Semiconductor Technology from A to Z

Analog & Mixed-Signal Design

Operational Amplifiers

Fundamentals & the Ideal Op-Amp

The operational amplifier (op-amp) is the central building block of analog circuit design. It amplifies the voltage difference between its two inputs – the inverting (–) and the non-inverting (+) input – by a very large factor, the open-loop gain A_0.

The ideal op-amp has four properties that serve as a simplification for hand analysis of circuits:

- infinite open-loop gain ($A_0 \to \infty$)
- infinite input resistance (no input current)
- zero output resistance (ideal voltage source at the output)
- infinite bandwidth

In a negative-feedback circuit (e.g. an amplifier with a resistor network), the high open-loop gain gives rise to the virtual ground: feedback drives the voltage difference between the two inputs to nearly zero, even though there is no direct conductive path between them. This principle underlies the analysis of nearly all basic op-amp circuits (inverting amplifier, non-inverting amplifier, integrator, difference amplifier).

Real op-amps deviate from this ideal: finite gain, finite bandwidth, input currents (bias currents drawn by the transistor input stage), and a non-zero output resistance. These non-ideal effects are quantified in the following sections.

Differential Pair & Current Mirror as Input Stage

The input stage of an op-amp is almost always a differential pair: two transistors whose sources (MOS) or emitters (bipolar) connect to a shared current source. The difference between the two input voltages steers how the shared bias current splits between the two branches.

In modern CMOS designs, the active load is typically a current mirror: a reference transistor sets a current that is mirrored to one or more further branches through the ratio of transistor widths (W/L). In the op-amp input stage, the current mirror converts the differential current signal into a single-ended voltage signal – this step largely sets the gain of the first stage.

Key parameters of this stage are the transconductance g_m of the input transistors (determining gain and noise performance) and the common-mode input range – the range within which both input voltages may lie without

pushing a tail-current or differential-pair transistor out of saturation.

Topologies: Miller OTA, Telescopic Cascode, Folded Cascode

For integrated op-amps (usually implemented as an OTA, Operational Transconductance Amplifier, without a low-impedance output buffer), several standard topologies have become established, each offering a different trade-off between gain, bandwidth, output swing, and area.

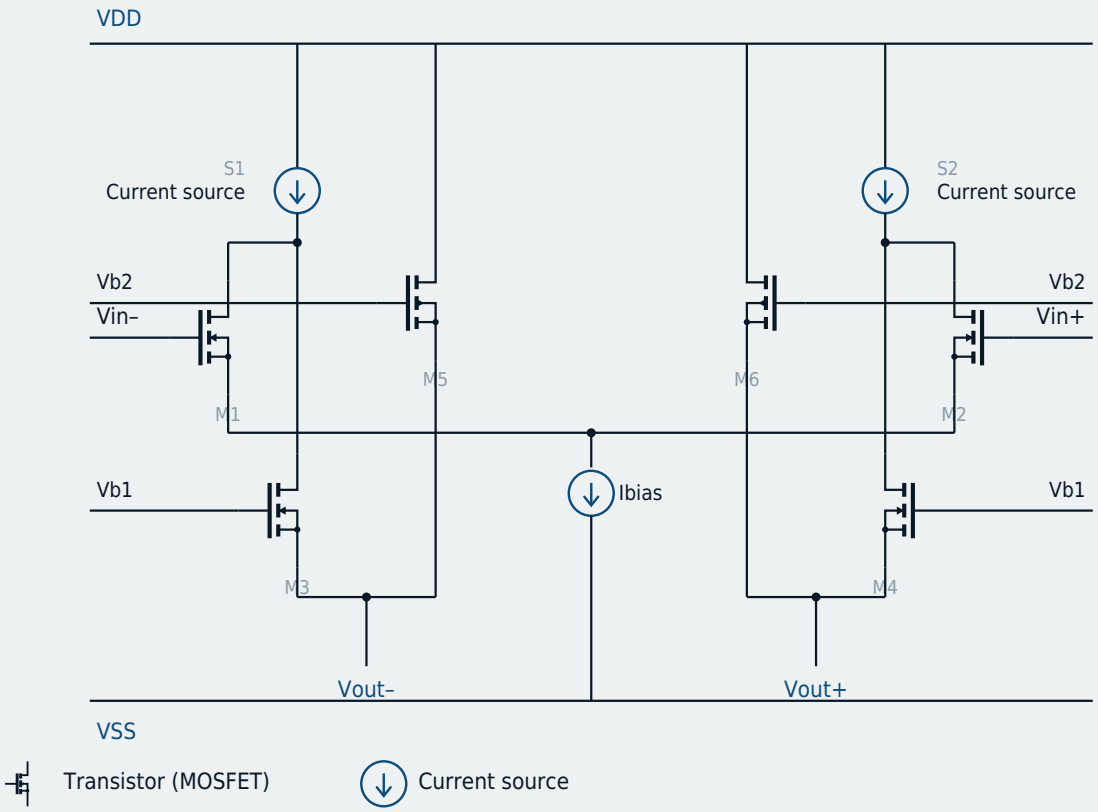
Two-stage Miller OTA: a differential input stage (high gain) followed by a second gain stage (large output swing), connected through a Miller compensation capacitor. This topology achieves very high gain and a large output voltage swing, but is harder to compensate because of the second stage.

Telescopic cascode: cascoding both the differential-pair transistors and the current-mirror load within a single stage substantially raises the output resistance and therefore the gain (typically 60–80 dB) without requiring a second stage. Drawback: the stacked transistors severely limit the usable output swing – unsuitable for low-supply-voltage designs.

Folded cascode: resolves the swing limitation of the telescopic cascode by "folding" the signal path after the input stage onto a complementary cascode branch. This allows a larger output swing at only a modest increase in current consumption, and is the single-stage topology most commonly used in practice.

Folded-Cascode Operational Amplifier

Simplified schematic (single-stage OTA)



The input pair M1/M2 is folded through current sources S1/S2 into cascode branches (M3–M6) toward the output nodes Vout–/Vout+.

Key Parameters: Gain, GBW, Slew Rate, PSRR/CMRR, Offset

Real op-amps are characterized by several parameters found in datasheets and used throughout circuit design:

- Open-loop gain (A_0): gain without feedback, typically 60–120 dB. Determines the accuracy of the closed-loop circuit (residual error $\approx 1/A_0$).
- Gain-bandwidth product (GBW): product of gain and bandwidth at a given operating point; approximately constant across frequency for single-pole behavior. Sets the maximum usable signal bandwidth for a given closed-loop gain.
- Slew rate (SR): maximum rate of change of the output voltage (V/µs), limited by the maximum available charge/discharge current into the compensation capacitor. For large signal steps, slew rate – not bandwidth – limits the actual rise time.
- PSRR (Power Supply Rejection Ratio) and CMRR (Common Mode Rejection Ratio): a measure of how well supply-voltage disturbances and common-mode input signals are rejected.

- Offset voltage: a systematic input voltage difference caused by device mismatch (threshold voltage, W/L ratios) that produces an output error even under ideal feedback.

Frequency Compensation

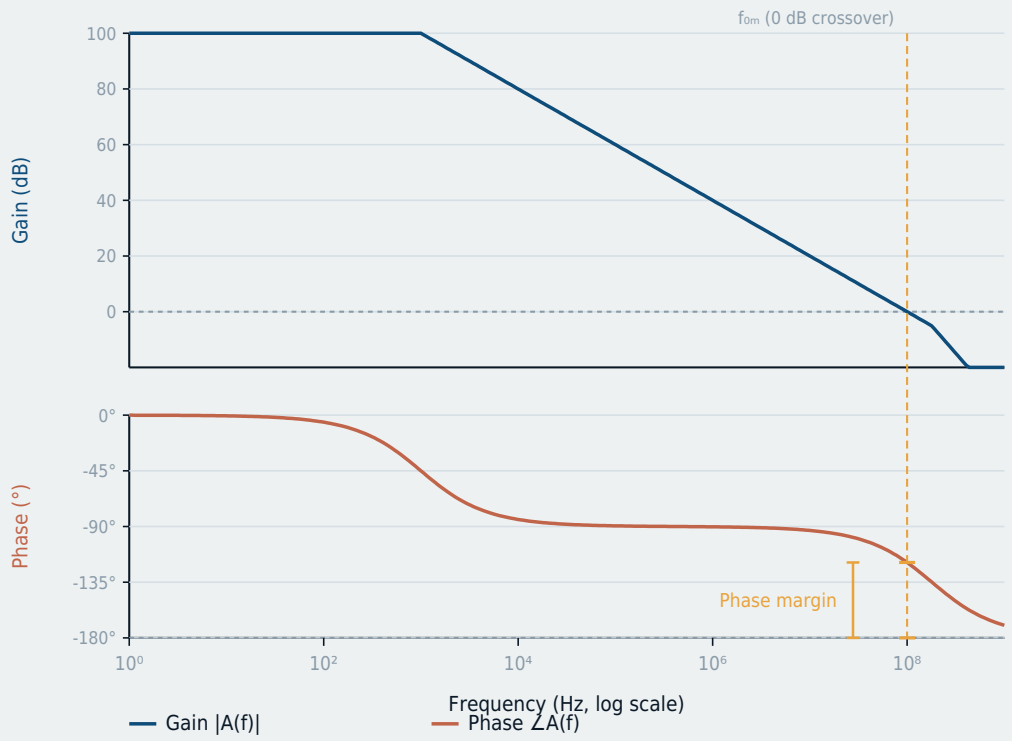
Multi-stage op-amps exhibit several poles in their frequency response. Left uncompensated, this leads to instability (oscillation) under feedback, since the phase shift can reach or exceed 180° at the frequency where the loop gain drops to 0 dB.

The phase margin – the difference between the phase and –180° at the 0 dB crossover frequency – is the central stability criterion. A phase margin of at least 45–60° is generally considered adequate for good settling behavior without excessive overshoot.

The most common technique is Miller compensation: a capacitor between the output and the intermediate node of the second stage exploits the Miller effect to push the dominant pole to a lower frequency (pole splitting), keeping the remaining poles outside the relevant bandwidth. A nulling resistor placed in series with the compensation capacitor cancels the right-half-plane zero it introduces, which would otherwise degrade the phase margin.

Bode Plot with Phase Margin

Gain and phase of a compensated two-stage op-amp



Phase margin = distance of the phase curve to –180° at the 0 dB crossover frequency.

Practice: Common Basic Circuits

Op-amps form the basis of numerous basic analog circuits found in nearly every mixed-signal system:

- Voltage follower (buffer): unity gain, extremely high input and low output resistance – used for impedance transformation between circuit stages.
- Integrator: resistor at the input, capacitor in the feedback path – produces an output voltage that is the time integral of the input; a core building block of filters, sigma-delta modulators, and PLL loop filters.
- Instrumentation amplifier: a combination of several op-amps for precise, low-noise amplification of small differential signals with high common-mode rejection – typically used in sensor signal conditioning.

These basic circuits connect the Operational Amplifiers chapter to the upcoming ADC/DAC architecture and PLL chapters, where OTAs are reused as integrators, gain stages, and charge pumps.

ADC Architectures

Fundamentals: Sampling, Quantization, Resolution

An analog-to-digital converter (ADC) converts a continuous-time, continuous-amplitude signal into a sequence of discrete numerical values. This process consists of two separate steps: sampling in time and quantization of the amplitude.

According to the Nyquist–Shannon sampling theorem, the sampling frequency must be at least twice the highest frequency component contained in the signal in order to reconstruct it without loss of information. Otherwise, aliasing produces indistinguishable mirror frequencies – in practice, an analog anti-aliasing filter therefore usually precedes the ADC.

Quantization rounds the sampled voltage to the nearest of 2^N representable codes (for N-bit resolution). The resulting unavoidable difference is the quantization error, which appears as quantization noise in the output signal. Key parameters for evaluating real converters are SNR (signal-to-noise ratio), SFDR (spurious-free dynamic range), and ENOB (effective number of bits) – the actually usable resolution accounting for noise and distortion.

SAR ADC (Successive Approximation Register)

The SAR ADC converts a sampled voltage into a digital code through a binary search. An internal register tests bit by bit, starting with the most significant bit (MSB), whether the corresponding code value lies above or below the input voltage, using an internal DAC to generate the comparison voltage and a comparator to make the decision.

For a resolution of N bits, the conversion requires exactly N comparison cycles – conversion speed therefore scales linearly with resolution. SAR ADCs feature small area, low power consumption, and moderate-to-high resolution (typically 8–18 bits) at medium sample rates (kHz to a few tens of MHz), making them today's most widely used ADC architecture in mixed-signal ICs.

Pipeline ADC

The pipeline ADC splits the conversion across several stages connected in series. Each stage converts a few bits with a simple flash sub-ADC, reconstructs a reference signal from these bits via an internal DAC, forms the difference from the input signal (the residue), and amplifies this residue for the next stage

(multiplying DAC, MDAC).

Because all stages work simultaneously on different samples (pipelining), this architecture achieves significantly higher throughput than a SAR ADC at comparable resolution – typically 8–16 bits at sample rates from tens of MHz to the GHz range. Drawbacks are the latency introduced by the pipeline stages and higher area and power requirements compared to a SAR ADC.

Sigma-Delta ADC

The sigma-delta ADC (also delta-sigma ADC) takes a fundamentally different approach: instead of high per-sample amplitude resolution, it uses strong oversampling (typically 32–256× the Nyquist rate) combined with a feedback sigma-delta modulator. At its core, this consists of an integrator, a coarse (often 1-bit) quantizer, and feedback through a DAC into the signal path.

The feedback pushes quantization noise to higher frequencies (noise shaping), which is then removed by a digital low-pass (decimation) filter that simultaneously reduces the high sample rate back to the actual Nyquist rate. Sigma-delta ADCs thereby achieve very high effective resolutions (beyond 24 bits) at low-to-moderate bandwidths and are standard in precision instrumentation, audio, and sensing.

