

# Superior Error Concealment and Robust Audio-Video Synchronization in Set-Top Box Chip

● Jiang Li ● Yan Yingrui ● Ni Xin

Fujitsu's MB86H06 set-top box (STB) chip is a high-density chip featuring high-performance, superior error concealment, robust audio-video (AV) synchronization, and low cost. It comprises a QFP256 package with DDR2 memory and supports a two-layer printed circuit board (PCB) layout for a low bill of material (BOM) cost. It is widely used in the China terrestrial television broadcasting (CTTB) market. In this paper, we discuss the practical realization of error concealment and AV synchronization under weak signal conditions.

## 1. Introduction

More and more set top box (STB) vendors and chip vendors are entering the digital broadcasting market in China as digital broadcasting becomes increasingly popular, from major global players like Broadcom and STMicroelectronics, to low-price local vendors. This increase in competition has led to more and more demands being placed on chip vendors. Chip designers are being challenged to meet multi-dimensional requirements, including increased integration, reduced chip cost, enhanced performance, lower power consumption, smaller packages, and more printed circuit board (PCB) layers. Moreover, to utilize intellectual property (IP) prepared by IP providers is making chip differentiation more and more difficult.

Although video decoder (MPEG-2, H.264, etc.) development has quite a long history, different markets still have different requirements. For example, high-performance high-definition (HD) decoders, especially dual-decoding ones, require reduced bandwidth and increased throughput. These requirements can be met through the use of cache technology; motion vector (MV) merge technology, and special memory mapping.

Fujitsu's MB86H06 STB chip is targeted at the China terrestrial television broadcasting (CTTB) market and has been widely implemented by many local STB vendors. It is designed for standard density (SD) decoders. Its use for terrestrial broadcasting means that its requirements are driven by the characteristics

of terrestrial broadcasting. A key requirement in terms of competitiveness is reducing mosaicing under weak signal conditions.

This paper introduces the problems presented by terrestrial broadcasting, overviews the MB86H06 STB chip, and discusses how superior error concealment and robust audio-video (AV) synchronization are achieved.

## 2. Problems presented by terrestrial broadcasting

Terrestrial broadcasting is characterized by weak (or no) signal conditions, for example, when a moving vehicle passes across the line of sight, when neighboring buildings create a complex RF signal environment, or when the broadcasting stations are distantly spread, as they often are in rural environments. Under such conditions, the signal may become broken up during a broadcast, creating mosaic images. In such cases, STB vendors and/or operators are likely to perform side-by-side comparisons of STB boxes from different vendors to see which ones have less mosaicing.

STB performance can be improved under such conditions by improving error concealment and by making AV synchronization more robust. Error concealment reduces the number of artifacts in the image, and robust AV synchronization reduces the AV synchronization time.

### 3. MB86H06 Overview

#### 3.1 Overview

Fujitsu's MB86H06 STB chip is a high-performance chip well suited for various applications, including digital video broadcast (DVB) satellite TV, terrestrial TV, and DVB terrestrial TV. It is widely accepted in the CTTB market due to its superior error concealment and robust AV synchronization.

The MB86H06's advanced features support multi-standard decoding: H.264, MPEG-2, MPEG-4, VC1, and Audio Video Coding Standard (AVS). The chip includes a 324-MHz ARC Tangent A4 CPU with 8kB/8kB instruction/data (I/D) cache that is connected to shared 8-bit DDR2 SDRAM (second-generation double-data-rate synchronous dynamic random access memory) through a 162-MHz bus.

The MB86H06 chip supports multiword direct memory access (DMA) for IDE (integrated drive electronics) devices and includes a USB high-speed on-the-go (OTG) link controller for connection to USB hosts and devices. It supports 1x/2x/4x series flash boot-up, which reduces the package pin count and increases the boot-up speed. Access was changed from 16-bit DDR1 to 8-bit IO DDR2 to reduce the pin count and bill of material (BOM) cost.

