

ACRONYMS — THIS CHAPTER

ACRONYMS

ADC	analog-to-digital converter
CC0	Creative Commons Zero (public-domain dedication)
CFA	color filter array
DNG	Digital Negative (raw file format)
DSLR	digital single-lens reflex
ETTR	expose to the right
EV	exposure value
EXIF	Exchangeable Image File Format
HDR	high dynamic range
HLG	hybrid log-gamma
HSL	hue, saturation, lightness
ISP	image signal processor
ND	neutral density (filter)
PQ	perceptual quantizer (HDR transfer function)
SNR	signal-to-noise ratio

Course notes · 37 sections from the book · 27 optional

[hide optional sections](#)[the full book ↗](#)

LECTURE NOTES 6 — HDR

Course notes cut from *Computational Photography, an AI-powered Slopendum* for this lecture. Each section links back to its home in the book; sections marked **optional** go beyond what the lecture needs.

► [Contents](#) · 31 sections on this page, from 15 chapters of the book

Image measurements as integrals over the plenoptic function

Noise, signal-to-noise ratio and dynamic range

Dynamic range

OPTIONAL
in the book [↗](#)

Two extremes bound what a single exposure can hold. At the top, a pixel's well saturates: once it has counted its full-well capacity of electrons, more light adds nothing and the highlight clips to white. At the bottom, the read-noise floor sets the dimmest signal still distinguishable from noise. The ratio of the two is the sensor's dynamic range: how much of a high-contrast scene a *single* exposure can capture. When a scene's range exceeds the sensor's (a bright window and a dim interior in one frame), no single exposure holds both ends, which is the motivation for high-dynamic-range (HDR) imaging, merging several exposures ([Chapter 17.1, Dynamic range and multiple-exposure capture](#); [Figure 2.8.14](#)). Like everything else a photographer counts, dynamic range is quoted in stops (factors of two), which connects it straight to the exposure controls.

One subtlety keeps dynamic range from being the hard number it looks like: the *bottom* is not a sharp cliff where signal abruptly vanishes, but wherever the signal-to-noise ratio falls below what you are willing to accept. Engineers usually quote the *engineering* dynamic range, placing the floor at $\text{SNR} = 1$ (signal equal to noise); that is the most generous possible bound, and full-well divided by read noise approximates this criterion when read noise dominates at the threshold. With shot noise included, the electron signal N at $\text{SNR } q$ solves $N/\sqrt{N + \sigma_r^2} = q$, giving $N_{\min} = (q^2 + \sqrt{q^4 + 4q^2\sigma_r^2})/2$. The usable range is then $\log_2(N_{\text{sat}}/N_{\min})$ stops, where N_{sat} is the effective saturation signal at the chosen gain. A photographer who wants genuinely *clean* shadows demands a far higher SNR, say 10 or 20, sets the floor higher, and counts fewer usable stops. This is why measurement sites report dynamic range at a stated quality threshold and their numbers disagree: the same sensor has a wide *engineering* range and a narrower *photographic* one. In this respect dynamic range behaves like depth of field: both carry a crisp-looking definition that actually rests on a chosen tolerance (an acceptable SNR here, an acceptable circle of confusion there), so the single figure you read off is only ever as hard as the criterion behind it.

There is one more limit that no sensor spec sheet lists: the *lens*. Bright parts of a scene scatter stray light inside the lens and camera body (faint reflections off every element surface, the diaphragm blades, and the barrel walls) that settle as a veil of light spread across the whole frame. This *veiling glare*, the diffuse cousin of lens flare, lifts the darkest tones off true black, adding a structured background and its shot noise. Even if the mean glare could be subtracted, its photon noise would remain; this differs from electronic read noise. It therefore caps the dynamic range a real photograph can reach *below* whatever the bare sensor could deliver, and it bites hardest exactly when dynamic range matters most, shooting a dim subject against a bright source. It is why lens coatings, hoods, and a clean front element buy real shadow contrast, and why the deepest blacks in a studio shot are easier to protect than those in a backlit scene. Flare and veiling glare get their full treatment as optical defects in [Chapter 22.1, Aberrations and optical challenges](#) (§10.1).

