G.722, A New CCITT Coding Standard for Digital Transmission of Wideband Audio Signals

Paul Mermelstein

Introduction

The rapid increase in digital connectivity of telephone networks, brought about by the gradual removal of most analog links, suggests a new look at enhancing the quality of audio transmitted over the telephone network. Pulse-code modulation (PCM) with 64 kbit/s μ-law or A-law (G.711) arose in response to the need for multiple analog-digital-analog conversions of standard 300–3,400-Hz audio signals. Such signals are considered here to be narrowband audio. Modern speech coding techniques permit reduction of the transmitted bit rate, while preserving audio quality, as in CCITT (International Telegraph and Telephone Consultative Committee) Recommendation G.721, where the customary 300–3,400-Hz-wide telephone signal is encoded at 32 kbit/s [1]. Alternatively, one can provide improved audio quality and maintain the transmission rate at 64 kbit/s. Such improvements are most important for audio or audio-visual conferencing applications where one would like to approach the quality of face-to-face communication. CCITT Study Group XVIII recognized the need for a new international coding standard on high-quality audio to allow interconnection of diverse switching, transmission, and terminal equipment and, thus, organized an Expert Group in 1983 to recommend an appropriate coding technique. Hummel [2] provides a good introduction to the working methods of the CCITT. The coding method described in this paper constitutes the group's recommendation, which was approved by the CCITT through an accelerated procedure in 1986. The algorithm represents the results of a joint effort of contributors from around the world,* and is best described in a series of papers presented at Globecom '86 [3]. This paper is meant to be a tutorial discussion, responsibility for which lies completely with the author. Bit-level particulars of the algorithm, although important for correct implementation of the standard, are not discussed in detail. For more complete information, the reader should refer to the forthcoming CCITT document.

Requirements

The main objective for the new standard is to allow speech transmission at 64 kbit/s with quality as high as poissible and significantly better than that provided by 8-bits/sample, 8-kHz sampled PCM coding. If the signal is sampled at 16 kHz, or twice the PCM rate, the spectrum of the signal to be encoded can be extended to about 7 kHz (3-dB point), and this results in a major improve-

---

*A complete list of the participants in the Expert Group can be found in Appendix 1 to the Report CCITT, COM XVIII-R 17-E, April 1986. Participating organizations included BNR, Canada; CNET, France; FTZ, Federal Republic of Germany; CSELT and SIP, Italy; NTT, Fujitsu and KDD, Japan; PTT, Switzerland; BT, United Kingdom; Bellcore and Comsat, United States. Technical contributions of all participants are recognized and acknowledged without specific credit on individual items.

0163-6804/88/0001-0008$01.00

ment in audio quality. For speech signals, little additional enhancement is achieved by extending the cutoff frequency even higher. Although the need to encode music signals was recognized by the designers, a further gain in quality of transmitted music, which might have been achieved through use of an even higher cutoff frequency, would have required a transmission rate exceeding the 64-kbit/s target and would have increased the cost and complexity of the coding system. Figure 1 compares the attenuation curve chosen for wideband audio to that used in traditional (narrowband) telephony. Note that the end-to-end digital transmission conditions permit the low-frequency response of the coding system to be enhanced further by extending the cutoff frequency to 50 Hz, which improves the naturalness of the audio and results in the wideband audio signal considered here.

Today's voice networks carry a fair amount of voiceband data and analog signaling information. Such traffic can be carried more economically in digital form wherever end-to-end digital connections are available. The codec performance requirements for voiceband data are substantially different from those of voice signals. If the codec were required to encode voiceband data signals as well, its cost and complexity would increase substantially, as evidenced by the G.721 32-kbit/s codec [4]. In very large scale integration (VLSI) realizations, the cost penalty for the additional complexity is less severe than for discrete implementations and manifests itself mainly in increased chip area. If the wideband codec is not required to carry voice-band data, it can be optimized for best performance on speech signals without such penalties.

