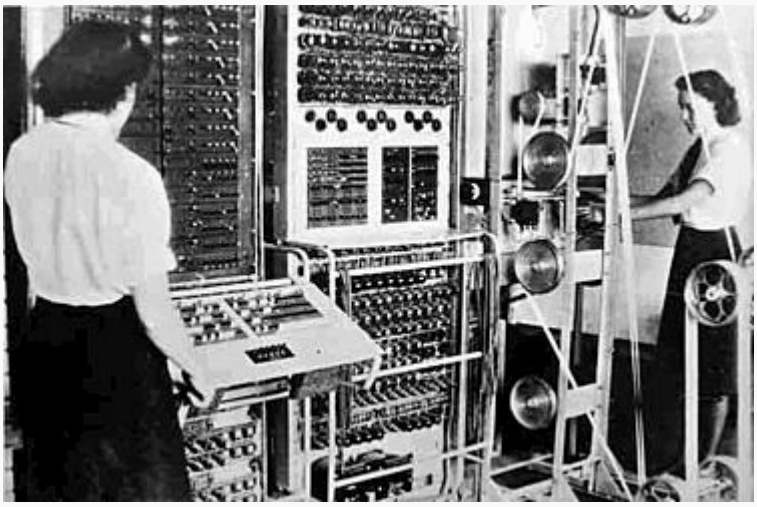
Colossus computer

From Wikipedia, the free encyclopedia

Not to be confused with the fictional computer of the same name in the movie Colossus: The Forbin Project.

Colossus computer



A Colossus Mark 2 computer being operated by Dorothy Du Boisson(left) and Elsie Booker. The slanted control panel on the left was used to set the "pin" (or "cam") patterns of the Lorenz. The "bedstead" paper tape transport is on the right.

Developer	Tommy Flowers assisted by Sidney Broadhurst, William Chandler and for the Mark 2 machines, Allen Coombs
Manufacturer	Post Office Research Station
Type	Special-purpose electronic digital programmable computer
Generation	First-generation computer
Release date	Mk 1: December 1943; Mk 2: 1 June 1944
Discontinued	8 June 1945

Units shipped	11
Media	Electric typewriter output Programmed, using switches and plug panels
CPU	Custom circuits using valves and Thyratrons. A total of 1600 in Mk 1 and 2400 in Mk 2. Also relays andstepping switches
Memory	None (no RAM)
Display	Indicator lamp panel
Input	Paper tape of up to 20 000 × 5-bit characters in a continuous loop
Power	7.5 kW

Colossus was the name of a series of computers developed for British codebreakers in 1943-1945 to help in the cryptanalysis of the Lorenz cipher. Colossus used thermionic valves (vacuum tubes) and thyratrons to perform Boolean and counting operations. Colossus is thus regarded[1] as the world's first programmable, electronic, digital computer, although it was programmed by plugs and switches and not by a stored program.

Colossus was designed by the engineer Tommy Flowers to solve a problem posed by mathematician Max Newman at the Government Code and Cypher School (GC&CS) at Bletchley Park. Alan Turing's use of probability in cryptanalysis[2] contributed to its design. It has sometimes been erroneously stated that Turing designed Colossus to aid the cryptanalysis of the Enigma.[3] Turing's machine that helped decode Enigma was the electromechanical Bombe, not Colossus.[4]

The prototype, **Colossus Mark 1**, was shown to be working in December 1943 and was operational at Bletchley Park on 5 February 1944.[5] An improved **Colossus Mark 2** that used shift registers to quintuple the processing speed, first worked on 1 June 1944, just in time for the Normandy Landings on D-Day.[6] Ten Colossi were in use by the end of the war and an eleventh was being commissioned.[6] Bletchley Park's use of these machines allowed the Allies to obtain a vast amount of high-level military intelligence from teleprinter messages between the German High Command (OKW) and their army commands throughout occupied Europe.