The dynamic-range problem: one high-contrast scene, no single exposure that holds both ends



Figure 2.8.14 (reproduced). *The dynamic-range problem. One high-contrast scene (a sunset over water) developed at three exposures: expose for the highlights and the foreground is crushed to black; expose for the shadows and the sky blows out to white; the compromise loses both ends. No single exposure holds the full range: the motivation for high-dynamic-range merging of multiple exposures.*

How wide a range, exactly? It helps to put numbers on it, because the dynamic ranges of real media and of the eye differ by a lot ([Figure 2.8.15](#)). A color slide is narrow, only about 5–6 stops; a reflective print about 6–7; a film negative is famously forgiving at roughly 12–13 stops of latitude; a phone sensor manages about 10–12 and a full-frame sensor about 14. The human eye sees about 10–14 stops *instantaneously*, but allow it to adapt (the slow re-centering of the retinal response covered with [Chapter 3.3, Visual processing](#)) and it spans 20+ stops over time; some animals do better still within their niche. Against all of these, an outdoor sun-and-shadow scene can exceed 20 stops in a single frame, which is precisely why no single capture holds it, and why HDR exists.

Dynamic range: a ladder of stops across media and the eye

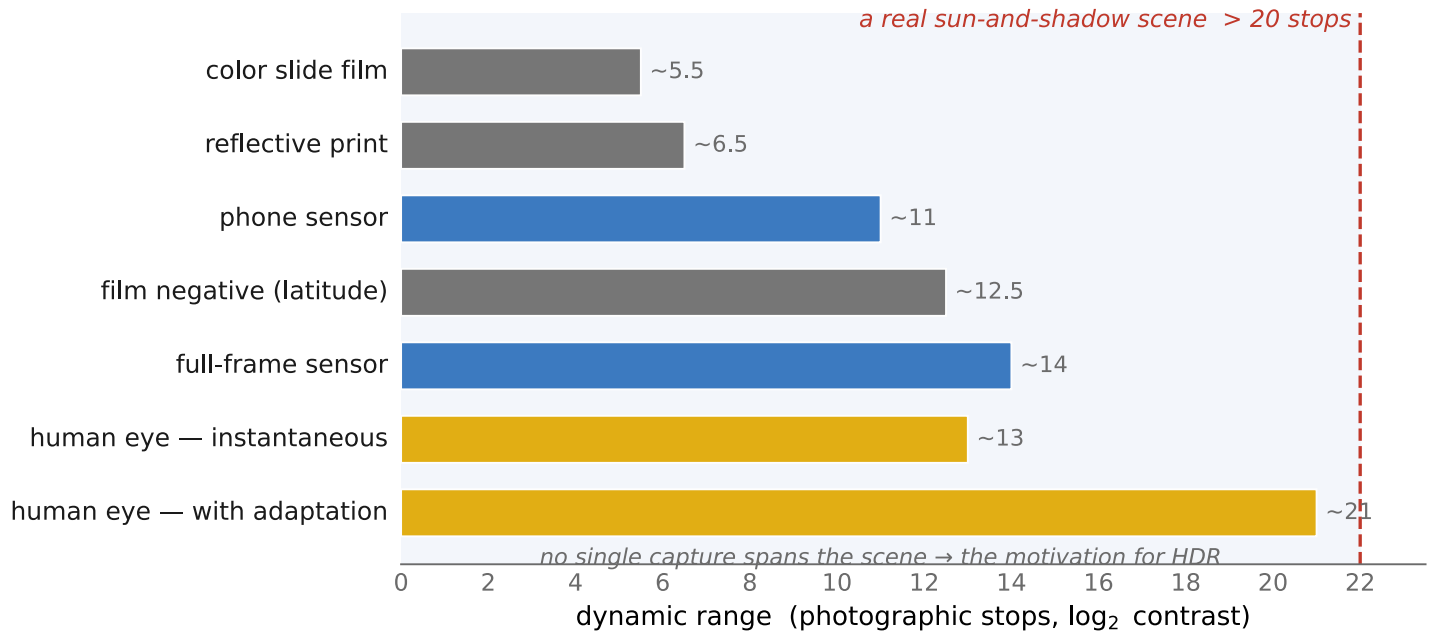


Figure 2.8.15 (reproduced). *Dynamic range across media and eyes, as a ladder of stops: color slide (~5–6 stops), reflective print (~6–7), film negative (~12–13 latitude), phone sensor (~10–12), full-frame sensor (~14), and the human eye (~10–14 instantaneous, ~20+ once adaptation is allowed), with an animal example, set against a real sun-and-shadow scene exceeding 20 stops, which no single capture spans: the motivation for HDR.