Speech quality is generally degraded through accumulation of quantization noise introduced at signal conversion points. Since the transmission systems envisioned are digital end-to-end, we require only one analog-digital conversion at the source and one digital-analog conversion at the destination. Robustness to multiple conversion sequenses of analog to digital-encoded representation is not required. However, interconnection of multiple sources of audio at conference bridges is best carried out with a uniformly quantized representation of the digital signal. To allow for multiple bridges in one connection, provision is made for a small number (up to three) of digital encoding/decoding sequences. Furthermore, both narrowband and wideband audio signals may arrive at audio bridges and bridge output should also be available in narrowband or wideband form.

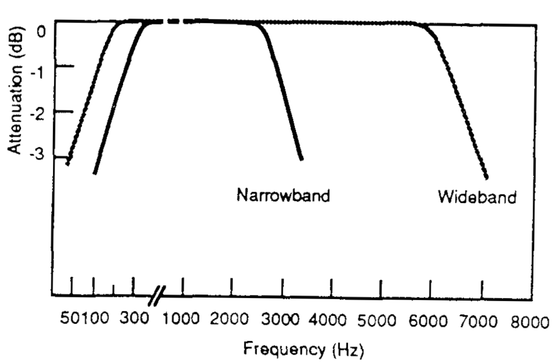

Fig. 1. Frequency Characteristics of Wideband and Narrowband Audio Channels.

The delay introduced by an encoding-decoding process should be limited to 4 ms. This requirement is imposed primarily for echo control purposes. Echo control is simplified by the fact that the common sources of echo on the telephone network, namely the hybrids at two-/four-wire conversion points, are absent in end-to-end digital connections. However, interconnection of existing narrowband links with wideband conferencing systems may introduce sources of echo whose control may be difficult unless the end-to-end delay of the wideband signal is carefully limited.

For some applications, it may be desirable to provide an "in-band" data channel, i.e., to replace part of the 64-kbit/s transmission channel used for speech by data. To permit the least degradation in audio quality consistent with a time-varying demand for data transmission, three speech transmission modes were defined: 64, 56, and 48 kbit/s. Where full 64-kbit/s transmission channels are available, this makes 0, 8, or 16 kbit/s available for data transmission. On those North American channels that are limited to 56 kbit/s, only 0 or 8 kbit/s of data can be transmitted.

An important practical requirement is that the coder provide acceptable performance (maintain intelligibility) for transmission bit error rates up to 10^{3}. This requirement ensures that performance degrades gently even under the worst transmission conditions that one may encounter on the telecommunication network.

Audio Coding Considerations

A tutorial review of modern speech coding techniques can be found in Crochiere and Flanagan [5]. Highly intelligible speech may be transmitted today at rates as low as 800 bit/s. However, most low-bit-rate techniques are restricted to the transmission of speech and are inappropriate for other forms of audio. If one wishes to encode any audio signal, only waveform coding techniques enter into consideration. Forward adaptive techniques require the transmission of some adaptation parameters in addition to the signal obtained by inverse filtering with a prediction filter. Transmission of such parameters, generally considered as side information, requires the use of special framing bits to allow identification of the parameter-carrying bits in the encoded signal. Backward adaptive techniques eliminate the need to transmit such framing bits. Once synchronization between receiver and transmitter has been achieved, groups of eight consecutive bits can be identified, each group carrying similar information about the signal. Recent work on the G.721 standard indicated that 4-bit sample adaptive differential PCM (ADPCM) coding results in toll-quality narrowband speech [6]. Making the quantizer more precise by increasing the number of bits allotted to each sample generally yields no audible improvement beyond 6 bits/sample. Thus, in the absence of coding for bit-rate reduction, 6-bit quantization at a sampling rate of 16 kHz or 96 kbit/s represents an upper limit to the transmission requirements.

