BAUDOT · ITA2

5-Bit Teleprinter Character Code

Historical Overview & Character Reference

Émile Baudot, 1870 · CCITT, 1930 · ISO/IEC 6937 legacy

> *"Before ASCII, before Unicode, before any computer existed in recognizable form, a French telegraph engineer named Émile Baudot worked out a way to represent every letter of the alphabet in just five bits. That code — refined, standardized, and renamed more than once over the following century — became the common substrate beneath every piece of telex equipment in this guide series. Vernam's cipher XOR'd it. Telex exchanges carried it. Punched key tape encoded it. To read any of it, you read Baudot."*

Part I — Historical Background

1.1 Émile Baudot and the Five-Bit Idea

In 1870, a French telegraph engineer named Jean-Maurice-Émile Baudot patented a new kind of telegraph system. The breakthrough was not the hardware — it was the code. Morse code, then dominant, used variable-length sequences of dots and dashes, which meant each character took a different amount of time to transmit and the operator had to count pulses to decode. Baudot's insight was to use a fixed five-bit pattern for every character — thirty-two possible combinations, each the same length, each decodable without counting. The machine could read the five bits in parallel, match them against a lookup wheel, and strike the corresponding letter automatically. No trained ear required.

Five bits gave 32 combinations, which is enough for 26 letters but not for both letters and digits in the same table. Baudot solved that with a shift mechanism: two special patterns switched the receiver between "letters mode" and "figures mode," doubling the effective character set while keeping the physical encoding at five bits. Every character pattern thus carried two meanings — one in letters shift, one in figures shift — and the receiver remembered which mode it was in until a new shift arrived. That dual-interpretation scheme survived essentially unchanged for the next 130 years.

1.2 Standardization: CCITT-1 and CCITT-2

Baudot's original code was adopted by the French telegraph administration and spread from there to other European networks, but with minor local variations. International teleprinter traffic demanded a single standard. The International Telegraph and Telephone Consultative Committee — the CCITT, ancestor of today's ITU-T — took on the task.

In 1926 the CCITT ratified the first formal international version: International Telegraph Alphabet No. 1, or ITA1 / CCITT-1. Four years later, an improved revision, ITA2 / CCITT-2, replaced it as the global standard. ITA2 retained Baudot's five-bit frame and the letters/figures shift scheme, but reorganized the character assignments to make the most-frequent letters use the most-reliable bit patterns — a kind of 1930s Huffman coding optimized for the electromechanical teleprinters of the day.

ITA2 became the workhorse code of twentieth-century non-military telex. News agencies moved their wire copy in it. Banks cleared their settlement instructions in it. Diplomatic missions filed their cables in it. The TW39 and ED1000 Telex networks described elsewhere in this guide series carried ITA2-encoded characters almost exclusively.

1.3 The American Variant: USTTY

The United States adopted a closely related but not identical code for domestic teleprinter use, known variously as USTTY, US-TTY, or simply “Teletype code.” Western Union, the Bell System, and later AT&T TWX all used it. The letter assignments are identical to ITA2 — so an A is an A and a Z is a Z on either side of the Atlantic — but the figures-shift assignments diverge on several characters. In USTTY, figures-shift S is the bell character (Klingel in the German chart), figures-shift ' is apostrophe, and the symbols $, !, &, #, " appear in positions where ITA2 reserves other punctuation. This difference is the reason European teleprinter messages sometimes printed garbled punctuation on American receivers and vice versa, and it is the reason telex gateways between continents often included a code-translation step.

> **Historical Note:** *The Baudot name stuck in everyday usage long after the original 1870 code was superseded, to the point where “Baudot code” and “ITA2” are often used interchangeably in casual writing even though they refer to different tables. The unit of signaling speed — the baud — is also named after Baudot and retains its independent meaning: one signaling event per second, regardless of how many bits that event carries. A 50-baud teleprinter, the standard speed of mid-century telex, transmitted 50 signaling events per second, which corresponds to about 6.67 characters per second or 400 characters per minute.*

Part II — How the Code Works

2.1 The Frame

Every character transmitted by a 5-bit teleprinter travels inside a fixed frame. The frame has three parts. First comes a single start bit, always in the space (zero) state, which tells the receiver that a character is about to arrive and gives it a timing reference. Second, the five data bits that actually encode the character. Third, a stop bit in the mark (one) state, or in some standards one and a half stop bits, which gives the mechanical receiver time to finish printing before the next character begins. The total frame length is therefore 7.5 bits at the standard ITA2 timing: one start bit plus five data bits plus one and a half stop bits.

At 50 baud, each bit takes 20 milliseconds to transmit. A full character frame takes about 150 milliseconds — the rhythmic chatter that anyone who has stood beside a working teleprinter remembers viscerally. The frame structure is why teleprinter transmission is called “asynchronous”: there is no continuous clock signal between sender and receiver. Each start bit resynchronizes the receiver for the duration of the following character, and the stop bit lets it return to an idle state between characters.

2.2 Reading a Punched Tape

On a paper tape — the physical medium that carries Baudot characters across most of this guide series — each character occupies one row across the tape's width. The five data bits run in parallel tracks along the tape's length. A punched hole is a mark (binary 1); the absence of a hole is a space (binary 0). A narrower track of smaller sprocket holes runs down the tape's center for mechanical transport, but carries no data.

A 5-level tape has five data tracks, though the physical tape often shows eight hole positions: five for data, one for the sprocket, and two reserved for formats like ITA5/ASCII or for specialized uses. The ability to distinguish a Baudot tape from an ASCII tape by eye is largely a matter of counting active tracks: if you see holes across only five positions with an obvious central sprocket, you are looking at Baudot.

2.3 The Letters/Figures Shift

The five-bit code by itself encodes only 32 symbols, far fewer than a complete English character set with letters, digits, and common punctuation. The solution built into ITA2 — and inherited from Baudot's original — is modal operation. Two reserved control codes, sometimes written LTRS and FIGS, sometimes Bu (Buchstaben) and Zi (Ziffern) in German usage, switch the receiver between two interpretation tables.

In letters shift, the five-bit code 00011 means A. In figures shift, the same five-bit code means a dash. The sender controls the mode; the receiver follows. A well-formed message typically begins with an explicit LTRS code to put both ends in a known state, switches to FIGS only when a run of digits or punctuation is needed, and returns to LTRS immediately afterward. If a receiver loses sync on the shift state — for example through a garbled control character — the next several characters can print as apparent gibberish until the next shift restores the mode.

2.4 Control Characters

CODE	NAME	MEANING
LTRS / Bu	Letters shift	Switch to letters mode (A-Z table)
FIGS / Zi	Figures shift	Switch to figures mode (digits / punctuation table)
SP / ZwR	Space	Space between words
CR / WR	Carriage return	Return print head to left margin
LF / ZL	Line feed	Advance paper by one line
BEL / Klingel	Bell	Audible alert at receiver (figures-shift only)
WRU / WerDa	Who are you	Query remote station's answerback
NULL	Null	No-op / padding / idle

The WRU control character — “who are you,” German WerDa — is particularly significant. When received, it causes the remote teleprinter's Kennungsgeber (answerback drum) to automatically transmit that station's identifier back to the sender. This is the same mechanism described in the ED1000 guide, where the exchange queries the called subscriber's Kennungsgeber during call setup to confirm that the connection reached the intended station. The WRU code is thus not merely a convenience but a security primitive built into the character set itself.

Part III — The Three Code Tables

3.1 Comparison Chart

The table below compares ITA1, ITA2, and USTTY side by side. The five-bit pattern is shown as a sequence of 1s and 0s, read left-to-right as bits 1 through 5. The letters-shift column shows what the receiver prints when in letters mode; the figures-shift column shows what it prints when in figures mode. Shaded rows indicate control codes that have no printable equivalent.

BITS	LETTER	ITA1 fig	ITA2 fig	USTTY fig
00011	A	-	-	-
11001	B	?	?	?
01110	C	:	:	:
01001	D	WerDa	WerDa	$
00001	E	3	3	3
01101	F	—	—	!