Recovering an underexposed shot: full-frame vs. phone. The clearest demonstration is to underexpose one. Shoot the *same* underexposed scene as raw on a full-frame camera and on a phone, then push the exposure up in software by three or four stops (Figure 2.8.16). The full-frame frame cleans up (its deep full-well and low read-noise floor give it the headroom to be lifted) while the phone frame breaks into shadow noise and banding, because lifting the exposure also amplifies its read-noise floor, which sat much closer to the signal to begin with. At matched framing, exposure, and output resolution, this illustrates the benefit a larger total collecting area can provide; it is not a universal rule that sensor size alone determines dynamic range, and exactly why phones lean on burst capture and HDR+ (Multiple-exposure part) rather than trusting a single exposure. It is also a coding exercise: see [Exercises & Experiments](#).

[figure fig-underexposure-recovery-ff-vs-phone not built]

Figure 2.8.16 (reproduced). (Deferred, to be generated.) The same underexposed scene, shot as raw on a full-frame camera and on a phone, then pushed up about 3–4 stops in software: the full-frame frame cleans up (deep full-well, low read-noise floor → wide dynamic range), while the phone frame reveals amplified shadow noise and banding (its read-noise floor lifted with the signal): dynamic range scaling with sensor size.

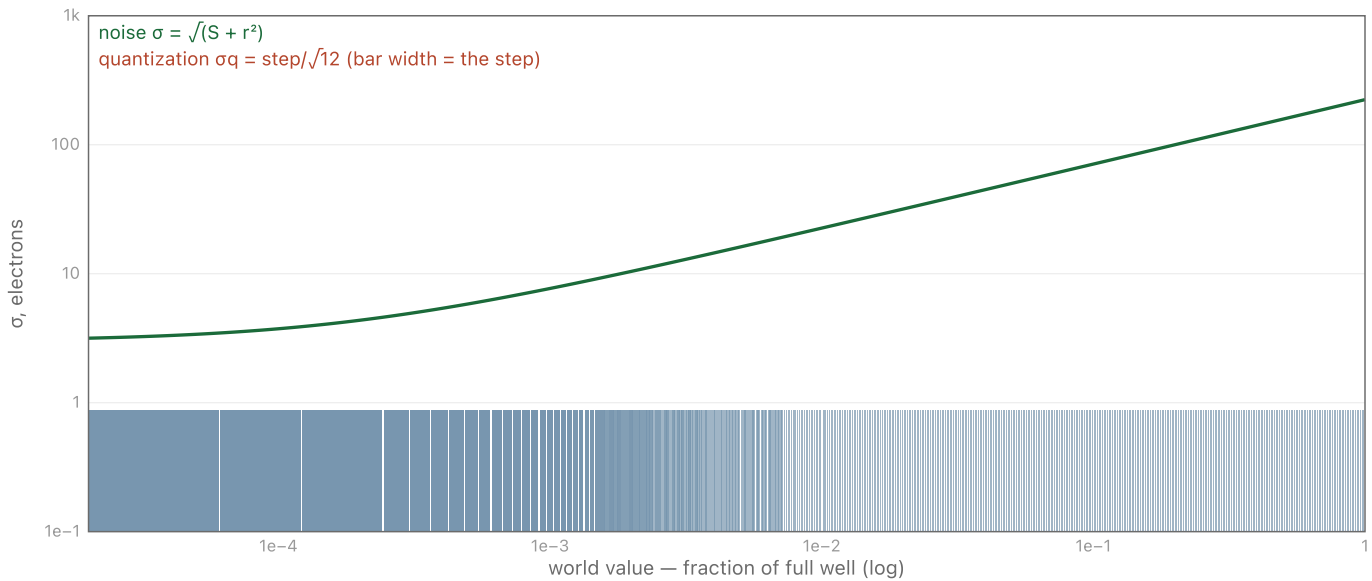
💡 **BIG LESSON (L2.46) — $\text{DYNAMIC RANGE} = \text{FULL-WELL CAPACITY} \div \text{NOISE FLOOR}$**