Subband coding techniques separate the signal into components occupying contiguous frequency bands and encode the components separately [7]. By appropriately

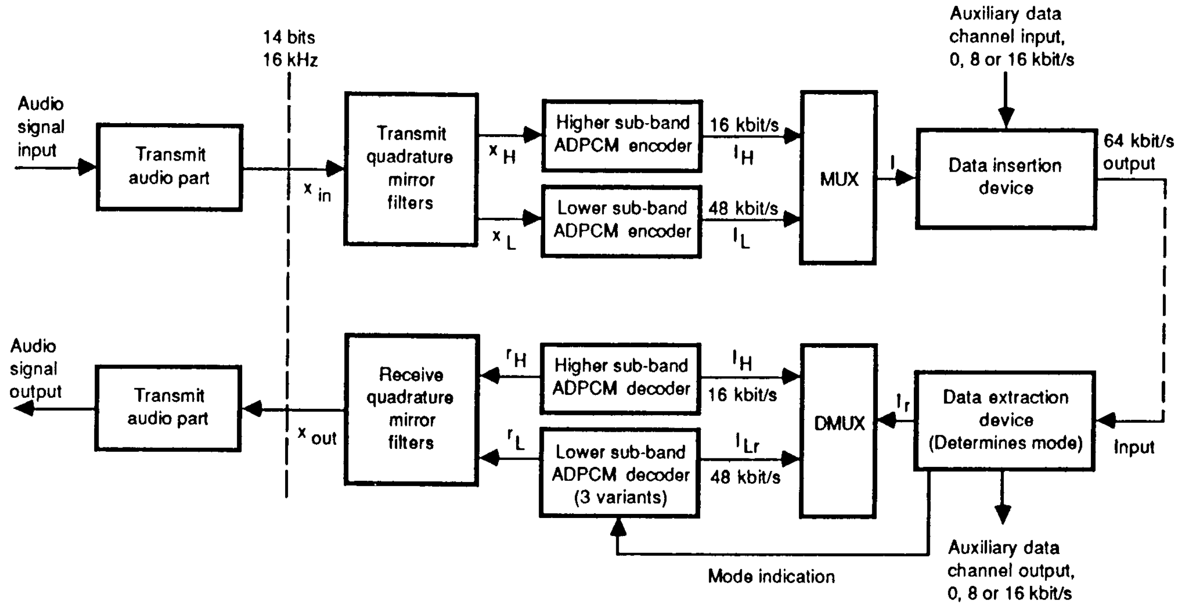

Fig. 2. Block Diagram of the 64-kbit/s (7-kHz) Audio Codec.

allocating the bits across the different bands, the error variance in the reconstructed signal can be shaped with frequency. With the audio signal subdivided into two 4-kHz-wide bands, a high signal-to-noise ratio in the lower band becomes perceptually more important than in the higher band. An advantage of a design that uses two equally wide subbands is that each component is subsampled to 8 kHz and the total transmission rate may be reduced in 8-kbit/s steps by reducing the number of bits assigned to samples of one or the other band. While the bit rate may also be reduced by reducing the sampling rate, those processes generally are more complex to implement. These considerations led to design and evaluation of two alternative subband ADPCM systems, one using 5 and 3 bits/sample for the low- and high-band components, respectively, the other, 6 and 2 bits/sample.

The G.721 ADPCM design employs an adaptive predictor with two poles and six zeros. A fixed predictor design was also tested for the wideband codec, but it led to a generally lower speech quality. A time-varying adaptive allocation of bits to the two subbands according to the short-time signal characteristics was also tried. For voiced sounds carrying significant low-frequency energy, one can assign additional bits to the low band; for fricative sounds, the additional bit may be assigned more advantageously to the high band. However, for the two-band, 4-kHz-wide subband design, the advantage of an adaptive bit assignment is only apparent at the 4 low-2 high bits/sample assignment and is found too small to warrant the additional complexity.

Overall block diagrams for the wideband encoder and decoder are shown in Fig. 2. These blocks are discussed in greater detail in the following sections.

Subband Filtering

The nominal 3-dB band of the codec was chosen as 50–7,000 Hz. Two sets of identical quadrature mirror filters (QMF) are used to divide the wideband signal sampled at a 16-kHz rate into two 8-kHz sampled components to be transmitted, a low band and a high band, and reconstruct the wideband signal from its received low- and high-band components. QMF filters are finite-impulse response, impose a fixed delay without phase distortion, and ensure that aliasing products resulting from subsampling the input signal at the transmitter are canceled at the receiver. However, quantization noise components introduced in coding the low- and high-band signals may not be eliminated completely by the receiver QMF filter. Because the level of the high-band component of the signal may be as much as 40 dB lower than the low-band component, aliasing noise introduced into the high-band frequencies due to coding the low-band signal might be inadequately masked by the high-band signal component. To achieve