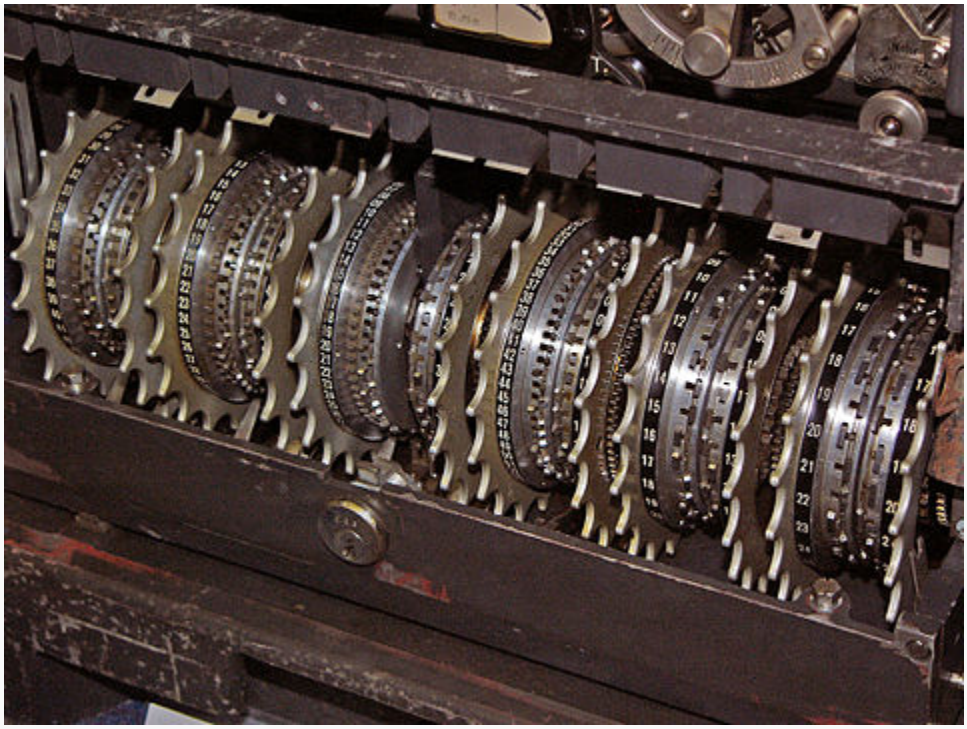
The destruction of almost all of the Colossus hardware and blueprints, as part of the effort to maintain a project secrecy that was kept up into the 1970s, deprived most of those involved with Colossus of credit for their pioneering advancements in electronic digital computing during their lifetimes. A functioning replica of a Colossus computer was completed in 2007 and is on display at The National Museum of Computing at Bletchley Park.[7]

Contents

[hide]

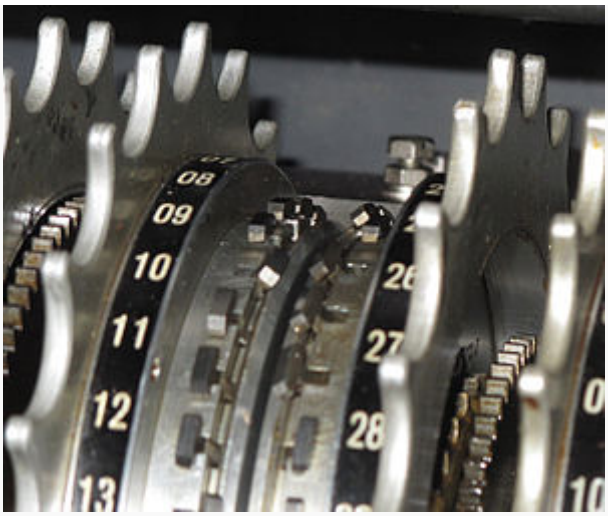
Purpose and origins[edit]

See also: Cryptanalysis of the Lorenz cipher



The Lorenz SZ machines had 12 wheels, each with a different number of cams (or "pins").

