Low-latency design considerations for video-enabled drones

Texas Instruments

Dennis Barrett
Product Marketing Manager

Prajakta Desai
Product Marketing Engineer
Texas Instruments

Introduction

Over the next decade, it is estimated[1] nearly USD $98 billion will be spent globally on aerial drones and other unmanned aircraft. These drones are not solely for hobbyist and government use—the commercial applications for drone technology are forecasted to drive an emerging global market for business services valued at over $127 billion[2]. Industries that can expect to be rapidly disrupted by drones include infrastructure, agriculture, transport, security, entertainment and media, insurance, telecommunications and mining.

As the presence of drones becomes commonplace, customizing them to address specific use-cases and industries is accelerating. A natural grouping or categorization is occurring based on key aerial features and flight requirements and a clear need has emerged for flight modes beyond visual line of sight (VLOS) operation.

A key requirement for enabling safe flight in drones across flight operation modes is low-latency video transmission. This paper will describe the design considerations engineers should take before developing such systems.

Drones take flight across an array of operating ranges

Drones can be categorized in various ways. For the sake of discussion in this paper, the categories are based on operating ranges between the remote pilot and the drone.

- **Visual line-of-sight (VLOS) operation:** The pilot on the ground is always able to see the drone without visual aids (like binoculars, telescopes, etc.) The drone should always be in unobstructed view of the pilot, where it moves behind trees or in thick clouds or fog. Certain countries have additional limits on the actual distance to be considered for VLOS operation.
- **Extended visual line-of-sight (EVLOS) operation:** The pilot commanding the drone

Figures 1 and 2: Drones serve a wide array of use-cases, including infrastructure, agriculture, transport, security, entertainment and media, insurance, telecommunications and mining.

may rely on other remote observers who are in visual line of sight of the drone. The remote observers must be able to relay critical flight information to the commanding pilot in real time.

- **Beyond visual line-of-sight (BVLOS) operation:** The drone is operated remotely based on instrumentation, control and communication between the aerial drone and a remote ground-piloting station. The drone is allowed to go beyond visual range. An on-board camera-based system (like first-person view operation) is usually employed but not sufficient to allow BVLOS operations. Additional levels of autonomy like "sense and avoid" technologies are deployed on these systems for safety. This also requires higher operator or pilot qualification and experience.
- **First-person view (FPV) operation:** The remote pilot utilizes on-board video cameras to provide a real-time view from the unmanned system and operates it based on this video. FPV operation is popular with recreational flyers. This type of drone is also used to collect sensor and imagery data while in flight.

Requirements for video-enabled drones

A key feature enabling drone operations beyond direct visual contact is the on-board camera with real-time transmission of the video to the remote operator. The most important requirements for this type of system in unmanned drones are mentioned below.

- **Low-power consumption:** The amount of time a drone can stay up in air all depends on how much power the system consumes. With this in mind, low-power consumption by all of the on-board components is essential. The power consumption rate directly impacts the time a drone can fly. Some drones are fitted with a "safe landing" feature when the system is close to running out of power to prevent the drone from falling out of the sky when it fully depletes its' battery.
- **Low latency:** In order to have control of the drone, the remote operator needs to react quickly to events happening around the drone. This means that the latency starting from the data collection on the drone to that being received by the remote operator must be minimized. This is especially true for video-enabled drones. In order to effectively transmit and reduce bandwidth, developers often employ video compression techniques. The designers therefore need to choose components that enable low-latency video compression and transmission.
- **Wireless link robustness:** The wireless link between the controller and drone serves as the foundation for accurate and responsive drone control. A wireless link must ensure frame delivery, even in the most congested RF environments. This can be achieved by either using a "clean" channel (for example, the less-congested 5-GHz band) or advanced rate control and receiver algorithms that ensure in-time delivery with minimal packet loss.
- **Range:** It is more desirable for a drone to be able to maximize operational range for the sake of flexibility and mission performance. For line-of-sight and video-enabled drones, the wireless communication link must enable very-long-range transmission and reception. This is done by using a connectivity solution with very-low sensitivity threshold on one hand, but high transmission power on the other.