A single exposure records from a **top** (the photosite's **full-well capacity**, where it saturates and clips to white) down to a **bottom** (the **noise floor**, the read noise that drowns the shadows). Their **ratio is the dynamic range**, quoted in stops. That floor is **not absolute**: it sits wherever the SNR drops below what you deem acceptable, so, just as with **depth of field** and its circle of confusion, the number depends on the quality criterion you pick (SNR = 1 gives the generous *engineering* range; a stricter SNR gives the narrower *usable* one). You **widen** it with a **bigger well** (larger photosites, why a full-frame sensor out-ranges a phone) or a **lower floor** (cooling, lower read noise), and you **beat** it altogether by merging **multiple exposures** (HDR). This is the capture-side companion to the quantization lesson below: what bounds a single shot is this *range*, full-well over floor, not the number of bits.

This is also the right moment to deflate a worry beginners often have about digital capture: that the image is *quantized* into a finite number of levels, and that this stair-stepping is what limits quality. With adequate raw bit depth and noise comparable to a code step, quantization is often a smaller limitation than sensor noise. Aggressive tonal edits or insufficient precision can still expose banding.

💡 **BIG LESSON (L2.47) — QUANTIZATION IS RARELY THE REAL PROBLEM**

With enough bits and a sane encoding, what limits image quality is **noise and dynamic range** (the two quantities of this section), **not** the number of levels the analog-to-digital converter (ADC) offers. The noise floor and the full well bound what you can record; the quantization step, once it sits below the noise, is invisible. (Indeed, a little noise *dithers* the quantization, hiding it further.) Levels matter in one place: **banding in the deep shadows** of a linearly-encoded image, too few codes where the eye, working in ratios, is most sensitive, and that is exactly the problem **gamma encoding** exists to solve, by spending code levels perceptually ([Chapter 4.1, Color space 101](#)). So "rarely" is not "never." We meet this lesson again in BASIC → Image representation → *Float vs 8-bit*, where floats with headroom are the sane default.



Sampled quantization steps lie below the sensor noise; their contribution is smaller.

14 bits $\rightarrow$ 16,384 levels $\cdot \gamma = 1.0 \cdot$ read $3.0 \text{ e}^- \cdot$ well $50,003 \text{ e}^- \cdot 3.1 \text{ e}^-$ per code at $\gamma = 1$

Figure 2.8.17 (reproduced). *The quantization step against the noise it has to beat (interactive). Both are standard deviations in the same units, so they can be compared directly: photon noise is Poisson, so $\sigma = \sqrt{S + r^2}$ electrons at a signal of S electrons with read noise r , while one code step contributes $\sigma_q = \text{step}/\sqrt{12}$. The **bars are the code steps** and their width is the step itself, uniform under $\gamma = 1$, squeezed into the shadows as γ rises. Where a bar pokes **above** the noise curve, quantization is adding something the sensor was not already contributing; everywhere else it is buried, and the noise dithers it away. The scenario buttons set plausible hardware: a **14-bit linear raw** from a good mirrorless, an **8-bit gamma-encoded JPEG** from the same camera, and an **8-bit phone** file whose pixel holds an order of magnitude fewer electrons. In the first three every step sits under the noise. Only **8-bit linear** crosses, and only in the deepest shadows, the one place the lesson allows for, and exactly what gamma encoding exists to fix. Read noise, full well, bit depth and gamma are all adjustable, and the vertical axis can be read in electrons, as a fraction of the well, or in JNDs.*

These three facts (that noise is the jitter of a photon count, that the signal-to-noise ratio is worst in the shadows, and that dynamic range is the full-well-over-floor a single exposure can hold) recur on nearly every later page: in how a photographer exposes to the right, in high-dynamic-range merging (Multiple-exposure part), and in the denoising of the basic pipeline. With measurement and its limits now in hand, the part turns next to what a *second* viewpoint buys you (depth), and then to the eye that will finally judge the result.