Wheel number	1	2	3	4	5	6	7	8	9	10	11	12
BP wheel name[8]	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	μ_{37}	μ_{61}	χ_1	χ_2	χ_3	χ_4	χ_5
Number of cams (pins)	43	47	51	53	59	37	61	41	31	29	26	23



Cams on wheels 9 and 10 showing their raised (active) and lowered (inactive) positions.

The Colossus computers were used to help decrypt radio teleprinter messages that had been encrypted using the electromechanical Lorenz SZ(*Schluesselzusatzgeraet*) in-line cipher machine.[9] In enciphering a message, the Lorenz machine combined the 5-bit plaintext characters with a stream of key ciphertext characters using the XOR Boolean function. This is a Vernam cipher and the deciphering process involved an identically setup Lorenz SZ machine generating the same key sequence and XOR-ing it with the received ciphertext to reproduce the plaintext. The keystream was generated using twelve pinwheels.

British codebreakers called encrypted German teleprinter traffic "Fish",[9] and called the SZ40/42 machine and the intercepted messages "Tunny". Colossus was used for finding possible Lorenz key settings – not completely decrypting the message. It compared two character streams, counting a statistic based on a programmable Boolean function. The ciphertext was read at high speed from a paper tape. The other stream was generated internally, and was an electronic simulation of the Lorenz machine. If the count for a setting was above a certain threshold, it would be sent as output to an electric typewriter.

The logical structure of the Lorenz machine was diagnosed at Bletchley Park without a machine being seen – something that did not happen until almost the end of the war.[10] First, John Tiltman, a very talented GC&CS cryptanalyst derived a key stream of almost 4000 characters from a German operating blunder in August 1941. Then Bill Tutte, a newly arrived member of the Research Section used this key stream to work out the logical structure of the Lorenz machine. He correctly deduced that it had twelve wheels in two groups of five, which he named the χ (*chi*) and ψ (*psi*) wheels, and the remaining two the μ *mu* or "motor" wheels. The *chi* wheels stepped regularly with each letter that was encrypted, while the *psi* wheels stepped irregularly, under the control of the motor wheels.[11]

In order to decrypt the ciphertext of the transmitted messages, there were two tasks that had to be performed. The first was "wheel breaking", which was the discovery of the cam patterns for all the wheels. These patterns were set up once on the Lorenz machine and then used for a fixed period of time and for a number of different messages. The second task was "wheel setting", which could be attempted once the cam patterns were known.[12] Each message encrypted using Lorenz was enciphered at a different start position for the wheels, and it was this start position of the *chi* wheels that Colossus was initially designed to discover.

The XOR function used in the Vernam cipher for both enciphering and deciphering could also be used to upset the cipher's obscuring of the characteristics of the plaintext in the ciphertext. This was discovered by Alan Turing in July 1942 when he was on loan from the German Naval Enigma section to the Research Section at Bletchley Park. He was studying Tunny and invented a method of wheel-breaking that became known as Turingery.[18] With a truly random key, the Vernam cipher removes the natural language property of a plaintext message of having an uneven frequency distribution of the different characters, to produce a uniform distribution in the ciphertext. Turing worked out that examining the character-to-character changes of character streams, instead of the frequency distribution of the characters in the ciphertext, showed a departure from uniformity which provided a way into the system. Providing the character-to-character changes was achieved by "differencing" in which each bit or character was XOR-ed with its successor.[15]

Notation [13]	
P	plaintext
K	key – the sequence of characters XOR'ed (added) to the plaintext to give the ciphertext
χ	*chi* component of key
ψ	*psi* component of key
ψ'	extended *psi* – the actual sequence of characters added by the *psi* wheels, including those when they do not advance [14]
Z	ciphertext
D	de-*chi*—the ciphertext with the *chi* component of the key removed[13]
Δ	any of the above XOR'ed with its successor character or bit[15]
$\oplus$	the XOR operation[16][17]

By using differencing and knowing that the *psi* wheels did not advance with each character, Tutte worked out that trying just two differenced bits (impulses) of the *chi*-stream against the differenced ciphertext would produce a statistic that was non-random. This became known as Tutte's "1+2 break in".[19] The process of wheel setting found the start position of the key wheels in relation to the start of the message. Initially Colossus was used only to work out the start positions of the *chi* wheels, but later, methods were devised for the other wheels. Later still an additional electronic unit was designed for wheel breaking, which was added to some Mark 2 Colossi.