- **Autonomy:** Drones are being equipped with additional sensors to increase their level of autonomy. Sensors can now determine not only obstacles in the drone's path, but also distances to those obstacles. These sensors include ultrasound, millimeter wave sensing, as well as vision-based stereo cameras. These sensing modalities are augmented using computer-vision-based algorithms to identify objects, and take appropriate actions to avoid them. Such a mechanism is often known as "sense and avoid".

Low-latency video compression and transmission

There are several ways that developers can introduce latency in a video compression and transmission system used in drones.

- **Video capture:** The higher the frame rate, the lower the capture time, T_{cap}. For example, a 30-fps camera takes 33 ms to capture each frame of video. This number reduces to 16.5 ms for 60-fps video capture.
- **Compression or encoding:** Compression techniques are used to reduce the data rate needed for transmitting video frames. The H.264 compression standard is a very common technique for recording and compressing video in drones. Compression is, in general, a compute-intensive task. The time required to encode, T_{enc}, depends on the choice of encoding engine and the features used.
- **Transmission:** The drone communicates to the ground station using a wireless communication mechanism like Wi-Fi® connectivity. The resulting transmission delay, T_{tx}, depends on the available data bandwidth. For example, if a 720p30 stream is encoded at 1 Mbps and the available bandwidth is 2 Mbps, the time taken to send a stream to the ground station is 16.5 ms.
- **Network:** Depending on the need, the aerial system may be connected to the remote ground station via a network. In that case, there may be an additional delay, T_{nw}, within the network.
- **Receive:** If the ground station is also wirelessly connected to the network, then additional latency, T_{rx}, similar to that of transmission is involved in the system.
- **Decompression or decoding:** The compressed video stream needs to be decompressed at the receiving station. Like encoding, this decoding process is also compute intensive. This will introduce a decoding delay, T_{dec}, in the system.
- **Display:** Just like video capture, there will be display latency, T_{disp}, depending on the display refresh rate.

Note that when the drone is communicating directly with the ground station, there is no network involved and there is only a single transmission delay, T_{tx}, present (i.e., T_{nw}=0 and T_{rc}=0).

On a system where the operations mentioned above are performed frame by frame, the total latency from capture to display is $T = T_{cap} + T_{enc} + T_{tx} + T_{nw} + T_{rx} + T_{dec} + T_{disp}$. This is illustrated in Figure 3 on the following page.

Table 1 (on the following page) illustrates the latency with a specific example.

