

## Autonomous Underwater Surveillance Robot

Shruti Nautiyal<sup>1</sup>, Anshuman Mehta<sup>2</sup>

<sup>1</sup>(UG Student, Department of ECE, Seemant Institute of Technology, Pithoragarh, Uttarakhand, India)

<sup>2</sup>(UG Student, Department of Mechanical Engineering, Sikkim Manipal Institute of Technology, India)

---

**Abstract:** The main objective is to create an interface that allows a person to drive a robot in water and capturing the view through a night camera (360 degrees). Mobile robotic platforms are becoming more and more popular, both in scientific research and in commercial settings. Robotic systems are useful for going places or performing tasks that are not suitable for humans to do. Robots are often able to precisely perform complicated or dangerous tasks with little or no human involvement. However, before a mobile robotic platform is able to be deployed, it must have a way of identifying where it is in relation to objects and obstacles around it. Often, this is being performed by using a visual system, such as a camera. However, just wiring a camera onto a robot is not sufficient. With many of the tasks that are given to a robotic system to perform, a great deal of precision is required to satisfactorily complete these tasks. This precision requires that the robot be given accurate information by the camera.

**Keywords:** AUVs; centre of gravity; buoyancy; thrusters; ARDUINO Board; L298; HT 12E Encoder and 12D Decoder; 434 MHZ TXR & RXR

---

### I. Introduction

#### 1.1 Theory

A robot is a mechanical or virtual artificial agent, usually an electro-mechanical machine that is guided by a computer program or electronic circuitry. Robots can be autonomous or semi-autonomous. The branch of technology that deals with the design, construction, operation, and application of robots, as well as computer systems for their control, sensory feedback, and information processing is Robotics. These technologies deal with automated machines that can take the place of humans in dangerous environments or manufacturing processes, or resemble humans in appearance, behaviour, and/or cognition. The world's oceans are home to the most strange and amazing creatures. What do scientists know about these deep-sea animals and how can they study them easily? Undersea operations are a great application for robotics to replace humans. Working underwater is both dangerous and difficult for humans. One way to learn about these animals in their homes is to use underwater robots. Underwater robots can record data that would be difficult for humans to gather.

#### 1.2 Motivation

The motive of this paper is to develop such a robot which can function under-water as well as on land with live streaming on a remote screen. Nowadays, Mobile robotic platforms are becoming more and more popular, both in scientific research and in commercial settings. Robotic systems are useful for performing tasks that are not suitable for humans or require less human effort. Robots are often able to precisely perform complicated or dangerous tasks.

#### 1.3 Problem Formulation

- For an under-water robot to function properly the main point to be kept in mind is its centre of gravity (C.G.) which helps to keep the robot balanced throughout its functioning period.
- To keep the project afloat in water it requires continuous moment of fans or of thrusters which helps the robot to cut through the water and go downwards to the base.
- For the functioning of the robot on land, it should be able to move in a balanced way on the surface. For this tyres with or without the conveyer belt are used.

#### 1.4 Previous Survey

**MIT's new underwater Robot Odyssey IV:** Chryssostomidis et.al mentioned in their journal that it can hover in a place like a helicopter -- an invaluable tool for deep water oil explorers, marine archaeologists, oceanographers and others. Odyssey IV is the latest in a series of small, inexpensive artificially intelligent submarines developed over the last two decades by the MIT Sea Grant College Program's Autonomous Underwater Vehicles Laboratory. The Odyssey series revolutionized underwater research in the 1990s by introducing the thrifty and highly capable underwater robots. But the previous Odyssey vehicles still had one significant limitation: Like sharks, they could only operate while continuously moving forward.

**Multi-ROV Missions Simulation Framework:** A. Sehgal et.al mentioned in their paper that an underwater communications simulation framework designed for the Unified System for Automation and Robotics Simulator (USARSim). This simulation tool is capable of modeling networked communications between, or with, AUVs by accurately characterizing the underwater acoustic channel. Details on this simulation framework are provided along with some results obtained during development of this tool.