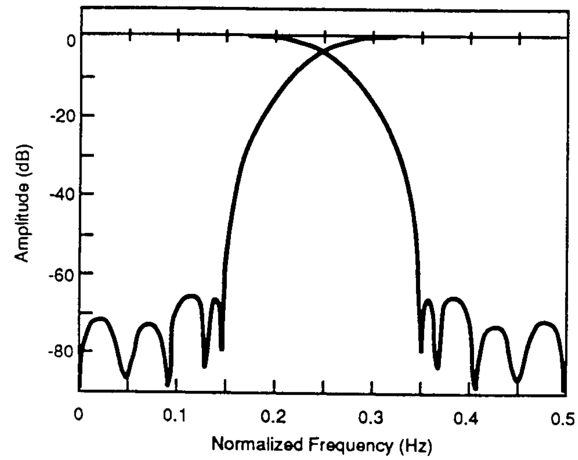

Fig. 3. Frequency Response of the QMF Filters.

a stop-band rejection of 60 dB, we employ a 24-tap filter design, introducing a total signal delay of only 3 ms (see Fig. 3). The resulting signal distortion is below 1 dB over the 100–6,400-Hz band.

The numerical precision with which the partial sums in the QMF filters are accumulated has an important bearing on the accuracy of the low- and high-band signal components generated. The overall goal for the wideband signal representation (after analog-to-digital conversion at the input and before digital-to-analog conversion at the output) is a precision of 14 bits. To this end, the internal coding computations are performed with 16 bits. For the subband signals to have 16 significant bits, the partial sum computations were found to require a precision of 24 bits. Since the wideband signal is accurate to 14 bits, the sum and difference signals to which the QMF filter coefficients are applied are only precise to 13 bits. To prevent the introduction of noise due to differently specified analysis and synthesis filters, the QMF filter coefficients are also represented with 13 bits.

ADPCM Coders

Two ADPCM coders are required, one for the low-band signal and one for the high-band signal. The coders employ identical adaptation strategies to modify the quantizers and predictors based on the previously observed characteristics of the input signal. The low- and high-band coders are very similar, except for small differences due to the need to vary the number of bits output by the low-band coder and the fact that the high-band quantizer output is always 2 bits/sample.

The adaptive predictor design is borrowed directly from that investigated in detail in developing the G.721 standard. The two-pole, six-zero design combines good prediction gain for speech with relatively simple stability control. Robust adaptation is assured by leaky integrators allowing the effects of transmission errors to dissipate rapidly [8]. Transmission errors may introduce differences between the predictor memories at the transmitter and receiver. Adapting the predictors and quantizers using the residual signal alone and not the reconstructed signal ensures that the predictors at the transmitter and receiver recover tracking rapidly for all signals [9]. The adaptive quantizer design is also borrowed directly from the G.721 standard since the low-band signal component resembles the narrowband speech signal in most of its properties. G.721 employs a dual-mode quantizer, a locked or slowly adapting mode for voiceband data signals, and an unlocked or rapidly adapting mode for speech signals. Since the G.722 standard was not required to encode voice-band data