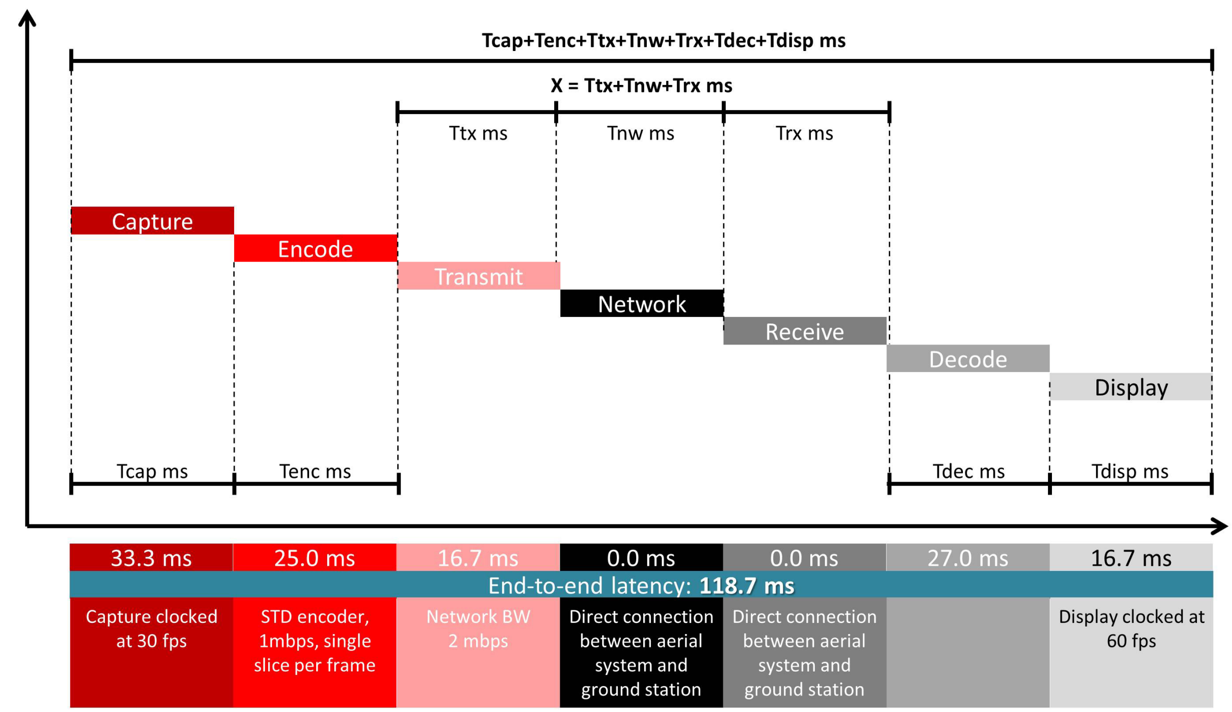

Figure 3: Video capture and display timeline.

System blocks	Latency in ms	Comments
T_{cap}	33.3	Capture clocked at 30 fps
T_{enc}	25	STD encoder, 1 Mbps, single slice per frame
T_{tx}	16.7	Network bandwidth 2 Mbps
T_{nw}	0.0	Direct connection between aerial system and ground station
T_{rx}	0.0	Direct connection between aerial system and ground station
T_{dec}	27	
T_{disp}	16.7	Display clocked at 60 fps
End-to-end latency	118.7	Assume zero network latency

Table 1: Frame-by-frame latency example.

This example provides a scenario with very high latency for controlling drone operation. It takes 118.7 ms for the operator to see the collected video. If the aerial system is traveling at 15 meters per second (approximately 30 miles per hour), the aerial system has moved 1.8 meters when the remote operator senses the need for a flight change and issues that directive. As a result, the moving drone could have reached an environment where the directive may induce erratic behavior. A sequence of erratic directives can cause loss of control and force the drone to land in an unauthorized territory or even hit an object.

In order to allow low latency video encoding, the H.264 standard introduces the concept of slices. A slice is composed of several macroblocks (a macroblock is a two-dimensional unit of a video frame) and is encoded independently, therefore it can be decoded by itself without reference to any other slice. The order of macroblocks inside a slice is quite flexible. However, to be most efficient in low-latency encoding, the natural row order slices are used. When the number of slices per frame is one, it reduces to frame-by-frame encoding discussed above.

However, when the number of slices in a frame is more than one, developers can reduce not only the encoding time but also the overall latency.

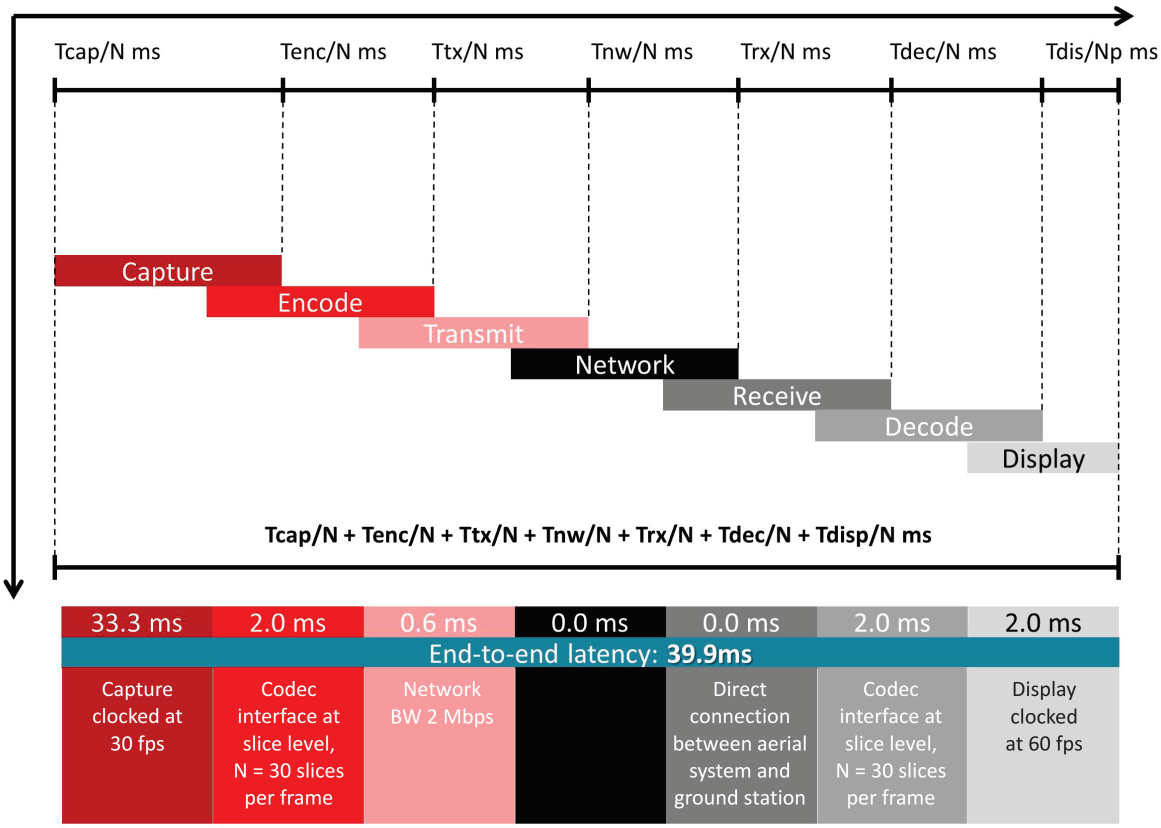

Figure 4: Slice-based impact on processing timeline.

The system doesn't have to wait for the whole frame to be captured before starting encoding. As soon as one slice is captured, the encoding process can begin. Similarly, as soon as a slice is encoded, its' transmission can begin and so on. The impact is that the capture, encode, transmission, receive, decode and display is no longer serial but can be parallelized with considerable overlap amongst these operations. At each step, it introduces a theoretical reduction of delay by a factor of N. The overall latency is then, $T = T_{cap} + (T_{enc} + T_{tx} + T_{nw} + T_{rx} + T_{dec} + T_{disp})/N$. This is illustrated in Figure 4.

Theoretically the effective time will be reduced by a factor of N from encode through display, however in practice, they may not scale linearly with the number of slices. This is due to overhead necessary in setting up and processing individual slices. In Table 2, an example of latency on slice-based encoding established on a rate of 30 slices per frame is shown.

System blocks	Latency in ms	Comments
T_{cap}	33.3	Capture clocked at 30 fps
Effective T_{enc}	2	Codec interface at slice level, N = 30 slices per frame
Effective T_{tx}	0.6	Network bandwidth 2 Mbps
T_{rx}	0.0	Direct connection between aerial system and ground station
Effective T_{dec}	2	Codec interface at slice level, N = 30 slices per frame
Effective T_{disp}	2.0	Display clocked at 60 fps
End-to-end latency	39.9	

Table 2: Slice-based latency example.

As illustrated above, even if an effective encoding-decoding time (representing 1/30 of a single slice) was not achieved, the latency with 30 slices per frame is 1/2 of that for frame-based encoding (single slice per frame). This allows the remote pilot (and therefore the drone) to react to events three times faster.

There is a trade-off on the number of slices and the amount of compression. The higher the number of slices, the faster it can be encoded and transmitted. However, it reduces the compression ratio, and increases the number of bits used for a slice and the effective transmission time for each slice. The drone designer should have a choice to decide this parameter so he can optimize his end-to-end system. Any solution needs to provide the requisite flexibility so as not to limit this choice for the designer.