An implementation of ROS on the YELLOWFIN autonomous underwater vehicle (AUV): DeMarco et.al mentioned in their paper that the design, testing, and mission execution of a network of autonomous underwater vehicles (AUV) is a difficult process. The design of low-level controllers requires high-fidelity hydrodynamic models for simulation, but the testing of a large network of AUVs with high-order models is computationally challenging. Also, efficiency is achieved when developers can reuse components already implemented and tested by others in the community. An integrated development system is discussed where the Robot Operating System (ROS) is used to interface a number of individual systems that could not natively communicate. The system integrates the low-level controller simulation, mission planning, and mission execution processes. Most importantly, ROS was integrated with the Mission Oriented Operating Suite (MOOS), which allowed for the use of both ROS and MOOS applications within the same robotic platform via the MOOS/ROS Bridge application. Also, the 3D globe mapping program, NASA World Wind, was interfaced to ROS via ROSJAVA. The target AUV for the ROS implementation was the GTRI YELLOWFIN, which was developed for multiple AUV collaborative missions

### **1.5 Literature Gap**

Hundreds of different ROVs have been designed over the past 50 or so years, but only a few companies sell vehicles in any significant numbers. Vehicles range in size from man portable lightweight ROVs to large diameter vehicles of over 10 meters length. Large vehicle have advantages in terms of endurance and sensor payload capacity; smaller vehicles benefit significantly from lower logistics (for example: support vessel footprint; launch and recovery systems). ROVs have been used for a limited number of tasks dictated by the technology available. With the development of more advanced processing capabilities and high yield power supplies, ROVs are now being used for more and more tasks with roles and missions constantly evolving.

As of 2008, a new class of ROVs are being developed, which mimic designs found in nature. Although most are currently in their experimental stages, these biomimetic (or bionic) vehicles are able to achieve higher degrees of efficiency in propulsion and manoeuvrability by copying successful designs in nature. Two such vehicles are Festo's AquaJelly (ROV) and Evologics' Bionic Manta (ROV).

Today, while most ROVs are capable of unsupervised missions most operators remain within range of acoustic telemetry systems in order to maintain a close watch on their investment. This is not always possible. For example, Canada has recently taken delivery of two ROVs (ISE Explorers) to survey the sea floor underneath the Arctic ice in support of their claim under Article 76 of the United Nations Convention of the Law of the Sea. Also, ultra-low-power, long-range variants such as underwater gliders are becoming capable of operating unattended for weeks or months in littoral and open ocean areas, periodically relaying data by satellite to shore, before returning to be picked up.

## **II. Design And Implementation**

This project can be used to carry out undersea operations. There are various dangerous places where sending a person can be life threatening. This robot can be used for the same. It could be sent to certain deep dark trenches for the research purposes. ROVs are also used extensively by the science community to study the ocean. A number of deep sea animals and plants have been discovered or studied in their natural environment through the use of ROVs: examples include the jellyfish Bumpy and the eel-like halosaurs. This project can be used for the same.

The design of the structure would be optimized with the following features:

- Buoyancy centred along length of the ROV
- Minimal drag for forward travel
- Adequate buoyancy and room for batteries
- Centre of thrust close to the centre of drag for forward and sideways movement
- Waterproof pass-through attached to end-caps of dry tube (rather than through tube itself)
- Accommodations for four thrusters (two vertical/side thrusters and two horizontal thrusters)
- Easy access to thrusters but thrusters protected from collisions and fingers by barrier
- Structurally sound (not fragile)
- Minimal parts

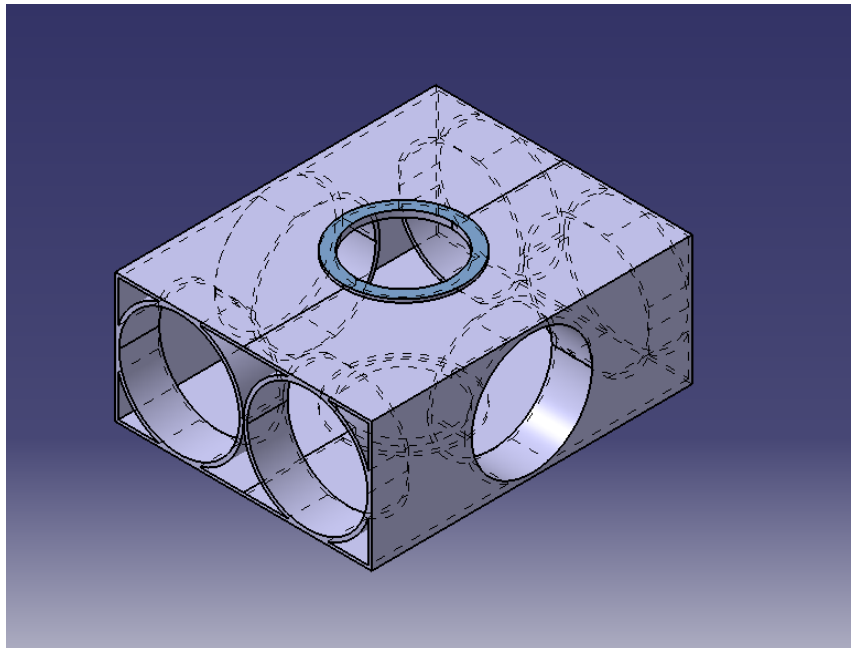


Figure 1: Outer body

### 2.1 Block Diagram

The input in the robot is given through a remote transmitter where it is encoded and transmitted to the receiver which is present in the ROV. The decoder interprets the logic from the signal and passes it to the dual H bridges which help to run the motor.

Simultaneously the wireless camera in the ROV streams the video to its receiver which is connected to a screen. Block diagram is shown below:

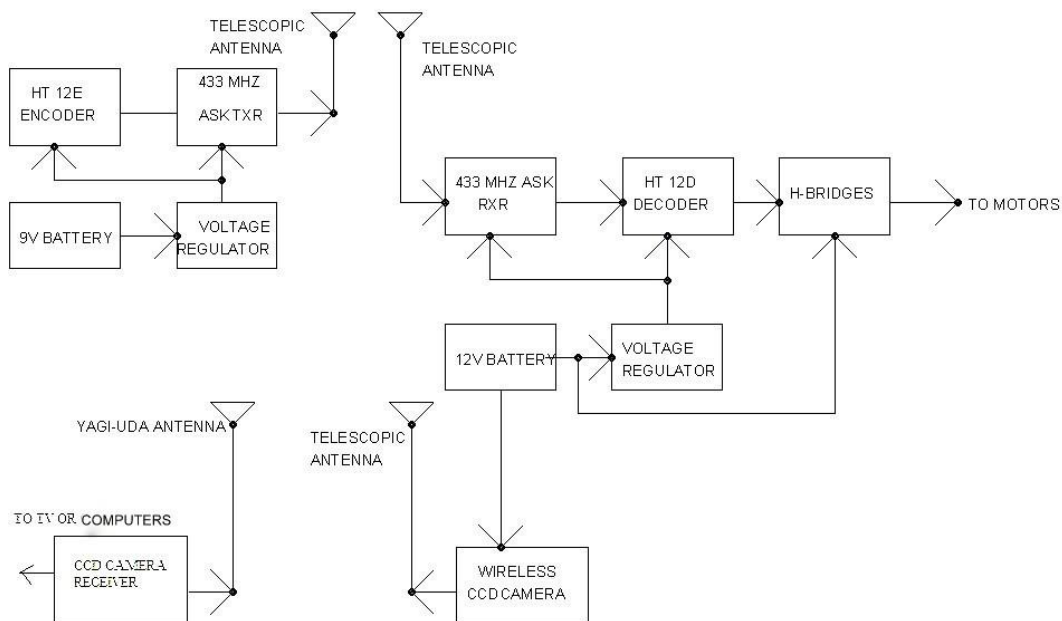


Figure 2: Block Diagram

The voltage regulator is designed to automatically maintain a constant voltage level. It may use an electro-mechanical mechanism for electric-electronic components.

Transmitter and receiver modules are tuned to work correctly at 433.92MHz. When a logic transmitted is received by the receiver, it is passed to the decoder. The decoder interprets the logic and pass it on to the desired H-bridge (L298 IC) which runs the motor connected to it. An additional 12V supply input is provided to the L298 IC so that the logic works at a lower voltage. The wireless camera is powered by a 12v battery and streams live video through its antenna. There is a dedicated receiver for the camera which is connected to the screen for display.

## 2.2 Circuit Diagram

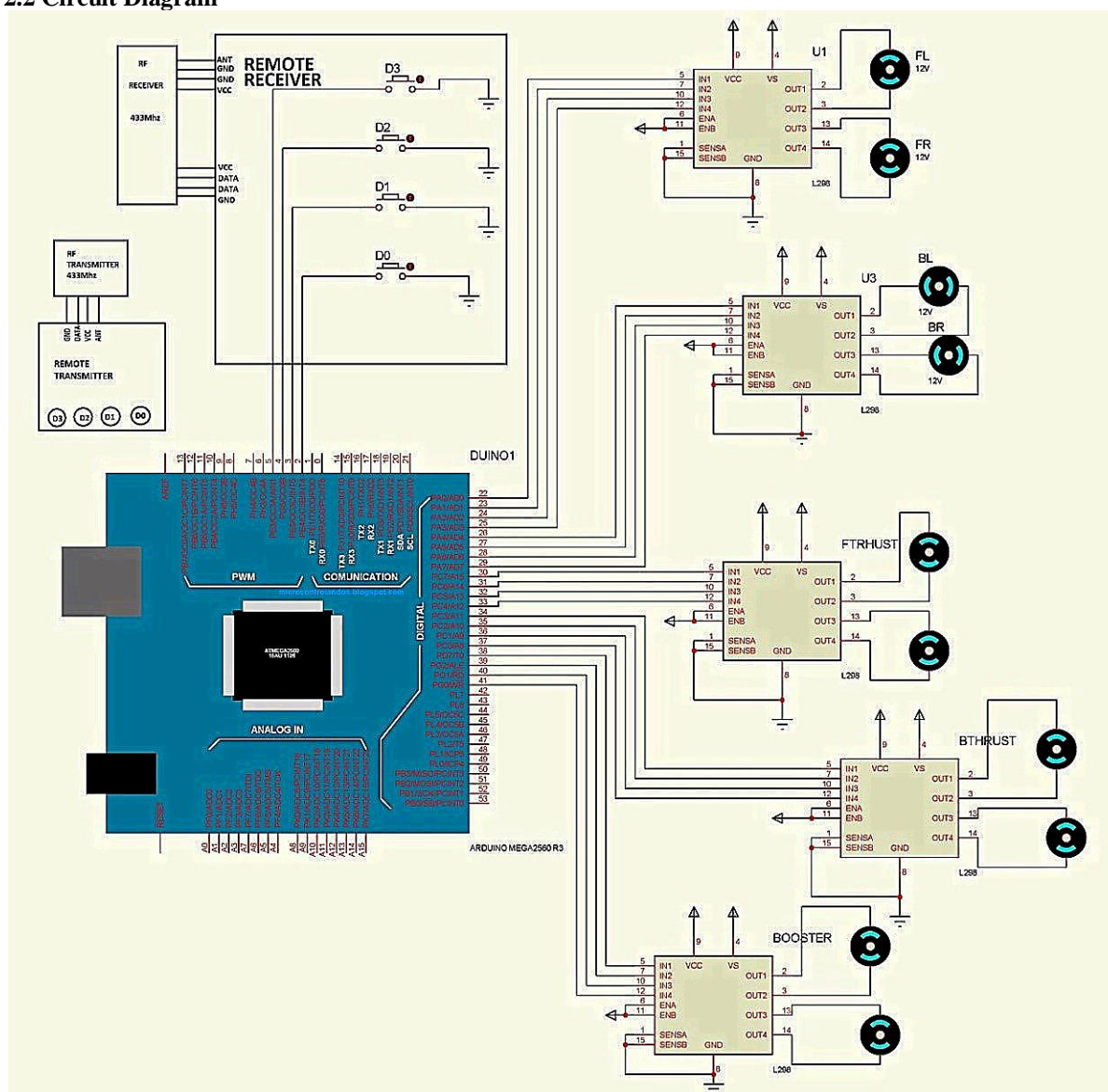


Figure 3: Circuit Diagram

Circuit diagram is shown in Fig. 2. Pins 2,3,4,5 are used as input ports that are connected to wireless receiver. The input is received by ARDUINO and it passes it to the output ports. Pins 22-44 are used as output ports. The output ports are connected to L298 motor driving IC which are further connected to the dc motors. An additional 12V supply input is provided so that the logic works at a lower voltage.

Two enable inputs are provided on L298 IC to enable or disable the motors independently of the input signals. The polarity of the motors depends on the input logic from the remote transmitter. The input logic also decides the L298 IC and motor required for the action. For example if logic provided from the remote transmitter is 1000 (i.e. D3 button is pressed on remote), two L298 ICs are used which are connected to the two motors each. These motors are responsible for the movement of the ROV in forward direction.

## III. Hardware and Components

### 3.1 ARDUINO Board

The ARDUINO Mega 2560 is a microcontroller board based on the ATmega2560. It has 54 digital input/output pins (of which 15 can be used as PWM outputs), 16 analog inputs, 4 UARTs (hardware serial ports), a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with an AC-to-DC adapter or battery to get started.

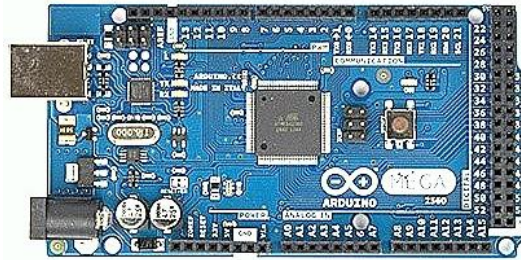


Figure 4. ARDUINO Mega2560 board

The board has following new features:

- **1.0 pin out:** added SDA and SCL pins that are near to the AREF pin and two other new pins placed near to the RESET pin, the IOREF that allow the shields to adapt to the voltage provided from the board. In future, shields will be compatible both with the board that uses the AVR, which operate with 5V and with the ARDUINO due that operate with 3.3V. The second one is a not connected pin that is reserved for future purposes.
- **Stronger RESET circuit.**

Table 1: Summary of Arduino Specifications Microcontroller ATmega2560	
Operating Voltage	5V
Input Voltage (recommended)	7-12V
Input Voltage (limits)	6-20V
Digital I/O Pins	54 (of which 15 provide PWM output)
Analog Input Pins	16
DC Current per I/O Pin	40 mA
DC Current for 3.3V Pin	50 mA
Flash Memory	256 KB of which 8 KB used by boot loader
SRAM	8 KB
EEPROM	4 KB
Clock Speed	16 MHz

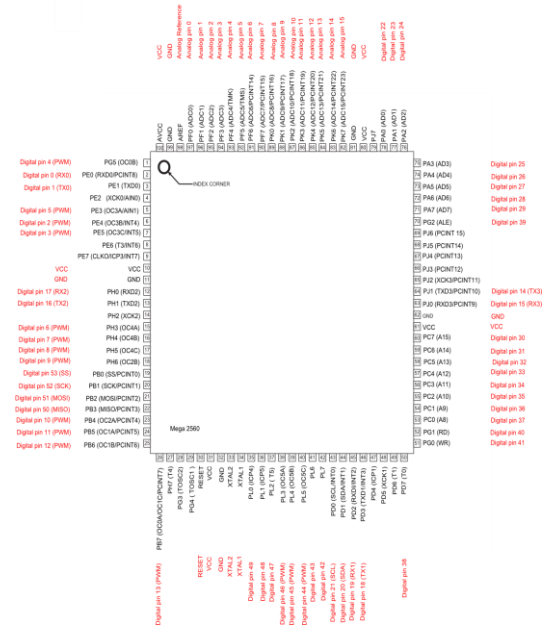


Figure 5: Pin description of Arduino Mega 2560

**Power:** The ARDUINO Mega can be powered via the USB connection or with an external power supply. The power source is selected automatically. External (non-USB) power can come either from an AC-to-DC adapter (wall-wart) or battery. The adapter can be connected by plugging a 2.1mm centre-positive plug into the board's power jack. Leads from a battery can be inserted in the GND and VIN pin headers of the POWER connector. The board can operate on an external supply of 6 to 20 volts. If supplied with less than 7V, however, the 5V pin may supply less than five volts and the board may be unstable. If using more than 12V, the voltage regulator may overheat and damage the board. The recommended range is 7 to 12 volts.

The power pins are as follows:

- **VIN.** The input voltage to the ARDUINO board when it's using an external power source (as opposed to 5 volts from the USB connection or other regulated power source). You can supply voltage through this pin, or, if supplying voltage via the power jack, access it through this pin.
- **5V.** This pin outputs a regulated 5V from the regulator on the board. The board can be supplied with power either from the DC power jack (7 - 12V), the USB connector (5V), or the VIN pin of the board (7-12V). Supplying voltage via the 5V or 3.3V pins bypasses the regulator, and can damage your board, the DC power jack (7 - 12V), the USB connector (5V), or the VIN pin of the board (7-12V). Supplying voltage via the 5V or 3.3V pins bypasses the regulator, and can damage your board. We don't advise it.

- 3V3. A 3.3 volt supply generated by the on-board regulator. Maximum current draw is 50 mA.
- GND. Ground pins.
- IOREF. This pin on the ARDUINO board provides the voltage reference with which the microcontroller operates. A properly configured shield can read the IOREF pin voltage and select the appropriate power source or enable voltage translators on the outputs for working with the 5V or 3.3V.

**Memory:** The ATmega2560 has 256 KB of flash memory for storing code (of which 8 KB is used for the boot loader), 8 KB of SRAM and 4 KB of EEPROM (which can be read and written with the EEPROM library).

**Input and Output:** Each of the 54 digital pins on the Mega can be used as an input or output, using pin Mode (), digital Write, and digital Read functions. They operate at 5 volts. Each pin can provide or receive a maximum of 40 mA and has an internal pull-up resistor (disconnected by default) of 20-50 kOhms. In addition, some pins have specialized functions:

- **Serial:** 0 (RX) and 1 (TX); Serial 1: 19 (RX) and 18 (TX); Serial 2: 17 (RX) and 16 (TX); Serial 3: 15 (RX) and 14 (TX). Used to receive (RX) and transmit (TX) TTL serial data. Pins 0 and 1 are also connected to the corresponding pins of the ATmega16U2 USB-to-TTL Serial chip.
- **External Interrupts:** 2 (interrupt 0), 3 (interrupt 1), 18 (interrupt 5), 19 (interrupt 4), 20 (interrupt 3), and 21 (interrupt 2). These pins can be configured to trigger an interrupt on a low value, a rising or falling edge, or a change in value. See the attach Interrupt () function for details.
- **PWM:** 2 to 13 and 44 to 46. Provide 8-bit PWM output with the analog Write () function.
- **SPI:** 50 (MISO), 51 (MOSI), 52 (SCK), 53 (SS). These pins support SPI communication using the SPI library. The SPI pins are also broken out on the ICSP header, which is physically compatible with the Uno, Duemilanove and Diecimila.
- **LED:** 13. There is a built-in LED connected to digital pin 13. When the pin is HIGH value, the LED is on, when the pin is LOW, it's off.
- **TWI:** 20 (SDA) and 21 (SCL). Support TWI communication using the Wire library. Note that these pins are not in the same location as the TWI pins on the Duemilanove or Diecimila.

The Mega2560 has 16 analog inputs, each of which provides 10 bits of resolution (i.e. 1024 different values). By default they measure from ground to 5 volts, though it is possible to change the upper end of their range using the AREF pin and analog Reference () function.

There are a couple of other pins on the board:

- AREF. Reference voltage for the analog inputs. Used with analog Reference ().
- Reset. Bring this line LOW to reset the microcontroller. Typically used to add a reset button to shields which block the one on the board.

**Programming:** The ARDUINO Mega can be programmed with the ARDUINO software. The ATmega2560 on the ARDUINO Mega comes pre-burned with a boot loader that allows you to upload new code to it without the use of an external hardware programmer. It communicates using the original STK500 protocol (reference, C header files). You can also bypass the boot loader and program the microcontroller through the ICSP (In-Circuit Serial Programming) header using ARDUINO ISP or similar; see these instructions for details.

The ATmega16U2/8U2 is loaded with a DFU boot loader, which can be activated by:

- On Rev1 boards: connecting the solder jumper on the back of the board (near the map of Italy) and then resetting the 8U2.
- On Rev2 or later boards: there is a resistor that pulling the 8U2/16U2 HWB line to ground, making it easier to put into DFU mode. You can then use Atmel's FLIP software (Windows) or the DFU programmer (Mac OS X and Linux) to load a new firmware. Or you can use the ISP header with an external programmer (overwriting the DFU boot loader). See this user-contributed tutorial for more information.

**Automatic (Software) Reset:** Rather than requiring a physical press of the reset button before an upload, the ARDUINO Mega2560 is designed in a way that allows it to be reset by software running on a connected computer. One of the hardware flow control lines (DTR) of the ATmega8U2 is connected to the reset line of the ATmega2560 via a 100 Nano farad capacitor. When this line is asserted (taken low), the reset line drops long enough to reset the chip. The ARDUINO software uses this capability to allow you to upload code by simply pressing the upload button in the ARDUINO environment. This means that the boot loader can have a shorter timeout, as the lowering of DTR can be well-coordinated with the start of the upload.

This setup has other implications. When the Mega2560 is connected to either a computer running Mac OS X or Linux, it resets each time a connection is made to it from software (via USB). For the following half-



second or so, the boot loader is running on the Mega2560. While it is programmed to ignore malformed data (i.e. anything besides an upload of new code), it will intercept the first few bytes of data sent to the board after a connection is opened. If a sketch running on the board receives one-time configuration or other data when it first starts, make sure that the software with which it communicates waits a second after opening the connection and before sending this data.

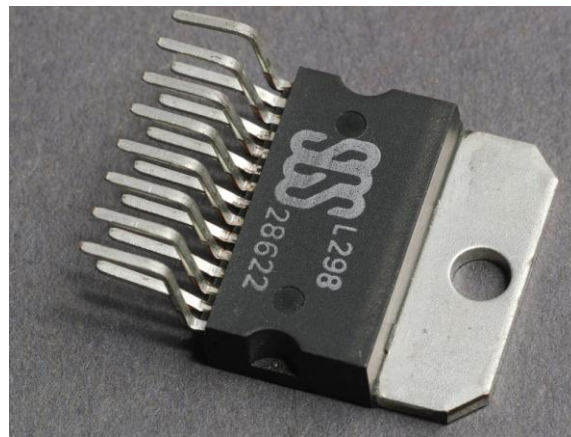
The Mega2560 contains a trace that can be cut to disable the auto-reset. The pads on either side of the trace can be soldered together to re-enable it. It's labelled "RESET-EN". You may also be able to disable the auto-reset by connecting a 110 ohm resistor from 5V to the reset line; see this forum thread for details.