##### 1) Features:

- Two high-performance transport stream decoders
- Multi-standard video decoder support for H.264 MP@L30, MPEG-2 MP@ML, MPEG-4 SP/ASP, XVID, DivX, VC1, and AVS
- Flexible video scaler
- Three-layer on-screen-device (OSD) controller
- High-performance 2D graphics engine
- De-interlacer for interlace to progressive conversion
- Flexible frame rate conversion, flicker, and flutter fixer for better image quality
- PAL/NTSC/SECAM digital encoder, including Teletext/WSS/VPS/CC/VBID insertion, RGB de-matrix
- Audio decoding DSP support for MP1/2/3, AAC, HE-AAC v1/v2
- Embedded quadrature amplitude modulation (QAM) demodulator

##### 2) Interface:

- High-Definition Multimedia Interface (HDMI) link and physical (PHY) interface
- ITU-R 656 video input/output
- Four video digital audio converters (DACs) for

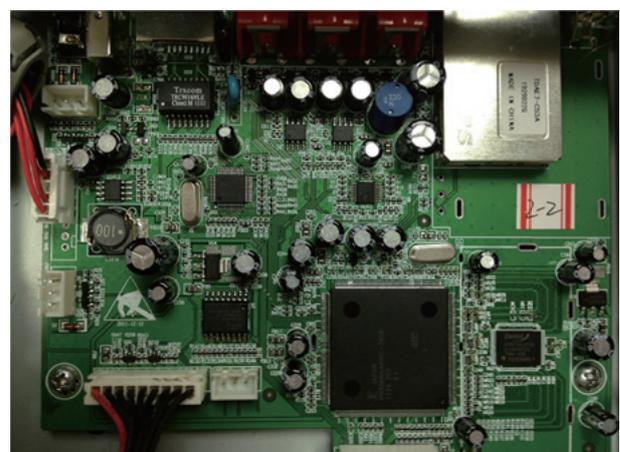
YPbPr/RGB, YC, and CVBS input

- S/P DIF output for PCM/AC3/Dolby 5.1
  - USB 2.0 including PHY interface (host and device)
  - Ethernet 10/100 base-T MAC with Reduced Media Independent Interface (RMII)
  - 59× shared general purpose input and output (GPIO), with universal asynchronous receiver and transmitter (UART), 2× smart card, I<sup>2</sup>C standard, pulse-width modulation (PWM), front panel, infrared receiver, secure digital input output (SDIO)
- ##### 3) Package / Technology:
- QFP256 package
  - Ambient temperature: -40~+80°C;
  - Fujitsu 90-nm technology 1.2-V core

#### 3.2 Video decoder

The video decoder is a small high-performance integrated package that supports dual HD decoding at 200 MHz. However, in this application, it is used only as an SD decoder running at very low frequency (27 MHz) to save power. Several HD decoding features (e.g., video data cache) are cut to reduce cost. A photograph of the MB86H06 PCB board is shown in **Figure 1**. A block diagram of the MB86H06 set-top box chip is shown in **Figure 2**, and the video decoder structure is shown in **Figure 3**.

It is embedded in an ARM7 CPU and processes the video stream received from the packet elementary stream (PES) layer. Decoder firmware handles the decoding of the picture layer and the layers above it as



**Figure 1**  
MB86H06 system PCB board.

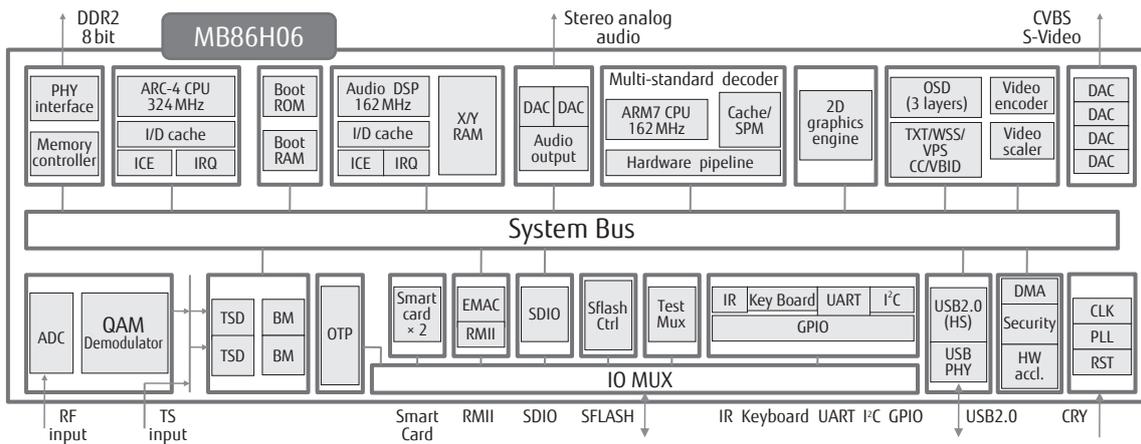


Figure 2 MB86H06 system block diagram.

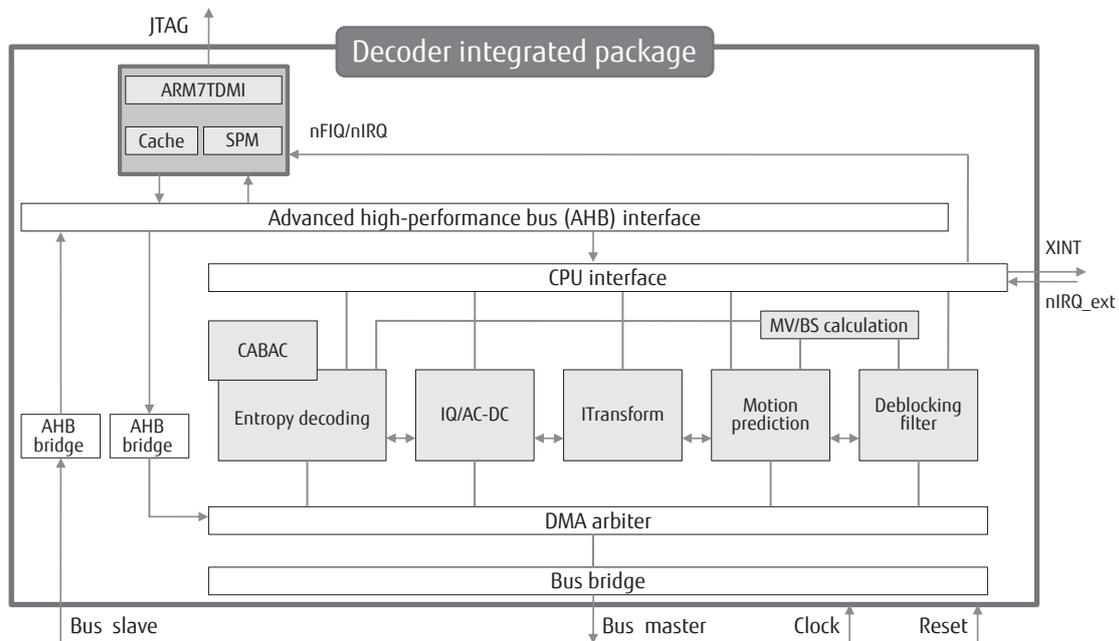


Figure 3 Video decoder structure.

well as AV synchronization control. The hardware (HW) pipeline handles the decoding of the slice layer and macro-block (MB) layer.

## 4. Error concealment and AV synchronization

### 4.1. Error concealment

The MB86H06 is partitioned into two parts (HW and software) to achieve error concealment.

HW is responsible for detecting errors below the

MB layer. This detection includes

- Copying the reference MB of the nearest MV when an error stream is detected to avoid mosaicing for each MB layer
  - Checking the syntax range
  - Internal error checking of coefficient numbers larger than 64 in an 8×8 block
  - Checking the decoding error status of consumed bits beyond the start-code boundary.
- When such errors occur, a strategy for concealing

them is realized by firmware. For some errors, decoding is stopped, firmware interrupted, and an indication from firmware on what to do next awaited. For other less serious errors, the system is configured to continue decoding and to set an error flag. Most of these detected errors have on/off switches and are configurable in registers by firmware. HW is also responsible for data processing. For example, it supports several different modes of reference frame data copying to the missing parts of the pictures.

Software (SW) is the "brain," telling HW how to conceal the errors. It first prioritizes the different types of reported error conditions and treats the errors accordingly. It helps locate errors in the MB layer above the syntax. For example, a discontinuity in the frame numbers in the picture parameter indicates one or more missing frames. MB86H06 supports three levels of error severity: 1st level includes recoverable errors (the errored data value can be discarded, given a default or guessed value, and decoded again), 2nd level includes data-recovery-needed errors (system stops decoding and re-synchronization on a small scale, i.e., the next slice boundary, to identify the missing parts and then copies reference data to recover the picture data), and 3rd level includes unrecoverable errors (system tries to re-synchronize at the next I-frame and group of pictures [GOP] and discard all pictures between them; display may be stalled for a while). Of course, these strategies are configurable by the system CPU.

Compared to MPEG-2, H.264 has a special context-based adaptive binary arithmetic coding (CABAC) entropy decoder. The key problem with CABAC is the lack of an error detection method. When a stream enters the CABAC engine, no matter whether it is errored or not, it has no boundary, and the CABAC decoder has to support unlimited bin-number syntax decoding. This means that there is no awareness of whether an error occurs, so neither hardware nor CPU software can know what happened in the stream decoding. A method for CABAC error detection must be created for it to become practical. We used two strategies to enable such detection.

- The value range of almost all parameters is checked. If the value of a decoded parameter exceeds its range, an interrupt request (IRQ) is generated to pass this information to the CPU.
- Additional conditions like the boundary

completeness of the trailing bits are added, which enables knowledge of error occurrence to be obtained more precisely and quickly.

Of course, a solution cannot be determined in advance for each case in the real world, but, thanks to our flexible HW/SW architecture, errors can always be handled quickly and easily.

- Some decoders do not test their MV range and occasionally their MVs may be out of the specified range. Although this is a real error according to the AVS specifications, the HW in our video decoder can easily toggle-off each check, including this one, so it can easily handle this case.
- A frame's presentation time stamp (PTS) may be completely wrong, but the use of SW to control display means that firmware can easily find this error and use a dummy PTS value for proper display.
- The MB86H06 supports skipping the display of single non-reference frames when the error is infrequent and does not propagate, which helps to prevent an undesired single mosaic, resulting in a smooth display for the user.
- The MB86H06 supports total control of the select/copy reference data area (thus the mosaic display area) in firmware, so more concealment strategies can be subsequently added.

## 4.2. AV synchronization

### 4.2.1 Conventional DVB receiver system

A digital multimedia stream can be transferred to a receiver through various links: satellite networks, cable networks, terrestrial networks, and other types of networks. Both the video and audio signals are conventionally formatted using the MPEG series standard, in which a program clock reference (PCR) is transmitted periodically. The PCR gives the receivers the correct time for use in decoding and/or display. Conventional DVB uses a PCR to synchronize the receivers to the transmitter. Each receiver determines the difference between the system time clock (STC) and PCR and uses it as a reference for adjusting its voltage controlled oscillator (VCO). As a result, the receiver's digital phase lock loop (DPLL) has the same frequency as the transmitter, so AV synchronization is achieved.

The conventional DVB system works well for isochronous and constant delay networks, like cable,

in which each packet is sampled before transmission, and the magnitude of the “gap” or “jitter” in the time domain is strictly guaranteed, so the precision achieved by using the PCR is adequate for playback of a multimedia stream.

However, the delay is variable in some networks (e.g., Ethernet), so the packet delivery time is not constant, and the difference between STC and PCR can vary greatly. The conventional DVB system fails in such situations: the receiver never becomes synchronized and/or the difference in the frequency domain is never eliminated. This can lead to two possibilities. 1) The receiver runs faster than the transmitter, so the amount of data in the buffer (the “fill level”) decreases, resulting in underflow and eventually video stutter. 2) The receiver runs slower than the transmitter, so the buffer fill level continues to increase, resulting in overflow and eventually video freeze. Both possibilities are unwelcome by the viewer.

