sensor Networks (BSN)*, 2011.

[32] R. H. N. S. G. a. M. S. Jafari, "Adaptive electrocardiogram feature extraction on distributed embedded systems.," *IEEE Transactions on Parallel and Distributed Systems,* vol. 17, no. 8, pp. 797-807, 2006.

[33] B. Murmann, "Digitally assisted Analog Circuits," *IEEE Micro,* pp. 38-46, 2006.

[34] A. S.-V. Pierlugi Nozzo, "Robustness in Analog Systems: Design Techniques, Methodologies, and Tools," *Symposium on Industrial Embedded Syst,* pp. 194-203, 2011.

[35] B. Murmann, "Digitally assisted Data Converter Design," *Proceedings of ESSCIRC,* pp. 24-31, 2013.

[36] E. Vittoz and J. Fellrath, "CMOS Analog Integrated Circuits Based on Weak Inversion Operations," *Solid-State Circuits, IEEE Journal of,* vol. 12, no. 3, pp. 224-231, 1977.

[37] M. Pelgrom and A. C. J. Duinmaijer, "Matching pries of MOS transistorsopert," in *ESSCIRC*, 1988.

[38] R. e. a. Perez-Aloe, "Programmable time multiplexed switched capacitor variable equalizer for arbitrary frequency response realizations," *Solid-State Circuits, IEEE Journal of,* vol. 32, no. 2, pp. 274-278, 1997.

[39] M. M. a. Y. Eldar, "Sub-Nyquist Sampling," *IEEE Signal Process. Mag.,* vol. 28, no. 6, pp. 98-124, 2011.

[40] D. H. a. D. J. Brady, "Compression at the Physical Interface," *IEEE Signal Process. Mag.,* vol. 25, no. 2, pp. 67-71, 2008.

[41] F. G. L. R. E. B. M. A. C. S. R. J. H. a. P. E. Pace, "A Nyquist folding analog-to-information receiver," in *In IEEE 42nd Asilomar Conference on Signals, Systems and Computers*, 2008.

[42] Instruments, Texas, "http://www.ti.com/tool/adc12j4000evm#technicaldocuments," [Online].

[43] R. Taft, C. A. Menkus, M. Tursi, O. Hidri and V. Pons, "A 1.8-V 1.0-GSPS 10b Self Calibrating Unified folding interpolating ADC With 9.1 ENOB at Nyquist Frequency," *IEEE Journal of Solid State Circuits,* vol. 44, no. 12, pp. 3294-3304, 2009.

[44] P. e. a. Amberg, "Digitally-assisted analog circuits for a 10 Gbps, 395 fJ/b optical receiver in 40 nm CMOS," in *IEEE Asian Solid-state circuit conference*, Jeju, Korea, 2011.