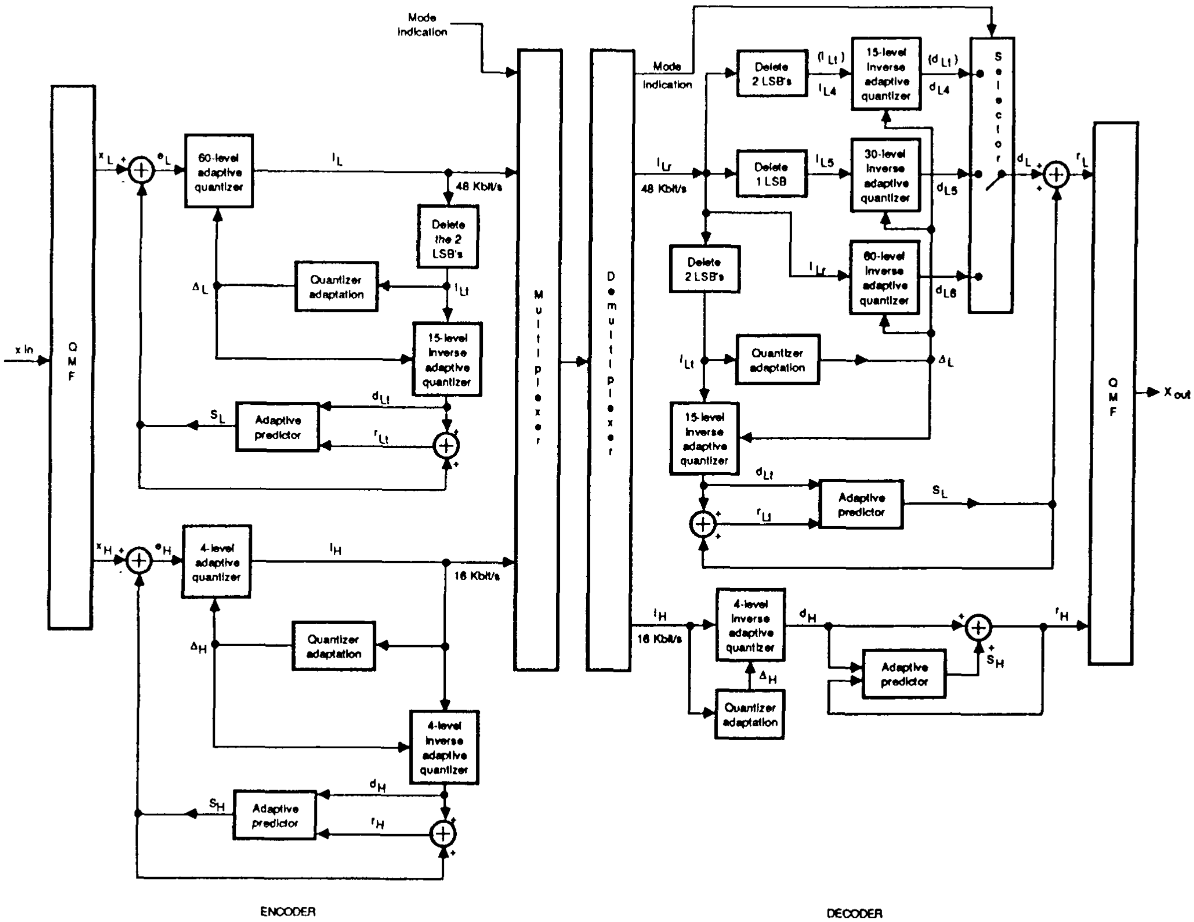

Fig. 4. Detailed Block Diagram of the Subband—ADPCM Encoder and Decoder.

signals, only a single rapidly adapting mode had to be implemented. The dynamic range of the low-band signal quantizer was set to be the same as for G.721, namely 54 dB. A higher dynamic range is allowed for the high-band quantizer, 66 dB, mostly to accommodate music signals. Robust adaptation is employed also for the quantizer scale factor to combat the effects of transmission errors.

Embedded quantizers allow for possible stripping of the less significant bits from the quantized signal during transmission by not making use of those bits in the quantizer adaptation process [10]. The low-band quantizer anticipates that the one or two least significant bits may be stripped from the transmitted code word and adapts the quantizer and predictor using only the four most significant bits. This results in 4- and 5-bit quantizers that are slightly suboptimal in quantization noise-to-signal ratio compared to the quantizers that may be designed without this constraint. The embedding property requires that the 4- and 5-bit quantization boundaries coincide with a subset of those employed for the 6-bit quantizer. Many transmission systems require a minimal number of zero-one alternatives to maintain synchronization. To prevent the all-zero code from appearing even in the 4-bit data representation, only 15 quantizer levels are used in that mode; this also restricts the higher modes to 30 and 60 levels. Experimental evaluations have shown only a fraction of a decibel is lost in quantizing speech signals with the embedded quantizer compared to an unconstrained quantizer design. In terms of subjective performance, the embedded design was not found to be significantly different from the nonembedded design, even after four transcodings.