**USB Over current Protection:** The ARDUINO Mega2560 has a resettable poly-fuse that protects your computer's USB ports from shorts and over current. Although most computers provide their own internal protection, the fuse provides an extra layer of protection. If more than 500 mA is applied to the USB port, the fuse will automatically break the connection until the short or overload is removed.

**Physical Characteristics and Shield Compatibility:** The maximum length and width of the Mega2560 PCB are 4 and 2.1 inches respectively, with the USB connector and power jack extending beyond the former dimension. Three screw holes allow the board to be attached to a surface or case. Note that the distance between digital pins 7 and 8 is 160 mil (0.16"), not an even multiple of the 100 mil spacing of the other pins.

### **3.2 L298 (Motor Driver Integrated Circuit)**

The L298 is an integrated monolithic circuit in a 15-lead multi-watt and PowerSO20 packages. It is a high voltage, high current dual full-bridge driver designed to accept standard TTL logic levels and drive inductive loads such as relays, solenoids, DC and stepping motors. Two enable inputs are provided to enable or disable the device independently of the input signals.



**Figure 6:** L298 Chip

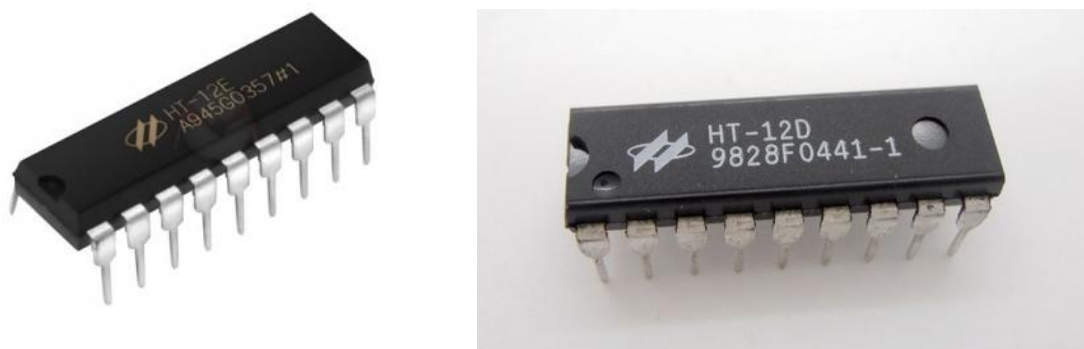
The emitters of the lower transistors of each bridge are connected together and the corresponding external terminal can be used for the connection of an external sensing resistor. An additional supply input is provided so that the logic works at a lower voltage.

**Key Features**

- Operating supply voltage up to 46V
- Total dc current up to 4A
- Low saturation voltage
- Over temperature protection
- Logical "0" input voltage up to 1.5 V (high noise immunity)

### **3.3 HT 12E Encoder and 12D Decoder:**

HT12E is an encoder integrated circuit of 212 series of encoders. They are paired with 212 series of decoders for use in remote control system applications. It is mainly used in interfacing RF and infrared circuits. The chosen pair of encoder/decoder should have same number of addresses and data format. Simply put, HT12E converts the parallel inputs into serial output. It encodes the 12 bit parallel data into serial for transmission through an RF transmitter. These 12 bits are divided into 8 address bits and 4 data bits. HT12E has a transmission enable pin which is active low. When a trigger signal is received on TE pin, the programmed addresses/data are transmitted together with the header bits via an RF or an infrared transmission medium. HT12E begins a 4-word transmission cycle upon receipt of a transmission enable. This cycle is repeated as long as TE is kept low. As soon as TE returns to high, the encoder output completes its final cycle and then stops.



**Figure 7:** Encoder IC and Decoder IC

HT12D is a decoder integrated circuit that belongs to 212 series of decoders. This series of decoders are mainly used for remote control system applications, like burglar alarm, car door controller, security system etc. It is mainly provided to interface RF and infrared circuits. They are paired with 212 series of encoders. The chosen pair of encoder/decoder should have same number of addresses and data format. In simple terms, HT12D converts the serial input into parallel outputs. It decodes the serial addresses and data received by, say, an RF receiver, into parallel data and sends them to output data pins. The serial input data is compared with the local addresses three times continuously. The input data code is decoded when no error or unmatched codes are found. A valid transmission is indicated by a high signal at VT pin. HT12D is capable of decoding 12 bits, of which 8 are address bits and 4 are data bits. The data on 4 bit latch type output pins remain unchanged until new is received.