Flash ADC

The flash ADC is the fastest but also the most resource-intensive ADC architecture: the input voltage is compared in parallel against 2^N – 1 reference voltages derived from the reference via a resistor ladder. A downstream priority encoder converts the resulting thermometer-coded comparator pattern into a binary code.

Because conversion happens within a single clock cycle, the flash ADC achieves the highest available sample rates (several GHz). However, the comparator count – and thus the required area and power – grows exponentially with resolution, limiting practical resolution to typically 6–8 bits. Flash ADCs are therefore used mainly in high-speed applications, often as a fast sub-ADC stage within a pipeline ADC.

Comparison & Application Areas

Architecture	Typ. resolution	Typ. sample rate	Typical application
SAR	8–18 bit	kHz – approx. 50 MHz	General purpose, sensing, mixed-signal SoCs

Architecture	Typ. resolution	Typ. sample rate	Typical application
Pipeline	8–16 bit	10 MHz – GHz	Communications, video, RF receivers
Sigma-Delta	up to >24 bit	Hz – a few MHz	Precision instrumentation, audio, sensing
Flash	6–8 bit	>1 GHz	High-speed digitization, oscilloscopes

Choosing an architecture is always a trade-off between resolution, speed, area, and power consumption. Modern SoCs also combine architectures depending on the application, such as pipeline stages with flash sub-ADCs, or SAR stages within time-interleaved ADC architectures for the highest sample rates at moderate resolution.

DAC Architectures

Fundamentals: Resolution, INL/DNL, Settling Time

A digital-to-analog converter (DAC) generates an analog voltage or current from an N-bit digital code, selecting from 2^N possible discrete output values.

Two parameters are used to evaluate linearity: DNL (Differential Non-Linearity) describes the deviation of the spacing between two adjacent output codes from the ideal LSB step – a DNL ≤ –1 LSB results in non-monotonic behavior. INL (Integral Non-Linearity) describes the cumulative deviation of the entire transfer curve from the ideal straight line.

For dynamic evaluation, the settling time is decisive: the time required for the output to settle within a defined error band (e.g. ±½ LSB) after a code change. It limits the maximum update rate of the DAC and is influenced by the bandwidth of the output stage (often an op-amp buffer).

R-2R DAC

The R-2R DAC uses a resistor network built from only two resistor values (R and 2R), which can be cascaded to any number of bits without – unlike a binary-weighted resistor network – requiring extreme resistance ratios between the MSB and LSB. At each node of the network, a switch connects either to ground or to the reference voltage depending on the bit value.

The major advantage is good manufacturability using just two precisely matched resistor values, which reduces area and matching requirements. Achievable

accuracy, however, depends strongly on the matching of the switch transistors and resistors; R-2R DACs are commonly used for resolutions in the 8–16 bit range at moderate speeds.

Current-Steering DAC

The current-steering DAC is the dominant architecture for high speed. It consists of an array of binary- or thermometer-weighted current sources that are steered, depending on the digital code, into one of two output branches (differential). An output resistor or transimpedance stage converts the resulting output current into a voltage.

Since no internal op-amps with limited bandwidth sit in the signal path, current-steering DACs achieve very high sample rates (up to the GHz range) at typical resolutions of 8–16 bits. The most significant bits are usually implemented with thermometer-coded segmentation to guarantee monotonicity and minimize glitches at code transitions, while the least significant bits are realized as binary-weighted (segmented architecture).

Sigma-Delta DAC

Analogous to the sigma-delta ADC, the sigma-delta DAC uses oversampling and noise shaping, but in the reverse signal direction: a digital interpolation filter first raises the data rate of the input signal, a digital sigma-delta modulator then reduces the word width (often to 1 bit) while pushing the quantization noise to high frequencies. A very simple, highly linear 1-bit DAC converts the resulting bitstream, and an analog low-pass filter smooths the result into the final analog signal.

Since the actual DAC core resolves only 1 bit, the matching problem of multi-bit resistor or current-source arrays is almost entirely eliminated – at the cost of a higher required clock rate and digital filter complexity. Sigma-delta DACs therefore dominate high-resolution audio applications (typically 16–24 bits).

Comparison & Application Areas

Architecture	Typ. resolution	Typ. speed	Typical application
R-2R	8–16 bit	low – moderate	General signal generation, control systems
Current-Steering	8–16 bit	up to GHz range	RF transmit chains, direct digital synthesis
Sigma-Delta	16–24 bit	low – moderate (audio bandwidth)	Audio DACs, precision signal generation

As with ADCs, the application determines the architecture: high speed at moderate resolution favors current-steering, high resolution at moderate bandwidth favors sigma-delta. R-2R DACs are found today mainly where simple, robust signal generation is sufficient without high-speed requirements.