The manual processes in decrypting messages were undertaken in a section at Bletchley Park led by Major Ralph Tester which was known as the "Testery". Colossus was developed for the "Newmanry",[20]the section headed by the mathematician Max Newman that was responsible for machine methods against the Lorenz machine. The Colossus design arose out of a prior project that produced a counting machine dubbed "Heath Robinson". The main problems with Heath Robinson were the relative slowness of electro-mechanical parts and the difficulty of synchronising two paper tapes, one punched with the enciphered message, the other representing the patterns produced by the wheels of the Lorenz machine.[21] The tapes tended to stretch when being read, at some 2000 characters per second, resulting in unreliable counts.

Design and construction[edit]

In 1994, a team led by Tony Sale (right) began a reconstruction of a Colossus at Bletchley Park. Here, in 2006, Sale supervises the breaking of an enciphered message with the completed machine.

Tommy Flowers was a senior electrical engineer at the Post Office Research Station at Dollis Hill who had been appointed MBE in June 1943. Prior to his work on Colossus, he had been involved with GC&CS at Bletchley Park from February 1941 in an attempt to improve the Bombes that were used in theCryptanalysis of the German Enigma cipher machine.[22] He was recommended to Max Newman by Alan Turing who had been impressed by his work on the Bombes.[23] The main components of Colossus's predecessor, Heath Robinson were as follows.

- A tape transport and reading mechanism that ran the looped key and message tapes at between 1000 and 2000 characters per second.
- A combining unit that implemented the logic of Tutte's method.
- A counting unit that had been designed by Dr C.E. Wynn-Williams of the Telecommunications Research Establishment (TRE) at Malvern which counted the number of times the logical function returned a specified truth value.

Stepping switch from an original Colossus presented by the Director of GCHQ to the Director of the NSA to mark the 40th anniversary of the UKUSA Agreement in 1986[24]

Flowers had been brought in to design the Heath Robinson's combining unit.[25] He was not impressed by the system of a key tape that had to be kept synchronised with the message tape and, on his own initiative, he designed an electronic machine which eliminated the need for the key tape by having an electronic analogue of the Lorenz (Tunny) machine.[26] He presented this design to Max Newman in February 1943, but the idea that the one to two thousand thermionic valves (vacuum tubes and thyratrons) proposed, could work together reliably, was greeted with great scepticism,[27] so more Robinsons were ordered from Dollis Hill. Flowers, however, knew from his pre-war work that most thermionic valve failures occurred as a result of the thermal stresses at power up, so not powering a machine down reduced failure rates very substantially.[28] Flowers persisted with the idea and obtained support from the Director of the Research Station, W Gordon Radley.[29] Flowers and his team of some fifty people in the switching group[30][31] spent eleven months from early February 1943 designing and building a machine that dispensed with the second tape of the Heath Robinson, by generating the wheel patterns electronically.

This prototype, Mark 1 Colossus, performed satisfactorily at Dollis Hill on 8 December 1943[32] and was taken apart and shipped to Bletchley Park, where it was delivered on 18 January and re-assembled by Harry Fensom and Don Horwood.[33][34] It attacked its first message on 5 February 1944.[5] As it was a large structure it was quickly dubbed Colossus by the WRNS operators. This machine contained 1600 thermionic valves (tubes).[30] and was soon followed by an improved production Mark 2 machine.[35] Nine of this version of the machine were constructed, the first being commissioned on 1 June 1944, after which Allen Coombs took over leadership of Colossus production.[36] The original Mark 1 machine was converted into a Mark 2 and an eleventh Colossus was essentially finished when the war in Europe ended.

The main units of Flowers' design were as follows.[26]

- A tape transport and photo-electric reading mechanism very similar to Heath Robinson's.
- A coder and adder that simulated the Lorenz machine using thyratron rings.
- A logic unit that performed Boolean operations.
- A master control that contained the electronic counters.
- A printer.

Most of the design of the electronics was the work of Tommy Flowers, assisted by William Chandler, with Sidney Broadhurst working on the auxiliary electromechanical parts.[37] The Mark 2 Colossus was designed while Mark 1 was being constructed. It contained 2400 valves and was both 5 times faster and simpler to operate than the original version.[38]

Flowers overcame the problem of synchronizing the electronics with the message tape by generating a clock signal from the reading of the sprocket holes of the message tape. The speed of operation was thus limited by the mechanics of reading the tape. The tape reader was tested up to 9700 characters per second (53 mph) before the tape disintegrated. So 5000 characters/second (40 ft/s (12.2 m/s; 27.3 mph)) was settled on as the speed for regular use.

The Mark 2 Colossus included the first ever use of what would now be called shift registers[39] one for each of the five channels of the punched tape. There were five parallel processing units each involving up to 100Boolean operations – although in normal operation fewer channels were examined in most runs. This five-way parallelism[40] enabled five simultaneous tests and counts to be performed. For each circuit of the tape, the shift register stored successive bits from each of the tape channels and delivered five successive characters to the processors, giving an effective processing speed of 25,000 characters per second. [39]

Operation[edit]

See also: Cryptanalysis of the Lorenz cipher

Colossus used state-of-the-art vacuum tubes (thermionic valves), thyratrons and photomultipliers to optically read a paper tape and then applied programmable logical functions to the bits of the key and ciphertext characters, counting how often the function returned "false".

Colossus was designed to perform the task of "Wheel Setting", that is determining the start point of the stream of key characters in relation to the characters of the enciphered message on the paper tape loop. Initially it was only the χ (*chi*) wheels that were examined. To keep the size of the task manageable, only two bits of the *chi*-stream were examined in the first run,[41] then progressively the other bits.[42] Success at this stage allowed the production of a version of the ciphertext from which the *chi* component of the key had been removed, the so-called "de-*chi*". This transformation allowed manual methods to be used to work out the settings of the ψ (*psi*) and μ *mu* "motor" wheels.

Later, methods were devised for using Colossus to determine the settings of the *psi* wheels. All of this required that "wheel breaking", the discovery of the cam patterns for all the wheels, had been successfully achieved. Later Mark 2 Colossi were equipped with a special unit to achieve this as

well. Programming Colossus was by setting switches and plugging appropriate units together. Sometimes, two or more Colossus computers tried different possibilities simultaneously in what is now called parallel computing, speeding the decoding process by perhaps as much as double the rate of comparison.[citation needed]

Influence and fate[edit]

Colossus was the first of the electronic digital machines with programmability, albeit limited by modern standards.[43]

- It had no internally stored programs. To set it up for a new task, the operator had to set up plugs and switches to alter the wiring.
- Colossus was not a general-purpose machine, being designed for a specific cryptanalytic task involving counting and Boolean operations.

A Colossus computer was thus not a fully general Turing complete machine. However, Professor Benjamin Wells of the Departments of Computer Science and Mathematics, University of San Francisco, has shown[44] that a Universal Turing Machine could have been run on the set of ten Colossus computers. This means that Colossus satisfies the definition of Turing completeness. Most of the other computing machines of this era were also not Turing complete (e.g. the Atanasoff–Berry Computer, the Bell Labs relay machines (by George Stibitz et al.), or the first designs of Konrad Zuse).[citation needed] The notion of a computer as a general purpose machine—that is, as more than a calculator devoted to solving difficult but specific problems—did not become prominent until after World War II.