BITS	LETTER	ITA1 fig	ITA2 fig	USTTY fig
11010	G	—	—	&
10100	H	—	—	#
00110	I	8	8	8
01011	J	Klingel	Klingel	'
01111	K	(	(	(
10010	L	)	)	)
11100	M	.	.	.
01100	N	,	,	,
11000	O	9	9	9
10110	P	0	0	0
10111	Q	1	1	1
01010	R	4	4	4
00101	S	'	'	Klingel
10000	T	5	5	5
00111	U	7	7	7
11110	V	=	=	;
10011	W	2	2	2
11101	X	/	/	/
10101	Y	6	6	6
10001	Z	+	+	"
00010	CR	WR	WR	WR
01000	LF	ZL	ZL	ZL
00100	SP	ZwR	ZwR	ZwR
11011	FIGS	1…	1…	1…
11111	LTRS	A…	A…	A…

Reading Note: *The "Klingel" (bell) sits at J in ITA1 and ITA2 but migrates to S in USTTY. The $, !, &, #, and " symbols exist only in USTTY. ITA2 keeps F, G, and H as reserved positions with no printable character assigned — an unused region of the table that some national variants filled with additional accented letters or local currency symbols.*

Part IV — Where Baudot Fits

4.1 Baudot as Cipher Substrate

The Vernam cipher, described in the first guide of this series, operates on Baudot characters directly. Vernam's 1917 insight was that if you XOR a plaintext Baudot character bit-by-bit with a random key character of the same length, the result is another valid five-bit pattern — and if that key is never reused, the result is mathematically unbreakable. The machine Vernam built did exactly this: two synchronized paper tapes fed through a pair of readers, the outputs XOR'd in real time, and the result punched onto a third tape for transmission. The cipher's elegance lies in its reuse of the existing teleprinter character frame. Ciphertext rides the wire in exactly the same 7.5-bit frame as plaintext, at the same 50 baud, through the same Telex exchanges, on the same Fernschaltgeräte. Only the meaning of the bits changes.

4.2 Baudot as Telex Payload

Every character transmitted across the TW39 and ED1000 Telex networks described elsewhere in this series was a Baudot character. The TW39 Fernschaltgerät cared only about whether the loop current was in mark or space at any instant — it knew nothing about character boundaries or alphabet tables. The ED1000 interface cared only about whether the FSK tone was at 2250 or 3150 Hz. Above both layers, though, the bits assembled into ITA2 character frames, and the receiving teleprinter's decoder mapped each five-bit pattern to the corresponding printable character or control code. ITA2 is what operators at both ends actually read and typed.

4.3 Baudot as Key Material

The punched key tape described in the KOI-18 guide of this series did not, strictly speaking, encode Baudot characters — a cryptographic key has no semantic meaning as text. But the physical format of the tape is a direct descendant of Baudot teleprinter tape. The hole positions, the sprocket, the paper-mylar-paper laminate construction, even the U.S. Teletype Corporation tape width — all trace back to the mechanical standards that Baudot and his successors established for the teleprinter industry. When an operator pulled a punched key tape through a KOI-18 reader, they were using equipment whose dimensional standards had been fixed before the First World War.

4.4 Afterlife

ASCII (1963) and later Unicode displaced Baudot and ITA2 for general computing use during the 1960s and 1970s. But ITA2 survived operationally in three important niches well beyond that transition. Maritime radio continued to use ITA2 for narrow-band direct-printing telegraphy — NBDP, SITOR, the Global Maritime Distress and Safety System — into the twenty-first century. Amateur radio teleprinter operation, known as RTTY, still uses ITA2 today and remains a popular

mode among shortwave enthusiasts. And the i-Telex enthusiast network described in the TW39 and ED1000 guides transmits ITA2 across TCP/IP, preserving the original 5-bit encoding end-to-end so that restored mechanical teleprinters can still exchange characters in their native alphabet.

4.5 Summary

Baudot / ITA2 is a 5-bit fixed-length character code, invented in 1870 and internationally standardized in 1930, that served as the primary character encoding for teleprinter communication across the twentieth century. Its letters/figures shift scheme extended its effective character set to roughly 60 printable symbols, enough to carry most business, diplomatic, and news correspondence of the era. It provided the character substrate on which the Vernam cipher, the German Telex network, and the mechanical teleprinter industry as a whole were built. Any chart comparing ITA1, ITA2, and USTTY side by side — like the one reproduced at the start of this guide — is therefore a key to reading the fabric of twentieth-century text-over-wire communication, plaintext and ciphertext alike.