A significant systems advantage resulting from the embedded coder design is that the encoder may operate without regard to the momentary data transmission requirements. The speech coding and data multiplexing operations are separated logically and possibly even physically. Thus, data may be introduced at a point downstream in the transmission path removed from the encoding terminal. The receiver or decoding terminal

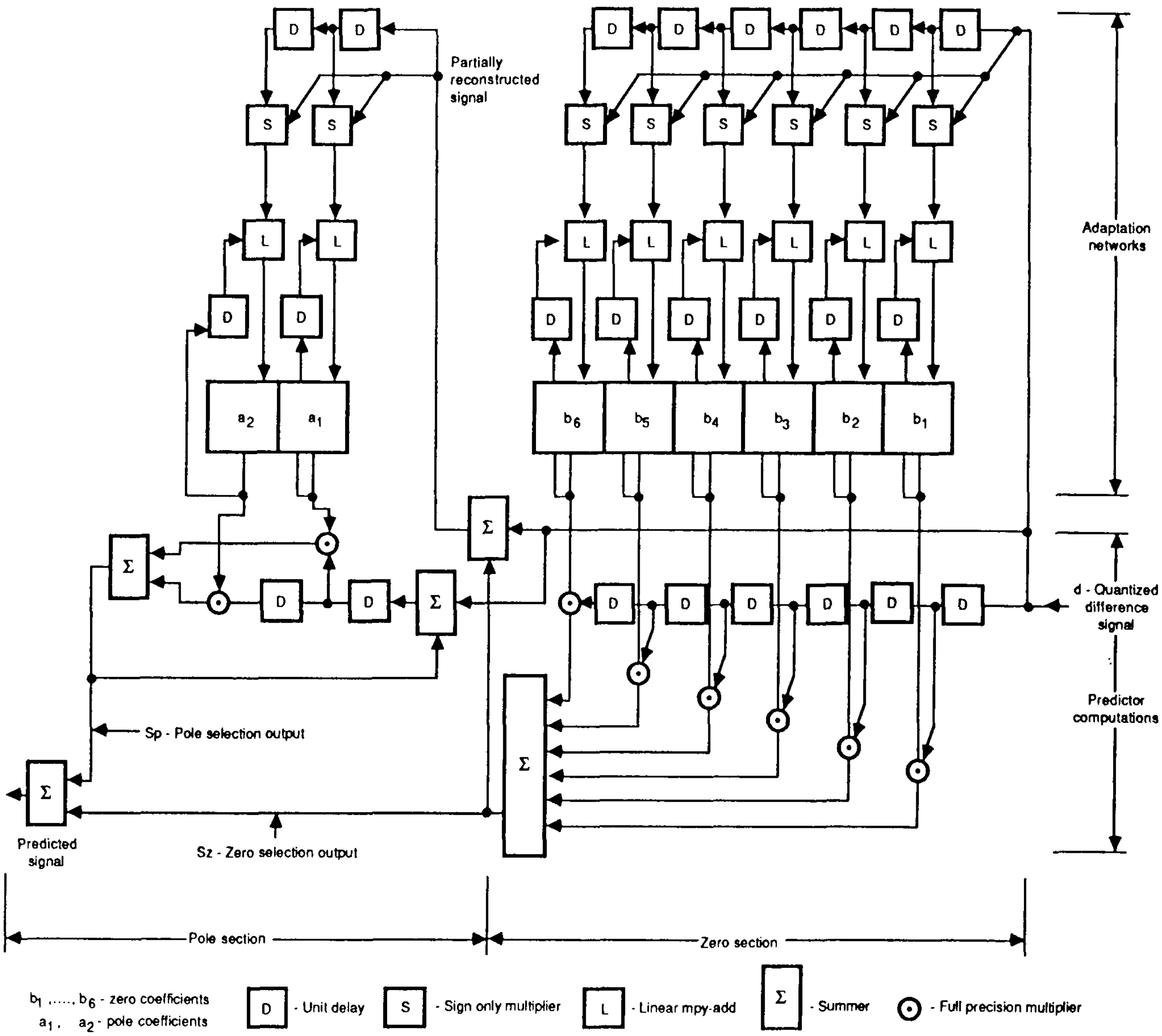

Fig. 5. Signal Flow Diagram for the Adaptive Predictors of each Subband.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE