Wideband Speech Telephony: from 1984 to 2007

a very personal experience

Karl.Hellwig@Ericsson.com

ETSI Workshop on Speech and Noise in Wideband Communication
22nd and 23rd May 2007 - Sophia Antipolis, France

ERICSSON
TAKING YOU FORWARD

Telephony: since 1876 … 1878 … 2004 … 2007

LM Ericsson's first telephone with signal trumpet, 1878.

1878

Telephon-band

300 Hz ... 3400 Hz

True for the mass market

Only some niche applications provide better quality

2004

ERICSSON

TAKING YOU FORWARD

The Situation in Year 1984

- CCITT G.711, log.PCM with 64 kbps, was ready since 1972.
  It was penetrating the long distance networks, but did not reach end-customers yet.

- CCITT G.726, ADPCM with 32 kbps, was just ready.

- Mobile Systems existed in analoge technology (e.g. NMT, AMPS and Net B, C)

- European-wide discussions were ongoing to standardize
  a Pan-European mobile system: the "*Groupe Spéciale Mobile* " had been created.

- The first single chip DSPs were on the market

- ISDN Standards were not ready yet and far from market introduction,
  but ISDN fans were seeking for applications to show the benefits of ISDN:
  - **Wideband speech** and
  - Video Telephony

- Karl Hellwig was a young engineer in digital signal processing, in a small team,
  headed by **Dr. Peter Vary** at **P**hilips **K**ommunikations **I**ndustry (PKI) in Nürnberg.

How "All" began with WB Speech (for me)

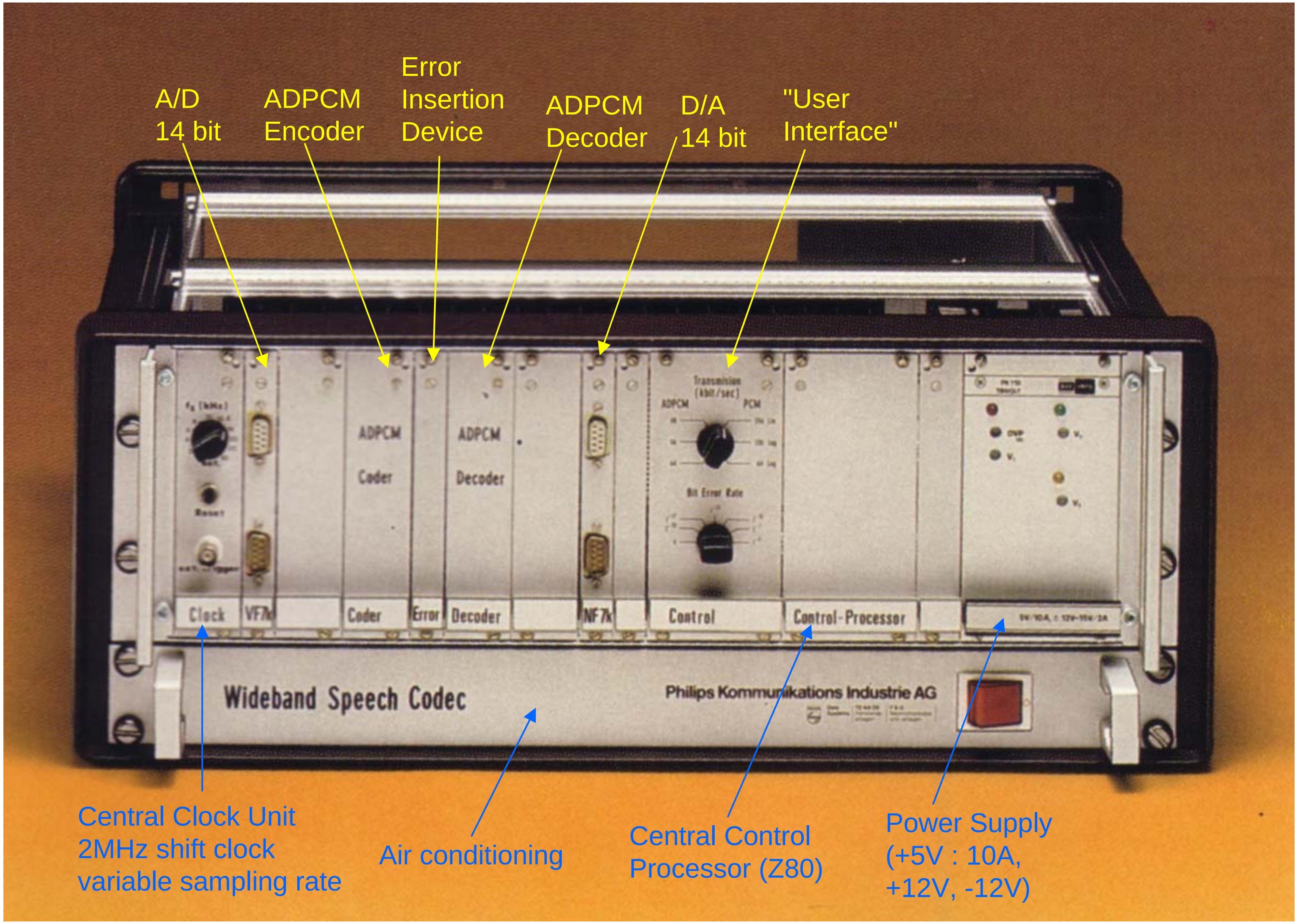
- In spring 1984 some experts from the
  Fernmeldetechnischen Zentralamt der Deutschen Bundespost in Bonn
  contacted Peter Vary at PKI for a small consultancy task:
  Implementation of a new WB Coding Algorithm, invented by **Dr. Manfred Dietrich**
  and a 7-page description was handed over. Objective: CCITT Standard.

- At that times it was not allowed for vendors to participate in CCITT standardisation, only the national **P**ost and **T**elecom and **T**elegraph monopolists were present.
  So we had no further background information and no insight into the CCITT discussions. At the end this turned out to be a substantial drawback.

- We had de facto about 3 months to realize two hardware models

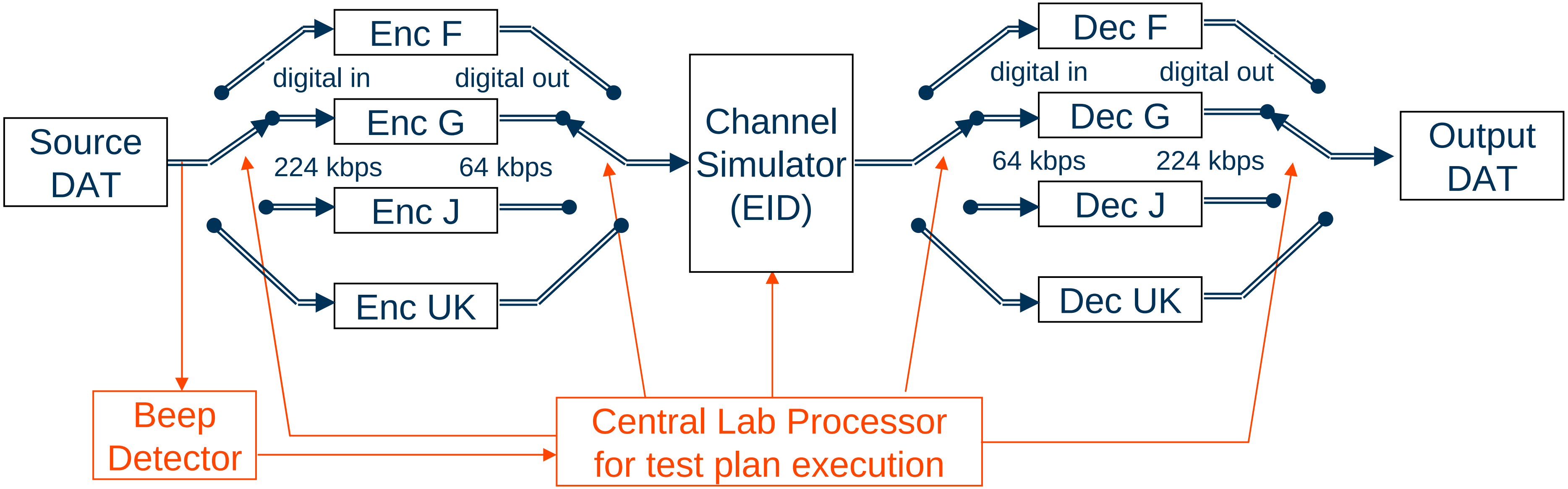
- A self-made modular "PKI DSP System" was ready at hand, which allowed to combine various analogue and digital modules in a flexible way to complex systems. It was based on the industry standard VME-bus, with some usefull extensions.

The German Candidate: modular 19" rack



The CCITT WB Competition Test 1984

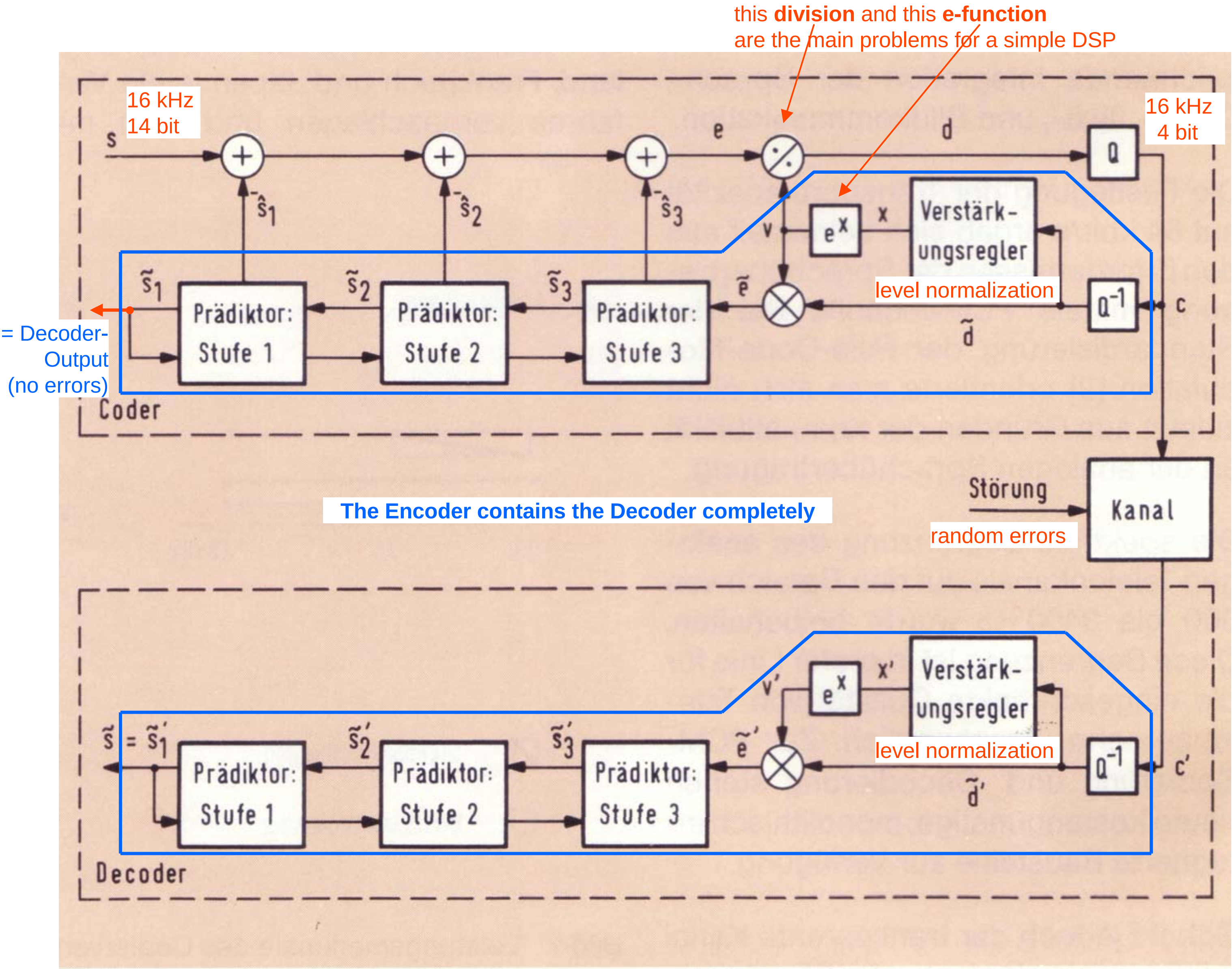
- The real-time laboratory session was scheduled for summer 1984 in Ipswich, UK.
- Four Nations had announced candidates for the WB Speech Contest: **France**, **Germany**, **Japan** and **United Kingdom**.
- All speech processings had to be realized in real-time hardware, because the speech material for source and output were recorded on Digital Audio Tapes (and no one had confidence in simulations ☺).



- The Host Lab had to provide the Central Lab Processor for automated test execution and the other equipment, notably the big switching matrixes.

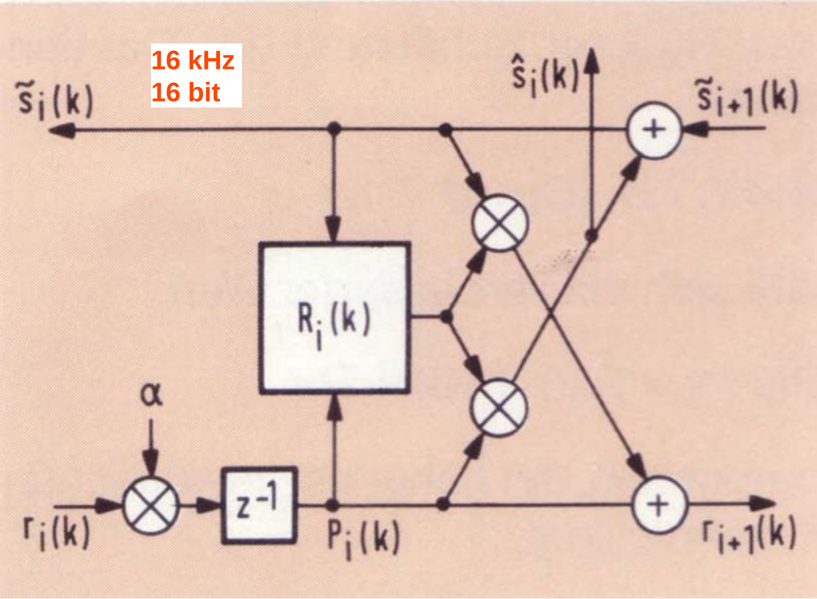
Testing Candiates for G.722 in Ipswich, UK, 1984

Functional Block Diagram of the German Candidate



Source:
PKI: TEKADE-
Fachbericht, 1985

One Predictor Stage



this **division** is a problem
for a simple DSP

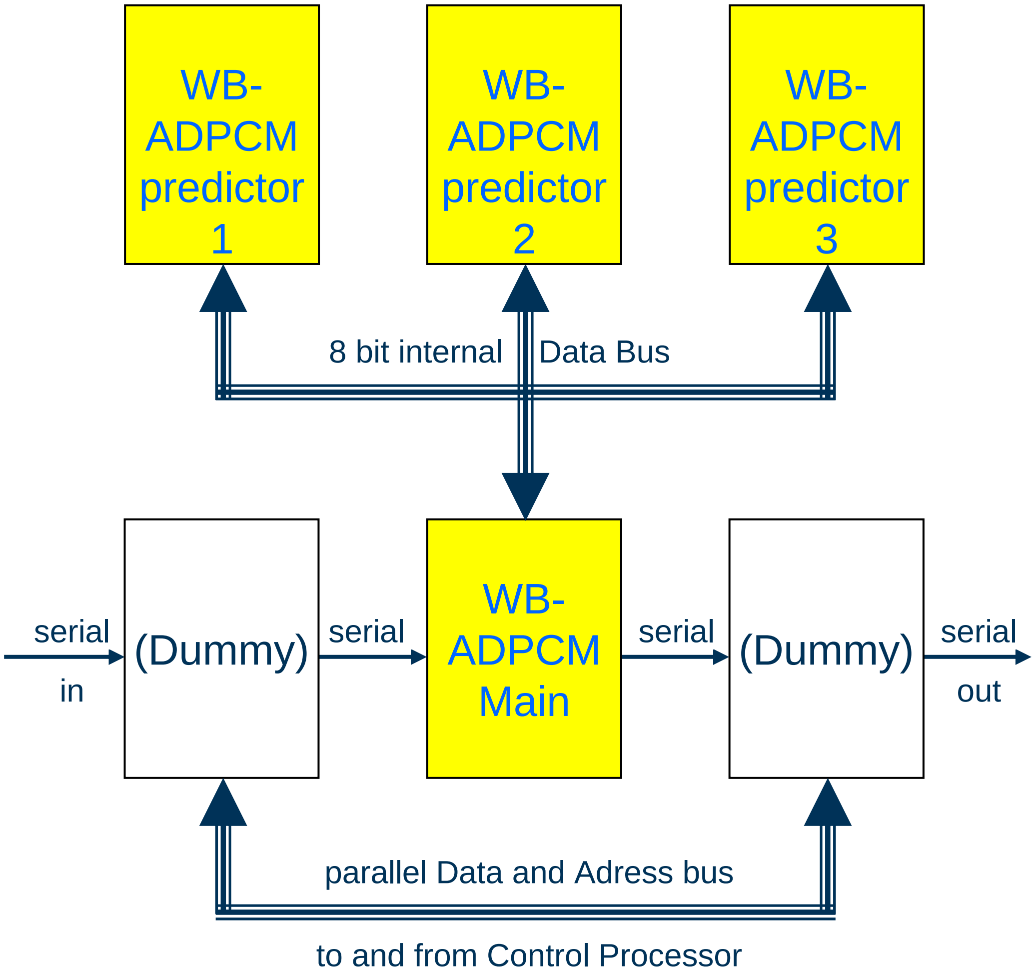
$$R_i(k) = a * C_i(k) \; / \; D_i(k)$$

$$C_i(k) = < S_i \; (k) \;\; * \, r_i \; (k) >$$
$$Di(k) = < S_i^2(k) + r_i^2(k) >$$

16 kHz
32 bit

Source:
PKI: TEKADE-
Fachbericht, 1985

One modular DSP Board with up to 6 DSPs



uPD 7720

Program-Eprom:	512 * 23
Data-Eprom:	510 * 16
Data-RAM:	128 * 16
Multiplier:	16*16
Cycle Time:	250 ns
Power:	1000 mW
Technology:	NMOS
Housing:	28 PIN DIP

This device was the first single-chip DSP, 1981, NEC

No Compiler, simple Assembler ever function carefully hand-optimized

Source:
PKI: TEKADE-Fachbericht, 1985

Result of the Subjective Listening Test

- In 6 of 9 conditions the German (Dietrich ADPCM) Codec was performing very well

- It failed in the other three conditions due to

  - a still **unknown implementation error** (mea culpa):
    for some specific input signals (Music and the **Swiss-German-dialect** !)
    the Encoder-Module produced a "perfect" Decoder output signal, but
    the Decoder-Module "faded away" even in transmission-error-free conditions:
    that is theoretically impossible !

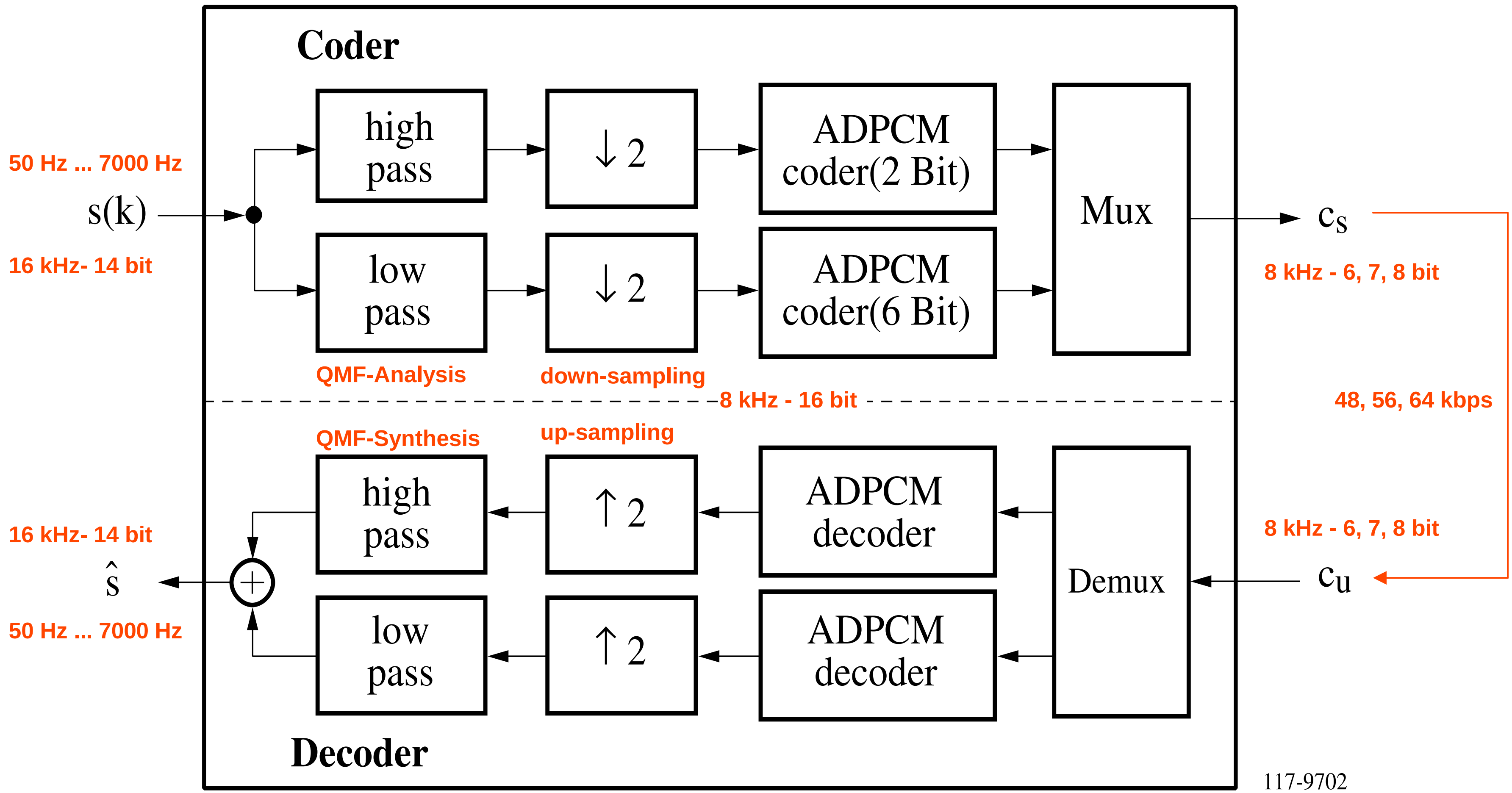
  - it was **not at all optimized for 56 kbit/s and 48 kbit/s**, because these Codec Modes
    were introduced very late in the CCITT discussion and we at PKI got notice very late

- At the end a combination of the French and the Japanese Codec candidates was standardized: G.722 (1988)

- This is still one of the best, low complex Wideband Speech Codecs, although at bit rates of 64, 56 and 48 kbps.

- So far it was not applied in any mass market application - what a pitty

ITU-T G.722: Split-Band WB-ADPCM



- Modes of operation:

wideband speech	64 kbit/s
wideband speech + 8 kbps data	56 + 8 kbit/s
wideband speech + 16 kbps data	48 + 16 kbit/s

Audio Demonstration G.722: WB-ADPCM

- G.711: log.PCM: 300 Hz - 3400 Hz, IRS 64 kbit/s
- G.722: WB-ADPCM: 50 Hz - 7000 Hz, 64 kbit/s
- G.722: WB-ADPCM: 50 Hz - 7000 Hz, 48 kbit/s
- Original speech: 50 Hz - 8000 Hz, 224 kbit/s

The Period "in between" (1)

- While our small team was working 1984 on one side of the room on WB speech, on the other side of the room the preparations began for the big GSM contest.

- At that time we had our own narrowband algorithms at bit rates of 8 to 16 kbps and got also the first version of the "Sluyter" Codec from the **Nat**uurkundig **Lab**oratorium of Philips in Eindhoven, which years later became GSM_FR.

- The difference in speech quality was enormous ! Even higher than today:
  WB (G.722, 64k) NB (GSM_FR, 13k) MS-to-MS call at 7dB C/I

- So the fascination of WB speech Coding for daily usage in telephony began in these days and never left me since then !

- But we had to push WB aside (we never looked for that implementation error) and we had to concentrate on GSM for the next 8 years (and more) to come

- It was in 1992 when we started again to dream the WB dream - a little bit, in internal pre-studies, but mainly in university cooperations (RWTH Aachen)

The Period "in between" (2)

- 1985: the NatLab/PKI Candidate for GSM wins the German competition
- 1986: the NatLab/PKI Codec is best in the European Contest
- 1988: the "Franco-German-Compromise" is ready (GSM_FR, 13 kbps), with main parts from the NatLab/PKI Codec and the Long Term Prediction from IBM, France.
- 1989: first workshop on Speech Coding for the Half-Rate Competition; concept to make this Codec adaptive (based on Prof. Hagenauer's suggestion), but it takes finally until 1995 to get a stable single-rate GSM_HR Standard (5.6 kbps).
- 1994: The performance of the GSM_HR in mobile-to-mobile calls is bad due to double, tandem coding (HR - PCM - HR), so the idea of a **Tandem Free Operation (TFO)** is born.
- 1995: US operators decide to adopt the GSM standard, but only if a better Speech Codec is introduced: fastest Codec selection ever! The bit rate must be below 13 kbps to reuse the channel coding of the GSM_FR.
- 1996: The "American Algorithm" was finally accepted as GSM_EFR Standard (12.2 kbps)
- 1998: After a period of 2 years of intensive discussions in GSM, the concept for an **Adaptive Multi-Rate Codec**, for **FR and HR**, with **NB and WB** components, is agreed. A real-life demonstration of WB Speech Coding by Prof. Vary's Institute in York, UK, gives the final push to accept the feasibility of WB speech in a GSM Full Rate Channel. But then the project is split into two phases: NB first, WB second.
- 1999: The AMR-NB is standard: 8 modes, from 4.75 to 12.20 kbps, for GSM and UMTS.
- 2001: The **AMR-WB** is standard: 9 modes, from 6.60 to 23.85 kbps, for GSM and UMTS. Short time later this Codec was also accepted in ITU-T as G.722.2 and in a slightly different version also in 3GPP2 for CDMA2000. **=> AMR-WB is a true global standard.**

The Period "in between" (3)

- 1992: the first GSM networks start commercial operations.
  GSM now means: "God, send Mobiles!"
- First research discussions start to develop a new radio interface:
  => Universal Telecommunication Radio Access Network (UTRAN)
  for a new Mobile System: => Univeral Mobile Telecommunications System (UMTS)
- Start of a long term research cooperation between PKI and the University of Aachen (RWTH, Prof. Vary)
  to develop WB Speech Coding algorithms for a future GSM / UMTS standard
- 1994: Ericsson opens a new site in Nürnberg and Karl Hellwig joins Ericsson.
  First task: support a big team to create a functional testbed for a pre-UTRAN system.
  Other main activity: push TFO and new architectures to allow WB speech communication
- 1996: Standardisation starts on UTRAN/UMTS;
  The new network architecture takes the Transcoder out of the Radio Access Network
  and into the Core Network: important step to design "Transcoder Free Operation" (TrFO),
  where Mobile-to-Mobile calls can operate with one Encoding-Decoding for optimal quality
- 1998: 3GPP (3rd Generation Partnership Project) is created to allow a worldwide participation.
- 1999: The AMR-NB is standard: 8 modes, from 4.75 to 12.20 kbps
  The signalling for the Maximum Rate Control is inband (very little overhead)
  and allows a Distributed Rate Decision along the whole speech path.
- 2001: The AMR-WB is standard: 9 modes, from <u>6.60, 8.85, 12.65</u> to 15.85 to 23.85 kbps
- TFO and TrFO standards are ready for end-to-end transcoding free operation in all mobile-mobile calls.
- TFO and TrFO are enablers for AMR-WB and optimal end-to-end Maximum Rate Control.

3GPP Codecs: suitable for Mobile Communication

All **3GPP Codecs** (FR, HR, EFR, AMR, AMR-WB)
have been designed with similar principles in mind:

- **20ms Framing:** good compromise between Bit Rate, Frame Rate and Delay.
  In contrast to that PCM (G.711), ADPCM (G.726) and WB-ADPCM (G.722), also LdCelp (G.728) and Acelp (G.729) have been designed for minimal delay, which is fine for serial circuit switched channels - such as ISDN, but which is useless for packet transmission.
  The RTP profiles for AMR and AMR-WB allow even higher packetization time (for some markets important)

- **Discontinuous Transmission** (DTX) is designed-in: saves 40% transmission costs.
  Very important for a costly radio interface and for cost efficient long distant trunking.
  DTX is not trivial for backward adaptive Codecs, such as ADPCM and WB-ADPCM

- **Error Robustness** is designed in, especially Frame Error Robustness.
  This is reflected in the Codec design and in the design of the Error Concealment.
  Codecs like ADPCM and WB-ADPCM are designed for random bit errors, not for burst errors.

- **Adaptivity** in AMR and AMR-WB allows efficient overload protection
  This idea has been "stolen" from DCMEs (Digital Circuit Multiplication Equipment), but the Maximum Rate Control is extented with a generalized Distributed Rate Decision.
  GERAN and the RTP Profile allow Rate Control by inband signalling, without any noticeable overhead, especially important, when a sudden congestion occurs.

- => Especially AMR and AMR-WB **are very well suited for IP Telephony**

Wideband Speech for Telephony

Wide *analogue Speech*-Band (100…7000Hz)

- substantially higher **speech** quality
- not highest priority for audio signals
- easier understanding, less stress during the call
- better conferencing (smaller near - far effect)
- preciser talker recognition and verification
- simpler speech / noise separation

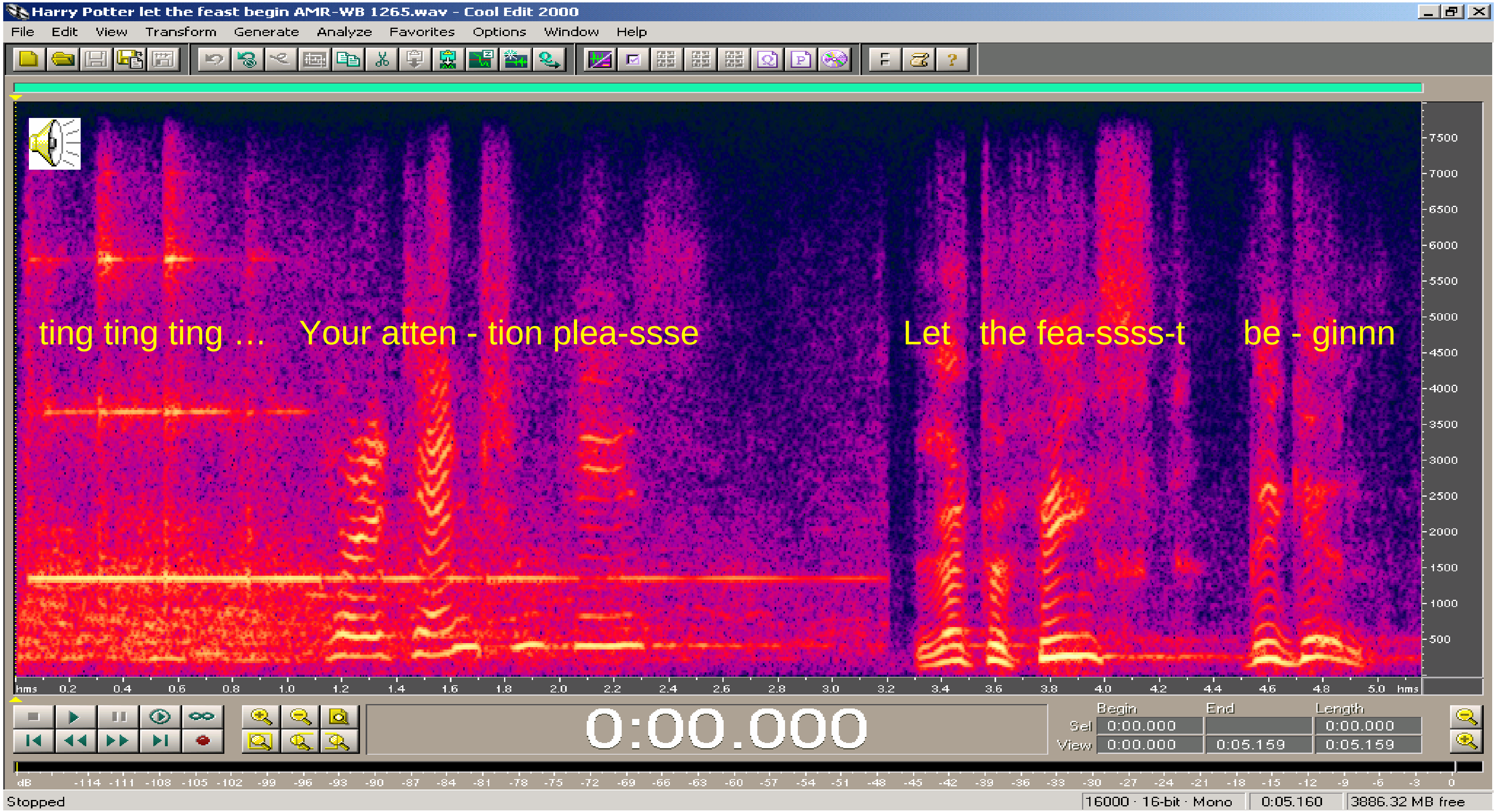
=> more comfortable, more attractive => more mobile minutes

- very easy to use: just call

- similar bit rates, similar transmission costs

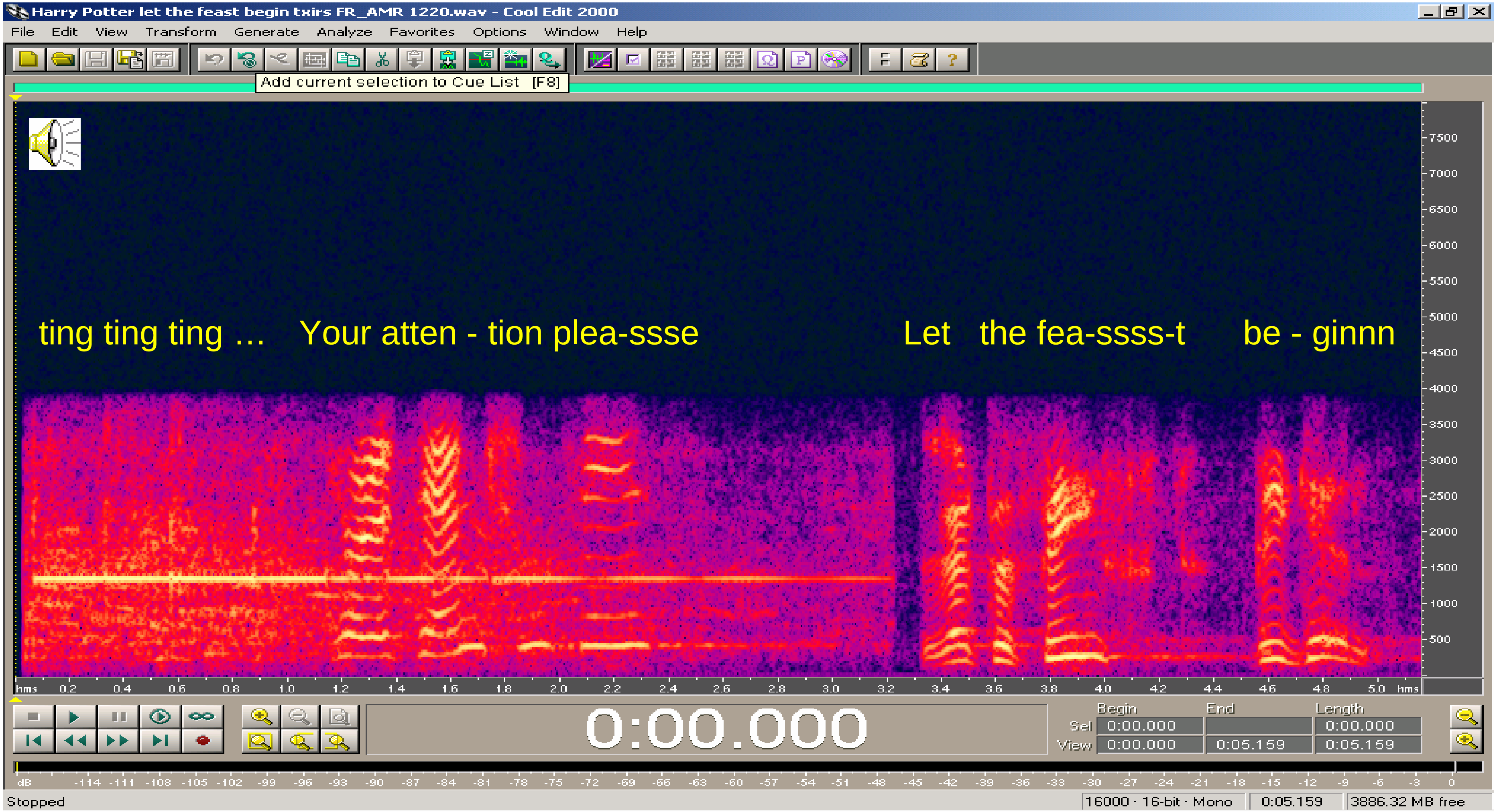
- => more revenue for operators

Spectrum of WB Speech (AMR-WB@12.65 kBit/s)



Spectrum of NB Speech (AMR@12.2 kBit/s)

original

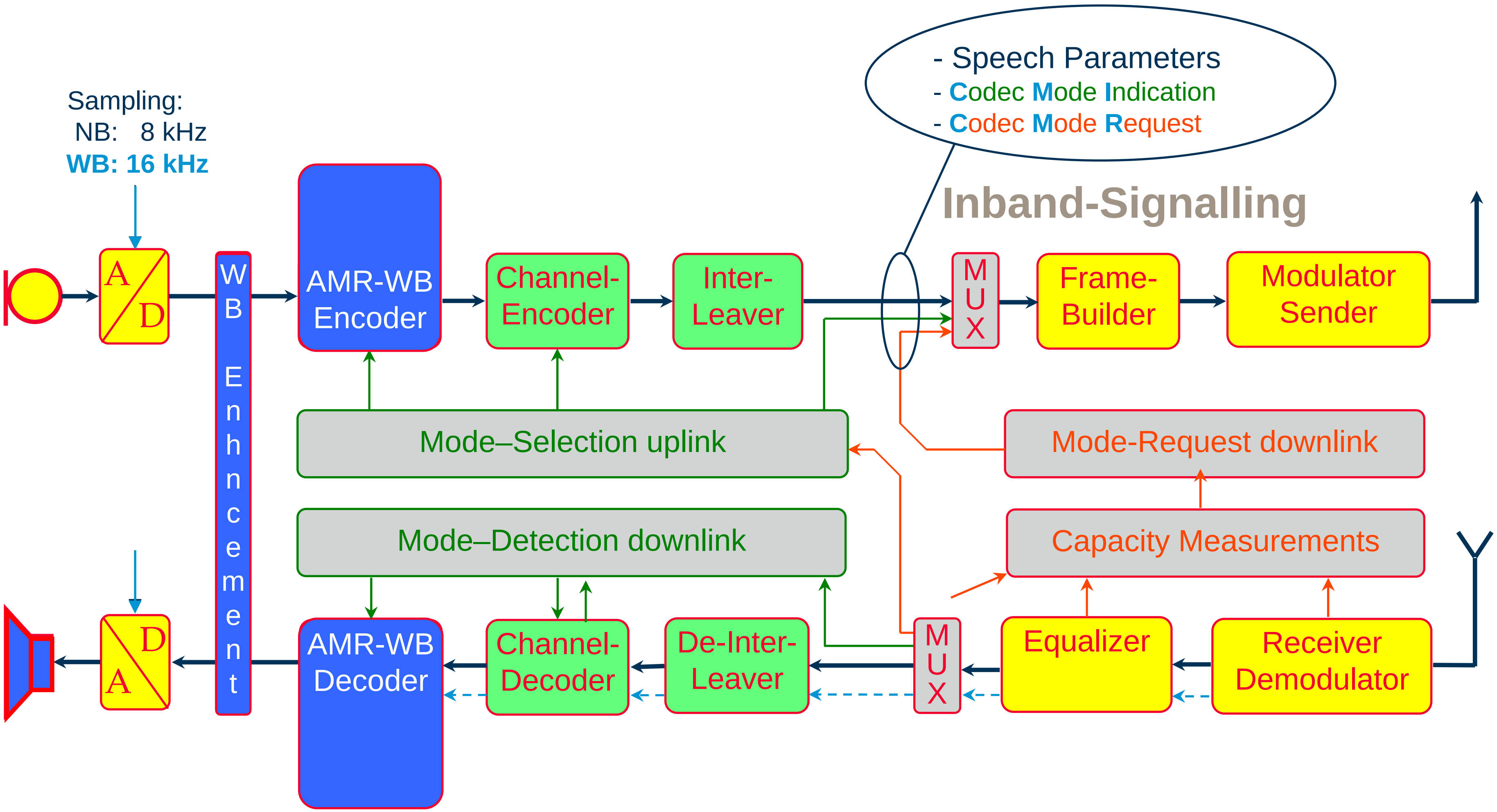


Speech Quality without transmission errors

	AMR-NB	
Bit rate	**Male**	**Female**
12.20		
10.20		
7.95		
7.40		
6.70		
5.90		
5.15		
4.75		

	AMR-WB	
Bit rate	**Male**	**Female**
23.85		
23.05		
19.85		
18.25		
15.85		
14.25		
12.65		
8.85		
6.60		

Adaptive-Multi-Rate Codec: **WB**
Principle for GSM terminal

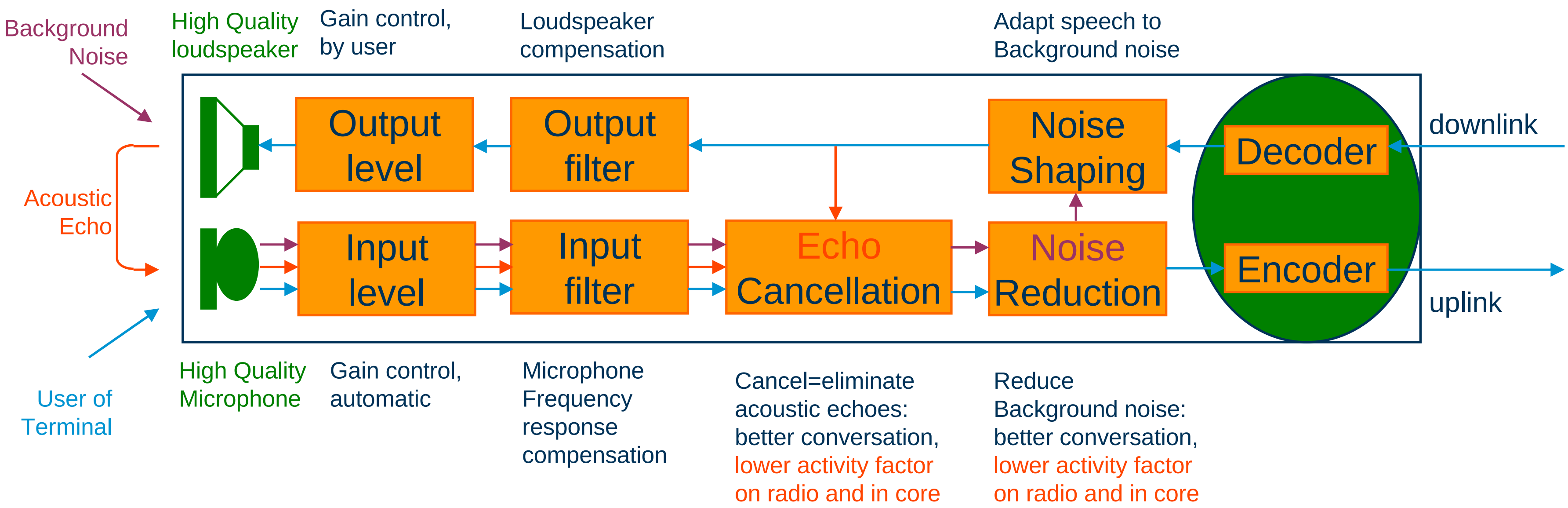


Codec Mode Adaptation: fast: CMR – CMI inband signalling

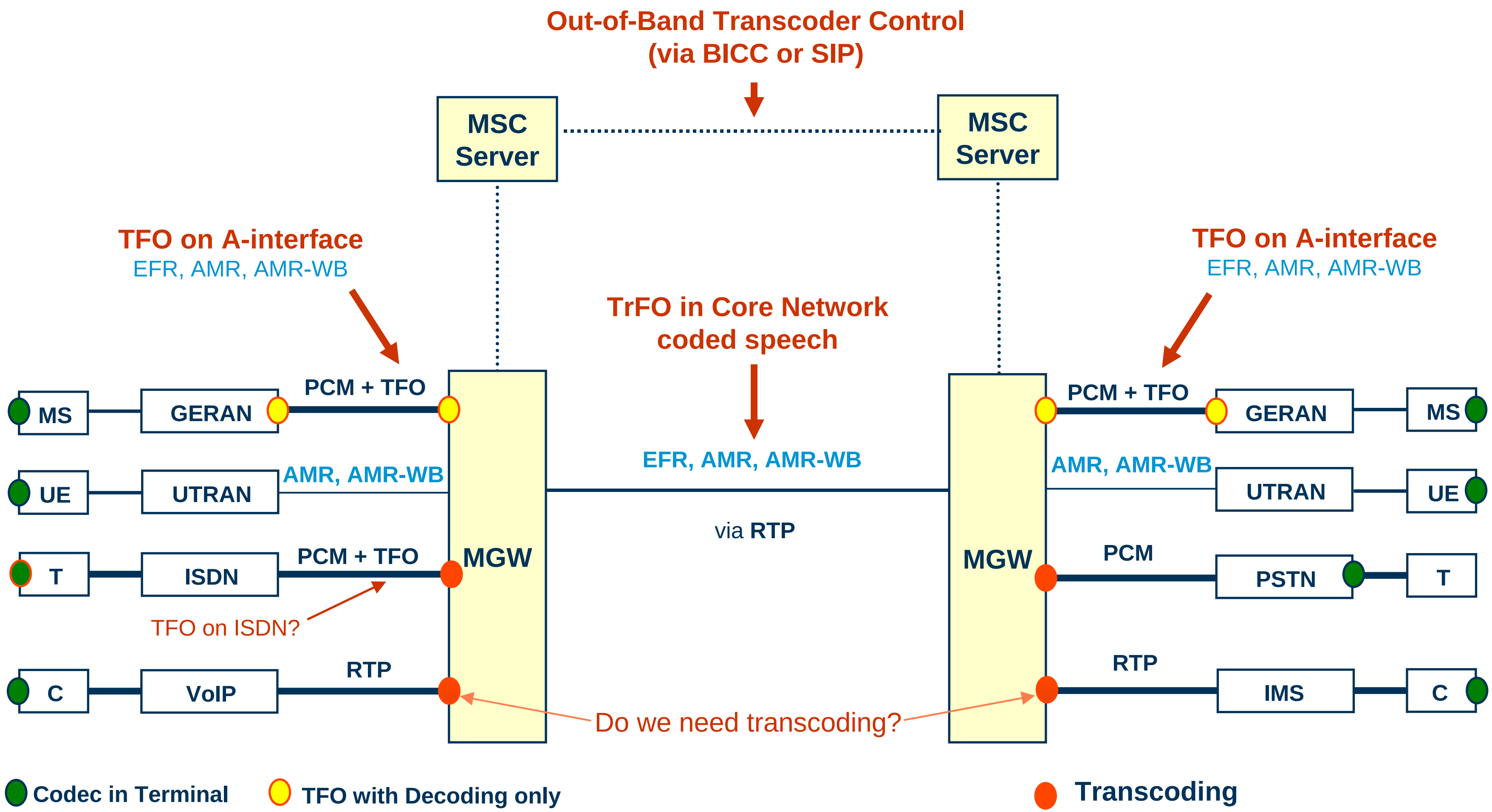
Codec Type and Configuration Modification: slow: out-of-band signalling

Voice Quality Enhancements: Optimal Terminal Block Diagram

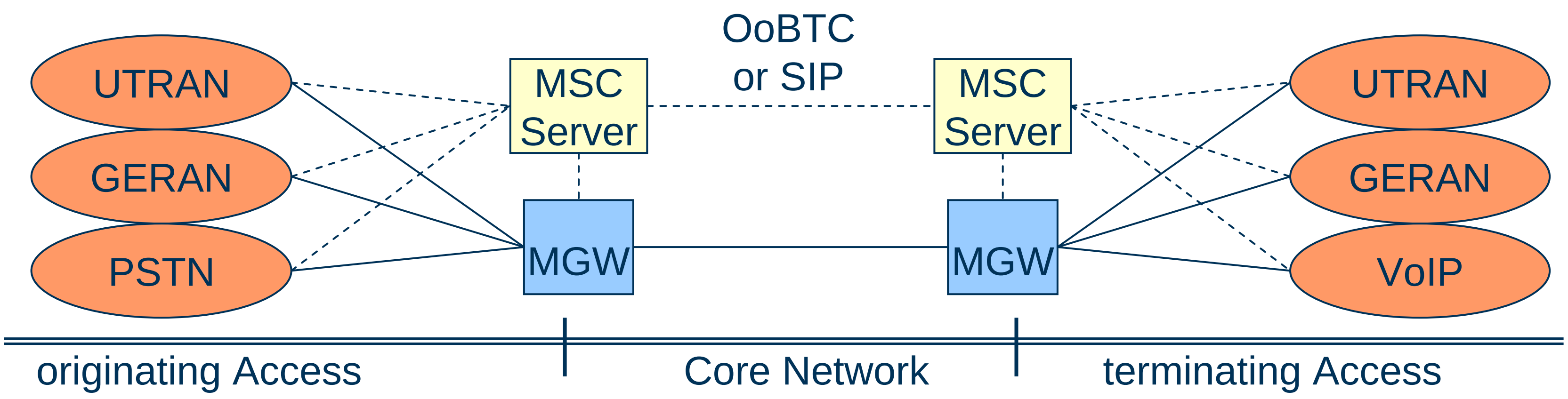
- Identical for all digital terminals, independent of access technology, or Codec Type or Codec Configuration. Also valid for VoIP and IMS terminals. Important:
- Voice Quality Enhancements shall be performed *before* speech compression in the terminal.
- Why: Speech Compression and Radio Errors distort the linear signal path and prevent
  optimal Voice Quality Enhancements in the Core Network.
  DSP resources are no issue any longer for terminals.
  The speech path activity factor goes down as early as possible in the path.



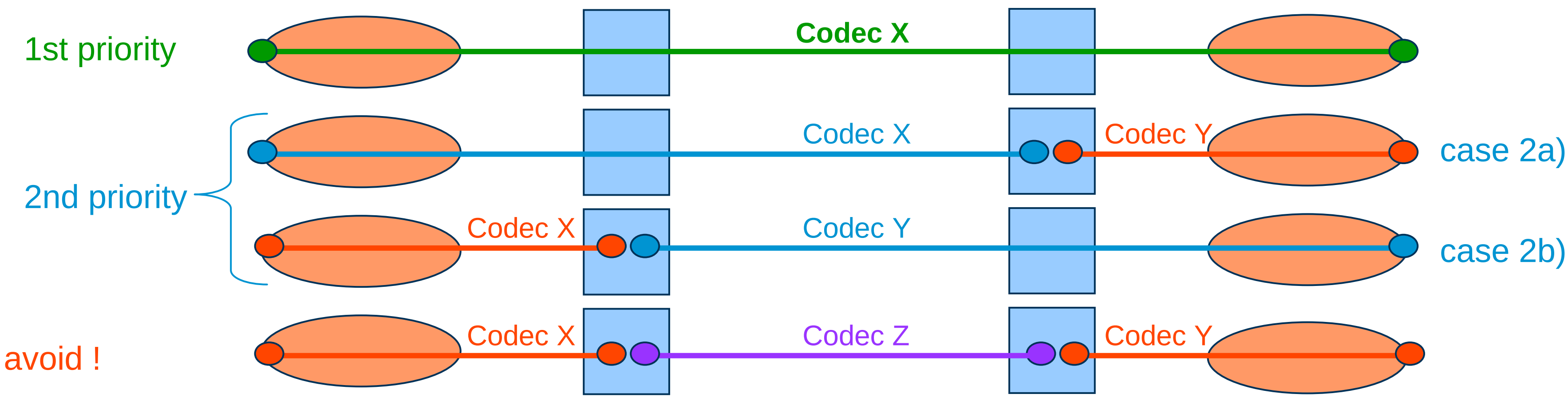
Transcoder-Free- and Tandem-Free-Operation (TrFO and TFO) in Layered Architecture Networks



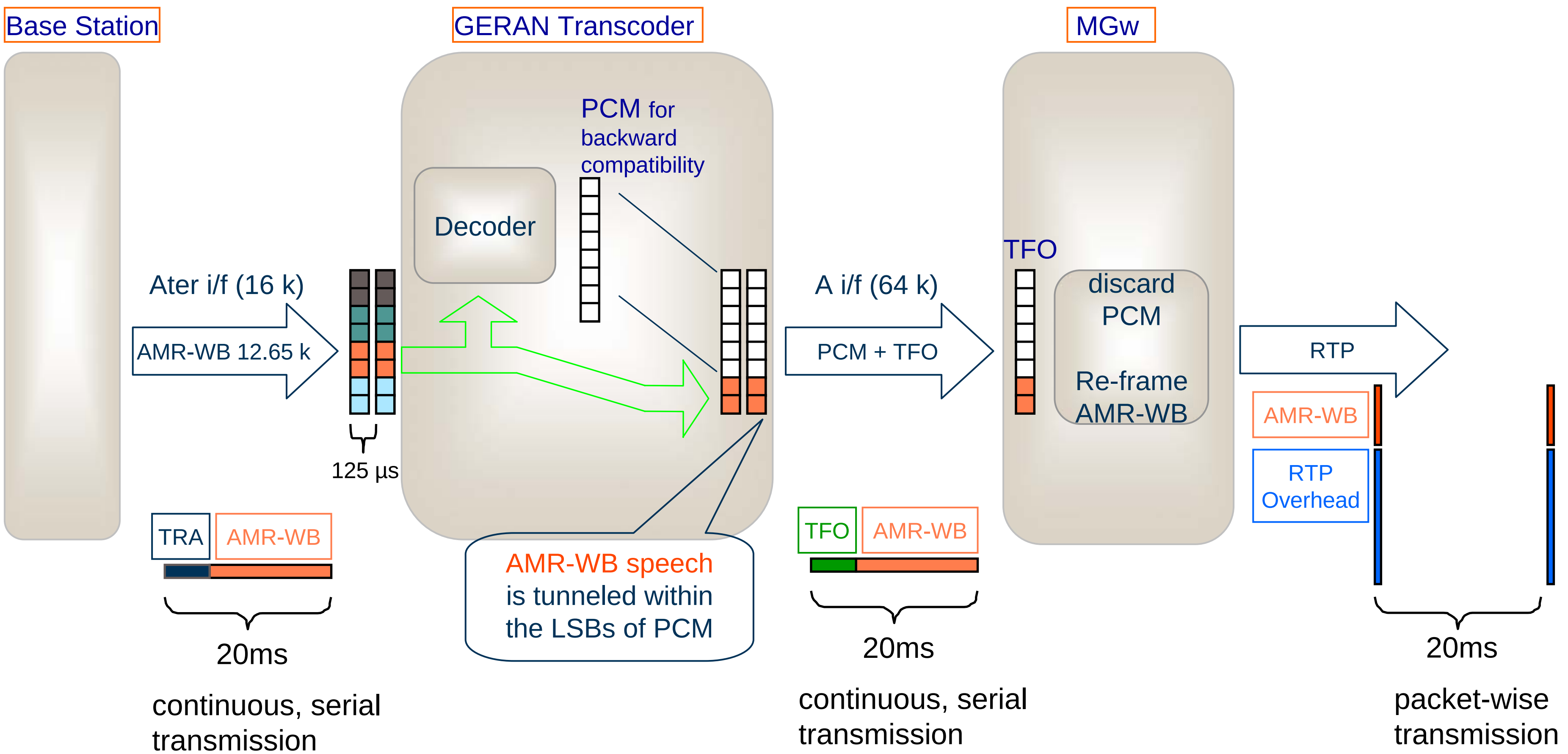
Codec Negotiation: Overall Objective



1. The (radio) Accesses have highest priority: radio resources are expensive and rare
2. Select that Codec-Combination that leads to highest speech quality for this call
3. Obey side conditions, such as Terminal Capabilities and low target bit rate in Core Network
4. Codec Selection shall be independent of call setup direction
5. Select the Codec(s) in such a way that the Core Network follows at least one of the accesses