Colossus was preceded by several computers, many of them first in some category. Zuse's Z3 was the first functional fully program-controlled computer, and was based on electromechanical relays, as were the (less advanced) Bell Labs machines of the late 1930s (George Stibitz, et al.). The Atanasoff–Berry Computer was electronic and binary (digital) but not programmable. Assorted analog computers were semiprogrammable; some of these much predated the 1930s (e.g., Vannevar Bush). Babbage's Analytical engine design predated all these (in the mid-19th century), it was a decimal, programmable, entirely mechanical construction—but was only partially built and never functioned during Babbage's lifetime. Colossus was the first combining *digital*, (partially) *programmable*, and *electronic*. The first fully programmable digital electronic computer was the ENIAC which was completed in 1946.

The use to which the Colossus computers were put was of the highest secrecy, and the Colossus itself was highly secret, and remained so for many years after the War. Thus, it could not be included in the history of computing hardware for many years, and Flowers and his associates were deprived of the recognition they were due.

Being not widely known, Colossus had little direct influence on the development of later computers; it was EDVAC that was the early design which had the most influence on subsequent computer architecture. However, the technology of Colossus, and the knowledge that reliable high-speed electronic digital computing devices were feasible, did have a significant influence on the development of some early computers in the United Kingdom and probably in the US. A number of people who were associated with the project and knew all about Colossus played significant roles in early computer work in the UK. In 1972, Herman Goldstine wrote that:

Britain had such vitality that it could immediately after the war embark on so many well-conceived and well-executed projects in the computer field.[45]

In writing that, Goldstine was unaware of Colossus, and its legacy to those projects of people such as Alan Turing (with the Pilot ACE and ACE), and Max Newman and I. J. Good (with the Manchester Mark 1 and other early Manchester computers). Brian Randell later wrote that:

the COLOSSUS project was an important source of this vitality, one that has been largely unappreciated, as has the significance of its places in the chronology of the invention of the digital computer.[46]

Colossus documentation and hardware were classified from the moment of their creation and remained so after the War. Tommy Flowers was ordered to destroy all documentation and burnt them in a furnace at Dollis Hill. He later said of that order:

That was a terrible mistake. I was instructed to destroy all the records, which I did. I took all the drawings and the plans and all the information about Colossus on paper and put it in the boiler fire. And saw it burn.[47]

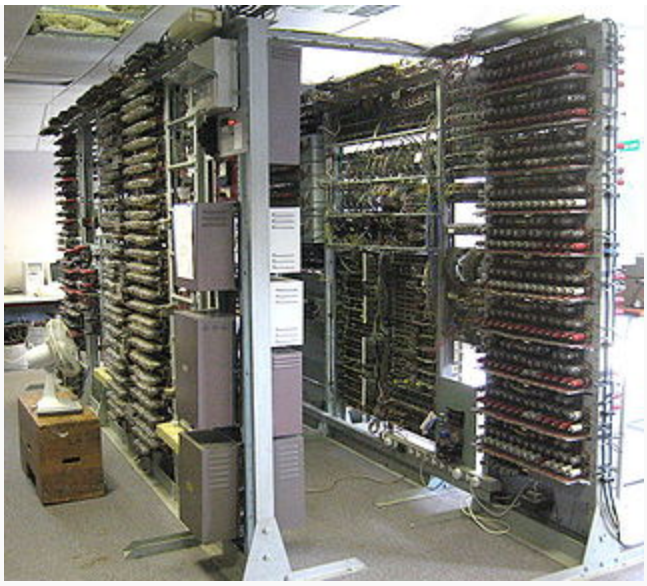
Some parts, sanitised as to their original use, were taken to Newman's Royal Society Computing Machine Laboratory at Manchester University.[48] Most of the Colossus computers were dismantled and parts returned to the Post Office. Two, along with two replica Tunny machines, were retained, moving to GCHQ's new headquarters at Eastcote in April 1946, and moving again with GCHQ to Cheltenham between 1952 and 1954.[49] One of the Colossi, known as *Colossus Blue*, was dismantled in 1959; the other in 1960.[49] There had been attempts to adapt them to other purposes, with varying success; in their later years they had been used for training.[50] Jack Good relates how he was the first to use it after the war, persuading the NSA that Colossus could be used to perform a function for which they were planning to build a special-purpose machine.[49] Colossus was also used to perform character counts on one-time pad tape to test for non-randomness.[49]