#### 4.2.2 MB86H06 chip

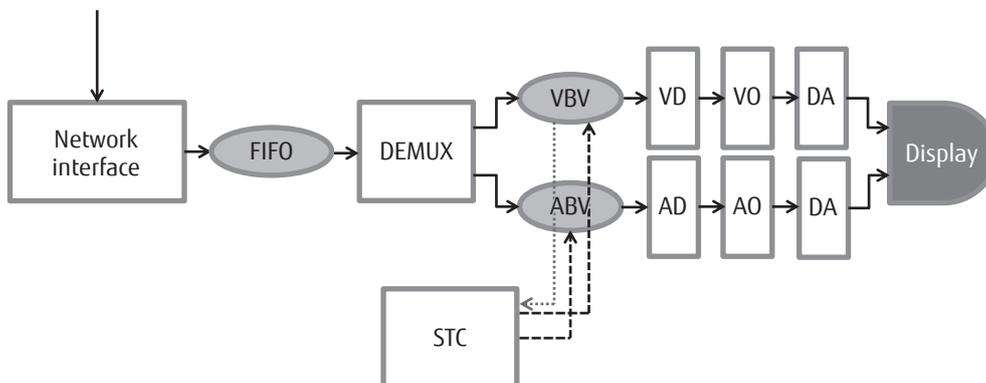
The target market for the MB86H06 is not only DVB but also Internet protocol switch networks, so AV synchronization must work for both isochronous and non-isochronous networks. Since the buffer fill level for AV data varies within a certain range if the system is synchronized, the clock frequency is determined by the buffer occupancy level. The frequency is reduced if the buffer level is relatively low and increased if it is relatively high. This approach is applied to both the audio and video buffers. Here we focus on video because there is relatively less video data than audio data for a given bitrate.

The AV synchronization process setup stage is illustrated in **Figure 4**.

First, the fill level of the video buffer is periodically monitored, and an average value is computed. This is done periodically to prevent excessive frequent adjustment because the video buffer level can be adjusted fairly quickly in a short observation interval. Once the fill level and bitrate have been determined, the threshold is determined taking into account the various bitrates used by different programs. In other words, at which level is the buffer level considered to be low so that the frequency needs to be decreased and at which level is the level considered to be high so that the frequency needs to be increased. This first phase may take 5–10 seconds.

Second, if video decoding has started and a PTS, which is usually attached to an I-frame, is received while doing the first step, the initial STC is set based on a different policy. One is use the PTS minus a fixed value as the STC. The other is a bit more complicated: decode until all the decoded picture buffers are full, then set the first PTS as the STC. Now we choose the simpler logic.

Third is long-term tuning. Once the second phase has ended, a reasonable threshold for the buffer fill level has been determined, and an appropriate STC initial value has been set. However, the local frequency may not be the same as that of the transmitter, so the receiver frequency is synchronized with that of the transmitter by adjusting the DPLL. The buffer fill level is periodically monitored, and, if the fill level is higher than the high threshold level, the frequency is increased in unit steps; if the fill level is lower, the



**Figure 4**  
AV Synchronization in MB86H06.

frequency is reduced in unit steps. We set a maximum adjustment because a large adjustment would affect system stability; moreover, if the difference was so big that synchronization was broken, another method would be needed to achieve synchronization. This phase is long term and keeps on working because the frequency of the oscillator cannot be expected to be exactly the same as that of the transmitter.

### 4.2.3 Special cases

There are several special cases that must be considered. Two in particular are discussed here: errors in transmission and PTS jitter, which causes discontinuity.

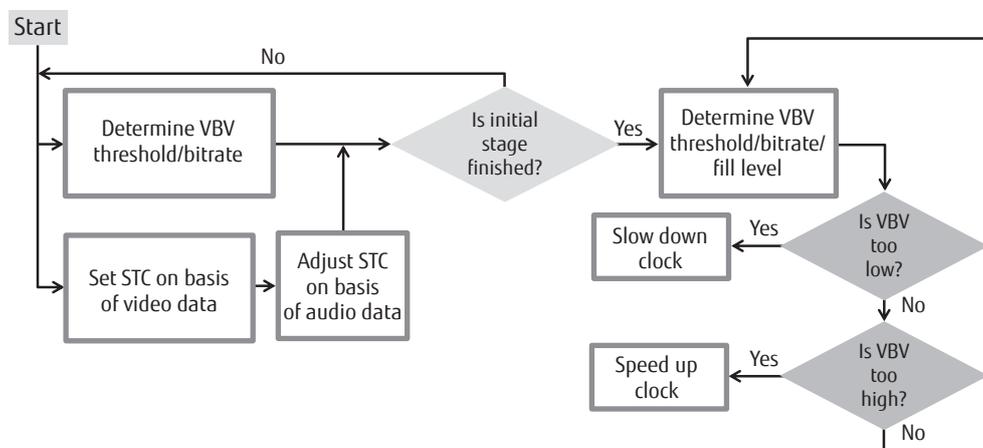
When there are errors in transmission, the initial STC must be adjusted. As shown in **Figure 5**, in MB86H06 AV synchronization, the STC is initially set on the basis of the video data. It is then adjusted on the basis of the audio data if needed. Generally, either there is no need to adjust for the audio data or the adjustment is trivial. However, in some cases, there is a big gap between the audio and video PTS, the audio adjustment must be made on the basis of the status of the video decoder buffer. If the fill level of the video buffer is relatively high, a decrease in the STC will probably cause video data to overflow. In contrast, an increase in the STC will probably cause video data to underflow. This means that some evaluation is needed before an adjustment is made. Since knowledge of the average bitrate for the current video stream has already been obtained, it is a simple matter to get the timing information. For example, how long will it take for the