Robust wireless link

Another important feature needed for low-latency video transmission is a robust wireless link. Here are several features in wireless links that are important to maintain robustness:

- **Antenna diversity:** This feature uses multiple transmit and/or receive antennas and is most commonly used with multiple transmit antennas. The link condition from a given transmit antenna to a given receive antenna may be noisy. Having multiple transmit antennas, along with the capability to switch between these antennas, allows the system to choose the best link available.
- **Maximum ratio combining (MRC):** MRC uses multiple receive antennas. Again, the receive signal in one antenna may be noisier than the others. The MRC technique is the optimal way to combine the signals from all the antennas in such a way that the combined signal is higher quality than the individual antenna signals.
- **Multi-input, multi-output (MIMO):** MIMO increases the number of channels between the transmitter and receiver. For example a 2×2 (two transmit and two receive) MIMO system has four channels compared to a single-input, single-output (SISO) receive system, therefore it will have theoretically four times the throughput of a SISO system, thereby effectively reducing the transmit time by one-fourth.
- **Rate adaptation:** Another important feature for a robust wireless link is effective rate adaptation. The wireless channel varies with time. A good link may become noisy for some time. In this case, a rate adaptation algorithm will change to a lower throughput but will maintain the link until the channel becomes good again. Without rate adaptation, the wireless connectivity will lose the link and the data.

A low-latency system for video-enabled drones

Texas Instruments' (TI) **DMx digital media processors** have integrated hardware video encode and decode engines coupled with frame-to-memory ISPs designed for low-latency encoding and decoding of videos using multiple slices per frame. These products scale across a broad range of price and performance options (Figure 5 on the following page) to allow designers to best meet a diverse set of drone requirements.

The **TMS320DM36x family** offers a low-cost solution which allows the drone to capture and transmit low-latency video thanks to the integrated frame-to-memory hardware imaging subsystem (ISS) which supports a parallel camera interface

	DM36x DM369 DM368 DM365	DM38x DM388 DM385	DM812x DM8127
Core	ARM® ARM9™ up to 432 MHz	ARM® Cortex™-A8 up to 1000MHz	ARM® Cortex™-A8 up to 1000MHz
Analytics (DSP)	-	-	C674x DSP up to 750MHz
Encode/ Decode Capability	Multi-format: H.264, MPEG4, MPEG2, MJPEG. H.264 BP/MP/HP up to **1080p30**	Multi-format: H.264, MPEG4, MPEG2, MJPEG. H.264 BP/MP/HP up to **1080p60**	Multi-format: H.264, MPEG4, MPEG2, MJPEG. H.264 BP/MP/HP up to **1080p60**
Multimedia	-	-	-
Peripherals	**Integrated HW ISP**, Parallel Camera Input, 3ch Video Output DAC, DDR2, USB2.0, 10/100 EMAC, SD/MMC	**Integrated HW ISP**, Parallel Camera Input, CSI-2, HDMI Output, DDR3/L, USB2.0, PCIe, 10/100/1000 EMAC, SD/MMC	**Integrated HW ISP**, Parallel Camera Input, CSI-2, HDMI Output, DDR3/L, USB2.0, PCIe, 10/100/1000 EMAC, SD/MMC
Applications	Sports and Action Cameras, Wearables, Drones, Automotive Camera, Video Security and Surveillance	Sports and Action Cameras, Wearables, Drones, Automotive Camera, Video Security and Surveillance	Industrial Machine Vision, Vision Analytics, Sports and Action Cameras, Wearables, Drones, Video Security and Surveillance
Package	13x13mm, 0.65mm BGA	16x16mm, 0.8mm BGA	23x23mm, 0.8mm BGA

Figure 5: TI Digital Media family scales across price/performance.

(ISIF). Coupled with the hardware video encode/decode engine (HDVICP) and dedicate display hardware, the processing requirements on the ARM CPU are greatly reduced, lowering power consumption and cost. If a MIPI (CSI2) camera serial interface is required, this may be supported with the **DM38x family** of devices. Figure 6 shows the TMS320DM368 digital media processor in a low-latency video-encoding Wi-Fi system for use in drones.

To allow the drone to capture video, the camera is interfaced to the digital media processor using one of the dedicated camera interfaces. The imaging subsystem (ISS) supports multiple image capture interfaces including a parallel camera interface (CAM), camera serial interface (CSI2) and an image sensor interface (ISIF).

The video feed can be transmitted to the ground control unit using 2.4- or 5-GHz Wi-Fi connectivity and showed on a display unit where the operator makes real-time adjustments to the drone flight path in order to avoid collision.

As an example, the **WiLink 8™ solution** is a family of Wi-Fi and *Bluetooth*® combo-connectivity devices from TI. It is equipped with the advanced features required for drones, such as antenna diversity, MRC, dual-band support (2.4- and 5-GHz bands), rate management and optimized data path. WiLink 8 devices are able to provide and maintain low-latency and robust link, even in very harsh environments.

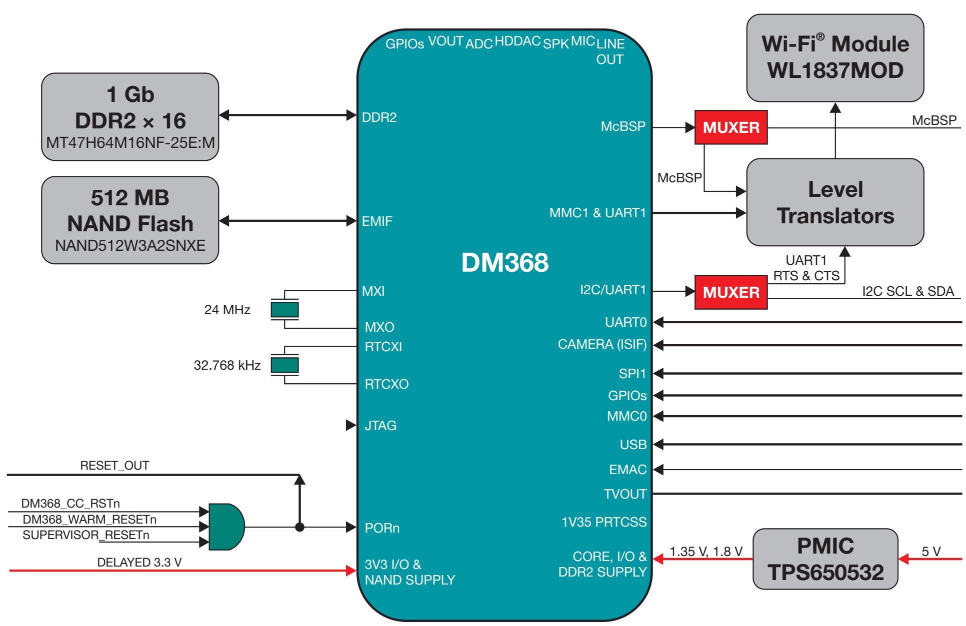

Figure 6: The TMS320DM368 digital media processor in a low-latency video-encoding Wi-Fi system for use in drones.

If local control of the aerial system is desired, using the UART interface exchanges control data with the drone's central control unit. This allows for locally generated collision avoidance control information (based on the video input plus other inputs) to be sent directly to the drone motor controller and allows for autonomous collision avoidance.

For applications that require clock synchronization between multiple wireless devices, such as a swarm of drones communicating over mesh, WiLink 8 can be used to achieve accuracy below 20 µsec.

Integrating the WiLink 8 solution is even easier than before with Linux®, as it is now included in the latest kernel sources.

Summary

Drones promise to enable new applications and fuel economic productivity; however they require modes of operation only possible with low-latency video delivery over a robust wireless connection. Texas Instruments has solutions available today in its' **DMx digital media processor family** to meet these challenges born from a pedigree in video processing spanning multiple decades. Flexible slice-based processing of the full video frame is required and supported across all TI DMx media processors, with examples of ultra-low-latency slice processing implemented in TI's reference design kits (RDK) that stream multiple channels of compressed video. In addition, TI's **WiLink 8** Wi-Fi and Bluetooth combination devices are able to provide and maintain low-latency and robust link, even in very harsh environments.

TI continues to focus on creating the innovative products required to solve the technical challenges of the future drone-based economy, including battery charging and management, embedded processors, sensor technologies, motor drivers, motor control, wireless connectivity and DLP® technology. Please visit **www.TI.com/product/tms320dm368** to learn more about how the TI products featured in this white paper are optimized for your design needs.

References

[1] Marcelo Balve, *Business Insider*, 13 Oct 2014 – COMMERCIAL DRONES: Assessing the Potential for a New Drone-Powered Economy – **www.businessinsider.com/the-market-for-commercial-drones-2014-2**

[2] *PwC* May 2016 – Clarity from above – PwC global report on the commercial applications of drone technology – **www.pwc.pl/clarityfromabove**