### 3.4 434 MHZ TXR & RXR:

Transmitter and receiver modules are tuned to work correctly at 433.92MHz. Transmitter can be powered from 3 to 12V power supply while receiver accepts 5V. 5V is common for AVR microcontrollers so no problems with interfacing.



**Figure 9:** Transmitter and Receiver IC

Modules don't require addition components. For better distances 30 – 35cm antennas needs to be applied. Modules use Amplitude-Shift Keying (ASK) modulation method and uses 1MHz bandwidth.

<b>Operating Voltage</b>	(DC) 3-12V
<b>Operating current</b>	5-45mA
<b>Standby current</b>	≤0.02uA
<b>Operating frequency</b>	433MHZ
<b>Mailing address</b>	Single shot
<b>Modulation rate</b>	3KHZ
<b>Modulation system</b>	ASK (amplitude modulation)



<b>Operating temperature</b>	-10°C - +60°C
<b>Dimensions</b>	19×19×8mm
<b>Input signal</b>	TTL level
<b>Radiation power</b>	20mW@5V
<b>Transmitting distance</b>	80 Meter
<b>Features</b>	Ultra small volume, power, normal fire head

**Table 3: Summary of ASK RXR Specifications**

<b>Modulation</b>	AM/ASK
<b>Sensitivity</b>	103dBm
<b>Data rate</b>	4.8K
<b>Nominal voltage</b>	DC (5V±0.5V)
<b>Operating current</b>	4mA@5V
<b>Operating frequency</b>	433MHZ
<b>Output data level</b>	TTL level
<b>Interface</b>	4PIN, (2.54mm spacing)
<b>Maximum rate</b>	1KHZ
<b>Receiver sensitivity</b>	105dBm
<b>Dimensions</b>	30×12×8mm
<b>Operating temperature</b>	-10°C to +60°C

#### IV. Results And Conclusions

The project has been developed for the surveillance purpose. It can be used by the researchers to carry out underwater operations. It can be sent to the places where human life cannot go. The extreme conditions, that causes a threat to human life, can be researched by using this robot. Applications are listed as: Fault analysis in sophisticated underwater systems

- Capturing High Resolution images of underwater flora and fauna.
- Temperature Measurement.
- Diving into unknown water body (“Trench”) explore human accessibility
- Detecting leakages in ships.
- Used for SPYING purposes

#### V. Scope of Future Work

In the future, it would be beneficial to have higher resolution cameras to allow the pilot to see the things clearly. Having HD 3D, stereoscopic cameras would be especially beneficial as it would allow for clearer view into the water while giving the pilot a better feel for how far away he/she is from items during the mission. Having digital cameras could also allow for computer assisted piloting that recognizes objects during the mission. This could then be used to help stabilize the ROV and centre on an object or to autonomously do tasks requiring a lot of precision. This would help alleviate stress the pilot may feel and allow for tasks to be completed faster.

#### References

- [1]. C Harrold, K Light and S Lisin. (1993). "Distribution, Abundance, and Utilization of Drift Macrophytes in a Nearshore Submarine Canyon System." In: Heine and Crane (eds). Diving for Science...1993. Proceedings of the American Academy of Underwater Sciences (13th annual Scientific Diving Symposium). Retrieved 2008-07-11. Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.
- [2]. Reed JK, Koenig CC, Shepard AN, and Gilmore Jr. RG (2007). "Long Term Monitoring of a Deep-water Coral Reef: Effects of Bottom Trawling." In: NW Pollock, JM Godfrey (Eds.) the Diving for Science...2007. Proceedings of the American Academy of Underwater Sciences (Twenty-sixth annual Scientific Diving Symposium). Retrieved 2008-07-11. Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
- [3]. Buhr, Sarah (November 23, 2013). "OpenROV wants to get to bottom of underwater mystery". USA Today: Tech. Retrieved November 24, 2013.
- [4]. McCracken, Dave. "Talking about putting yourself way out there beyond where you should not be!". The New 49'ers Club: Gold Mining Adventures. Retrieved November 24, 2013.
- [5]. Autonomous Vehicles at the Institute of Marine Technology Problems.
- [6]. Indian Underwater Robotics Society's Multiple AUVs.
- [7]. "AquaJelly" Festo Corporate, 2008.
- [8]. Dr. Steven W. Moore, Harry Bohm, and Vickie Jensen, 'UNDERWATER ROBOTICS: SCIENCE, DESIGN & FABRICATION, ISBN 978-0-9841737-0-9, Edition-V, 2013
- [9]. Sehgal, D. Cernea (June 2010). "A Multi-AUV Missions Simulation Framework for the USARSim Robotics Simulator". 18th IEEE Mediterranean Conference on Control & Automation. Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.