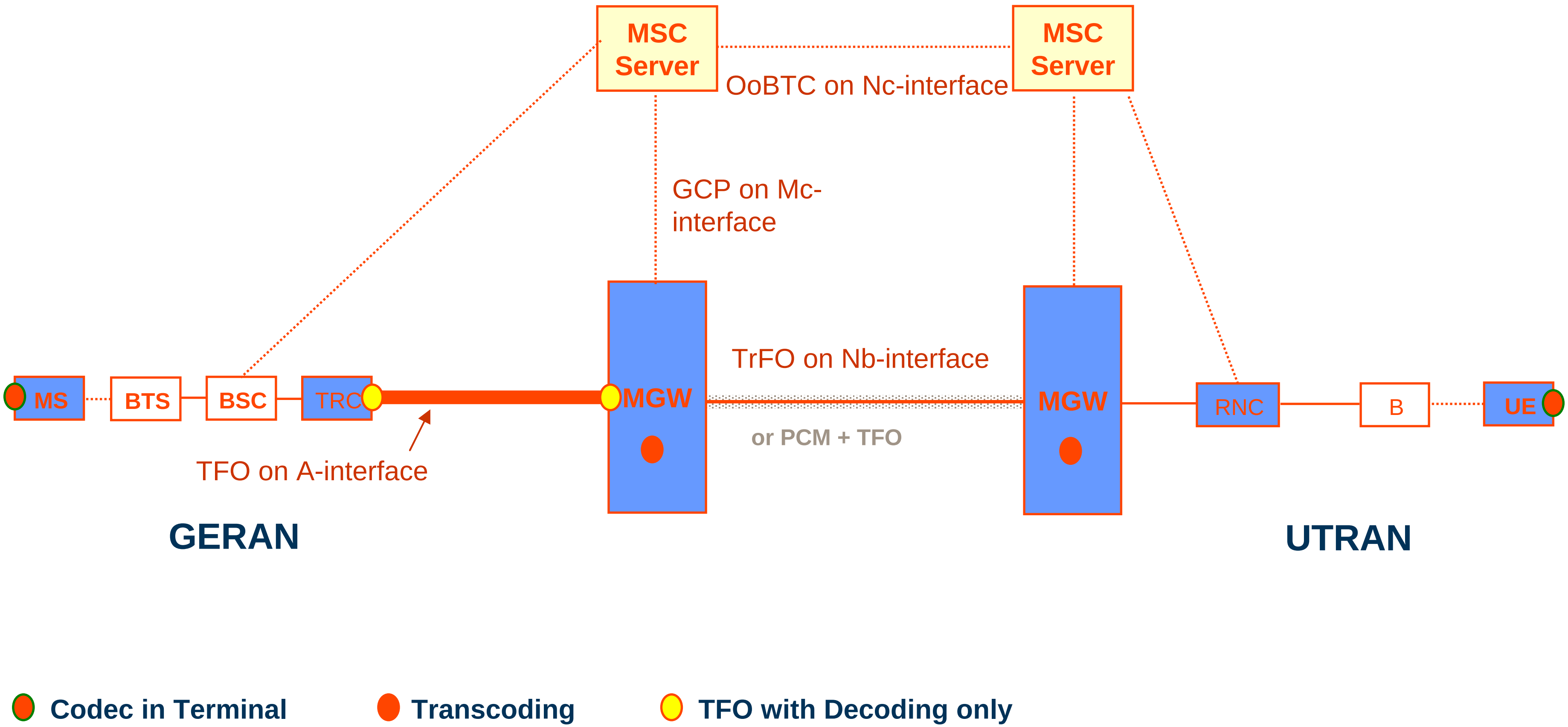


TFO/TrFO Interworking (GERAN to Core Network)



Only one way is shown

List of involved Network Elements



The long way of AMR-WB into commercial networks (1)

- 2001: The WB standard and all prerequisites (TFO, TrFO) are ready.
  GSM and UMTS operators have a unique chance
  to lauch a substantially improved Speech Telephony Service,
  far better than any existing mass market system,
  at same or lower operational costs,
  and with optimal conditions for fast market penetration.
- Operator interest ? ! ? Silence ! ? !
- Why ?
- 2000: The UMTS auctions in UK and Germany have taken about 100.000.000.000 €
  out of the telecom industry; big operators ramp down their investments to a minimum.
  The whole industry suffers and reduces their workforce, often by more than 50%!
  The dream of unlimited exponential growth in mobile communication is out.
- 2001 - 2004: Many presentations of AMR-WB to operators, in PPT slideware and in real-time.
  The reaction is always: Astonishment: Ahhh, what a good quality! We must consider that!
  What are the others doing? Nothing? Ohhh... ... ...biggest questions:
  Does the end-user care about quality? Will we get our investments back?
- These questions can not be answered without asking the end customers !

The long way of AMR-WB into commercial networks (2)

- 2004: T-Mobile Germany and Ericsson agree to undertake a larger field test with end customers.
- 2005: Ericsson, in cooperation with Teleca Systems, Nürnberg, develops a WB-Demo phone, based on a commercial NB-terminal (Z800 from Sony-Ericsson), which uses the existing Video-Telephony Service, but replaces AMR-NB by AMR-WB.
- This demo phone needs an UTRAN 64k bearer on both radio legs for mobile-mobile calls, but it does not need any modification in UTRAN or Core Network.
- 2006: Field tests in Köln and Hamburg, with about 150 users, show very encouraging results, although the WB-Demo phone had to live with a traditional NB acoustic design. Also not perfect UTRAN coverage for this bearer was an obstacle for many users.
- SYSTEMS, October 2006:
  T-Mobile announces firm interest to introduce **AMR-WB in GSM and UMTS.**
  This triggers an avalanche in Operator interest.
- 2007: Most likely we will see the first WB launches in life mobile networks at the end of 2007 or the begining of 2008.
- Done, uhh... !
- Done ? And the rest of the world?

What happens outside Mobile Communication?

- The proprietary **Skype** enters the VoIP market in year 2003. It allows "zero-cost" real-time telephony between Skype users, worldwide. The Skype Codec is adaptive to the IP-channel.
  It provides WB speech quality, if the IP-channel is good enough.

- Some of the "fathers" of the Skype Codec are former Ericsson experts, who believed that the traditional telecom industry is far too big and too slow for really good ideas ...
  well: at least Skype opened the competition and "smoothened" the way to better quality.

- **VoIP** penetrates the wireline world, using often the low bit rate Codecs of ITU-T:
  G.729 (8 kbps), G.723 (6.3 kbps) -> cheap service, but compromised quality.
  More and more the "good, old" G.711 (64 kbps) is used: better, well known quality.
  At this bit rate **G.722** is of course well feasible!

- ETSI updates the DECT standard and now G.722 can be used in Cordless Telephony .

- DSL and WLAN penetrate the home environment and more and more mobile terminals allow IP telephony via WLAN, soon also with WB Speech.

- 3GPP creates the **IMS Standard** (IP Multimedia System) to realize "All-IP" in Mobile Communication and ETSI adapts IMS for Wireline Application. Very good steps on the way to "Fixed-Mobile Conversion".
  Just now the "Multimedia Telephony Service on IMS" (MTSI) is ready in 3GP, including AMR-WB.

- ITU-T upgrades the G.729 standard with "bit slice layers" for improved NB speech (12 kbps) and extending it to WB speech (14 kbps to 32 kbps) (**G.729.1**).

- => The telecom world is changing fast and substantially:
  WB speech quality will be on the mass market very soon, but
- **Will we start with a Codec-Chaos?**

How many Codecs are reasonable?

WB Codec Candidates for important mainstream systems are:

- 3GPP: CS Telephony: AMR-WB (== G.722.2)
- 3GPP: IMS for mobile: AMR-WB
- ETSI: IMS for wireline: AMR-WB, G.722, G.729.1
- 3GPP2: CS and IMS: EVRC-WB

- Different Codecs in different networks require transcoding for interworking.
  Transcoding always costs
  - Speech Quality
  - Speech Delay
  - DSP Resources
  - End-to-end Encryption
  - Implementation and Verification effort
  - ...

  without any gain
- The Introduction of a new Codec must be justified by a substantial improvement !
- Different Codecs require more complex Codec Negotiation:
  how to find the optimal Codec Combination along a complex call path?
- Too many Codecs split the market and slow down the introduction

My Vision

- The Mobile Phone will be a true personal phone, with many rich services.
- The Mobile Phone outnumbers any other phone by far
- The Mobile Phone will be used always and everywhere, via different radio accesses, e.g.:
  - GERAN in sparcely populated (rural) areas (macro cells)
  - UTRAN and modern cellular RANs in cities (small / medium cells)
  - Picco-BTS, WLAN (or BlueTooth) at home and office (picco cells)
  - ...
- => No need to change the phone due to cost reasons or coverage reasons
- With good acoustic design (!) the mobile phone provides very satisfying quality
- => No need to change the phone due to quality reasons
- => The Mobile-Phone-inherent Codec(s) will dominate the market

  => few transcodings may be needed, that will be acceptable

A look into the future

	Todays mobiles 12 kBit/s	AMR-WB 12 kBit/s	AMR-WB+ 12 kBit/s	Original (1.5 Mbit/s)
Speech				
Musik				

AMR-WB+ Encoder is too complex for todays phones and has a too long coding delay to be used in real time (… maybe in 5 years?)

AMR-WB+ is an excellent compromise between cost and quality for music and news-clip streaming (just decode)

ERICSSON

TAKING YOU FORWARD