stream data remaining in the buffer to be consumed? If the change in the STC fulfills the requirement for preventing overflow and underflow, the STC is changed. Otherwise, the audio will enter into a free run state. However, the free run state is unexpected, so the above condition is checked periodically to see if the state can be left.

PTS jitter comes from two sources: stream errors and remultiplexing (remuxing). Normal stream errors need no special attention because stutter, distortion, and mosaicing are evident only at the error point. There is no obvious difference in the phase between the receiver and transmitter, so, after the error point, playback quickly recovers. In contrast, a remuxing error needs special treatment because it creates a bigger phase difference than the error duration. If the difference behaves like a gap, the fill level is increased; if it behaves like an overlap, the level is reduced. Since there is no difference in frequency, frequency adjustment cannot be used to restore the fill level. Moreover, such adjustment is very slow. If this situation is not dealt with, an overflow or underflow situation will eventually arise.

A two-step method is used for handling this situation in the MB86H06 chip. 1) The difference between two adjacent PTSs is checked. If the gap exceeds a predefined threshold, it is regarded as a discontinuity point. The threshold is set relatively large because a misjudgment may be made if it is small. 2) Recovery is attempted.

If the situation is not dealt with, the fill level of



**Figure 5**  
AV Synchronization flow chart.

the video buffer will eventually increase or decrease. This can be monitored, and appropriate measures can be taken. For instance, a skip/repeat can be done to force the fill level to increase or decrease as needed.

#### 4.2.4 Further improvement in AV synchronization

Although the AV synchronization in the MB86H06 chip has been shown to work well in practice, several improvements can still be made. The DPLL adjustment process can be made more flexible. Rather than using a fixed step to make the adjustment, the step could be adjusted on the basis of the buffer change rate to enable a quicker response. The treatment of discontinuities could be refined. Rather than continue

to decode the stream at the point of discontinuity, the decoding could be restarted in some cases to improve the user experience.

## 5. Conclusion

Our set-top box chip is widely used in the China terrestrial television broadcasting market due to its superior error concealment and robust audio-video synchronization under weak signal conditions. The error concealment is based on partitioning between hardware and software, and AV synchronization is based on periodic monitoring of the video buffer. Ways to make the synchronization even more robust are being investigated.



**Jiang Li**

*Fujitsu Semiconductor (Shanghai) Co., Ltd.*  
Mr. Jiang graduated and received the B.S and M.S degree from the Department of Electronics Engineer, Fudan University, China, received Doctor of Engineering degree from Tokyo Institute of Technology, Japan, in 1999. After that, he joined Fujitsu Laboratories Ltd. From 2005, he joined Fujitsu Semiconductor (Shanghai) Co., Ltd., and worked on STB LSI development.



**Ni Xin**

*Fujitsu Semiconductor Design (Chengdu) Co., Ltd.*  
Mr. Ni received the B.S. and M.S. degrees in information engineering from Zhejiang University, Hangzhou, China in 2004 and 2006. He joined Fujitsu Semiconductor Design (Chengdu) Co., Ltd., China in 2011 and was engaged in research and development of video codec driver & firmware.



**Yan Yingrui**

*Fujitsu Semiconductor (Shanghai) Co., Ltd.*  
Mr. Yan received his B.S. degrees from Shanghai Jiao Tong University. He joined Fujitsu Semiconductor (Shanghai) Co., Ltd. in 2005 and was engaged in research and development of IC's for STB video encoding/decoding system.