For nearly three decades after the war Colossus remained secret, long after any of its technical details were of any importance. The need for such secrecy ebbed away as communications moved to digital transmission and all-digital encryption systems became common in the 1960s. Information about Colossus began to emerge publicly in the 1970s, after the secrecy imposed was broken when Group Captain Winterbotham published his 1974 book *The Ultra Secret*. More recently, a 500-page technical report on the Tunny cipher and its cryptanalysis – entitled *General Report on Tunny* – was released by GCHQ to the national Public Record Office in October 2000; the complete report is available online,[51] and it contains a fascinating paean to Colossus by the cryptographers who worked with it:

It is regretted that it is not possible to give an adequate idea of the fascination of a Colossus at work; its sheer bulk and apparent complexity; the fantastic speed of thin paper tape round the glittering pulleys; the childish pleasure of not-not, span, print main header and other gadgets; the wizardry of purely mechanical decoding letter by letter (one novice thought she was being hoaxed); the uncanny action of the typewriter in printing the correct scores without and beyond human aid; the stepping of the display; periods of eager expectation culminating in the sudden appearance of the longed-for score; and the strange rhythms characterizing every type of run: the stately break-in, the erratic short run, the regularity of wheel-breaking, the stolid rectangle interrupted by the wild leaps of the carriage-return, the frantic chatter of a motor run, even the ludicrous frenzy of hosts of bogus scores.[52]

Reconstruction[edit]

Front view of the Colossus rebuild showing, from right to left (1) The "bedstead" containing the message tape in its continuous loop and with a second one loaded. (2) The J-rack containing the master control panel and jack field. (3) The K-rack with the large "Q" switch panel and sloping patch panel. (4) The double S-rack containing relays and, above the image of a postage stamp, five two-line counter displays. (5) The electric typewriter in front of the five sets of four "set total" decade switches in the C-rack.[53]



Rear view of the two bays of the Colossus rebuild, showing many of the 2400 vacuum tubes and thyratrons used.

Construction of a fully functional replica[54][55] of a Colossus Mark 2 was undertaken by a team led by Tony Sale.[56] In spite of the blueprints and hardware being destroyed, a surprising amount of material survived, mainly in engineers' notebooks, but a considerable amount of it in the U.S. The optical tape reader might have posed the biggest problem, but Dr. Arnold Lynch, its original designer, was able to redesign it to his own original specification. The reconstruction is on display, in the historically correct place for Colossus No. 9, at The National Museum of Computing, in H Block Bletchley Park in Milton Keynes, Buckinghamshire.

In November 2007, to celebrate the project completion and to mark the start of a fundraising initiative for The National Museum of Computing, a Cipher Challenge[57] pitted the rebuilt Colossus against radio amateurs worldwide in being first to receive and decode three messages enciphered using the Lorenz SZ42 and transmitted from radio station DL0HNF in the *Heinz Nixdorf MuseumsForum* computer museum. The challenge was easily won by radio amateur Joachim Schüth, who had carefully prepared[58] for the event and developed his own signal processing and code-breaking code using Ada.[59] The Colossus team were hampered by their wish to use World War II radio equipment,[60] delaying them by a day because of poor reception conditions. Nevertheless, the victor's 1.4 GHz laptop, running his own code, took less than a minute to find the settings for all 12 wheels. The German codebreaker said: "My laptop digested ciphertext at a speed of 1.2 million characters per second—240 times faster than Colossus. If you scale the CPU frequency by that factor, you get an equivalent clock of 5.8 MHz for Colossus. That is a remarkable speed for a computer built in 1944."[61]

The Cipher Challenge verified the successful completion of the rebuild project. "On the strength of today's performance Colossus is as good as it was six decades ago", commented Tony Sale. "We are delighted to have produced a fitting tribute to the people who worked at Bletchley Park and whose brainpower devised these fantastic machines which broke these ciphers and shortened the war by many months."[62]