Exposure and capture timing

Basic photography: exposure settings (shutter, aperture, and ISO)

OPTIONAL

For an unchanged scene and lens transmission, optical exposure scales as

$$H \propto \frac{t}{N^2}, \quad N = \frac{f}{D}. \quad (2)$$

Twice the integration time collects twice the light. Multiplying the f-number by $\sqrt{2}$ halves it. Thus f/2.8 at 1/250 s and f/4 at 1/125 s give approximately the same optical exposure, but the longer integration blurs motion more and the smaller aperture increases depth of field.

A stop is the base-two logarithm of a light ratio. The setting-based exposure value uses time in seconds:

$$\text{EV} = \log_2 \left(\frac{N^2}{t/(1 \text{ s})} \right). \quad (3)$$

Increasing this EV by one halves optical exposure. Positive exposure compensation is a different convention: it asks an automatic mode for a brighter result relative to its metered choice. In aperture priority at fixed ISO, +1 stop will ordinarily double time. In manual mode with Auto ISO it may instead double gain while leaving the light collected unchanged.

ISO is not a third optical control. At fixed aperture and time, increasing gain scales the signal and its shot noise together. It can reduce the importance of noise added downstream of that gain, or select a sensor conversion-gain mode with lower input-referred read noise. It can also reduce highlight headroom. On an ISO-invariant sensor over the relevant range, low-ISO raw data brightened later can be close to a higher-ISO capture. The comparison in [Chapter 2.8, Noise, signal-to-noise ratio and dynamic range](#) applies only after matching output brightness and checking clipping.

The common association between high ISO and noise usually includes a reduction in collected light: we raise ISO because the scene is dim or the shutter must be fast. Add light, open the aperture or lengthen time when the photograph permits it. Raising ISO alone does not improve photon statistics.

Reflected-light metering measures light coming from the scene; it cannot separate a dark object in strong light from a pale object in weak light without additional information. Spot metering makes the chosen patch explicit. Evaluative metering uses more of the frame and may use recognized subjects. A gray card is a useful reference under the subject's illumination, not proof that a meter assumes exactly 18% reflectance. Snow and dark clothing are good tests of when compensation is needed.

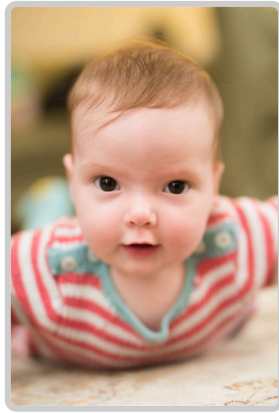
Expose to the right means collecting more light while retaining the highlights that matter. It is useful when time, aperture and saturation allow it; it is not a requirement to force every histogram against the edge. A longer exposure that blurs the subject is a poor trade for cleaner shadows. Inspect individual channels when possible, because a colored light can saturate one channel before the luminance preview looks white. Keep a margin when the camera's warnings refer to a processed preview rather than raw data.



Figure 5.2.1 (reproduced). Interactive: the faucet-and-bucket exposure analogy. Two buckets fill from faucets; the water caught is the exposure and the dashed line marks a correct one. The faucet diameter is the aperture (flow rate follows the opening's area, $\propto \text{Ø}^2$) and the fill duration is the shutter time. The left bucket is a fixed reference (wide faucet, short time); the right one you control, defaulting to the opposite (a narrow faucet open a long time), yet it catches the same water, because three stops of smaller aperture cancel three stops of longer time. Press to fill the

buckets and watch the narrow one keep pouring long after the wide one has finished; move either slider and the right bucket over- or under-fills (over- or under-exposes), with the swatch above it showing the resulting brightness: mid-gray when the exposure is correct, white when the bucket overflows into clipped highlights, dark when it is starved. A running stop tally makes the reciprocity explicit. The dashed line is one choice of target, and ISO, which changes how the reading is scaled rather than how much water is caught, has no place in the analogy.

Aperture and depth of field: a wide aperture isolates the subject; a small one keeps everything sharp. **Click a photo to enlarge.**



f/1.4 — wide aperture, shallow focus



f/22 — narrow aperture, deep focus

Photos credited in the book's image credits.

Figure 5.2.2 [\(reproduced\)](#). Aperture versus depth of field. Left: an infant portrait at f/1.4, the wide aperture throwing the background into a smooth blur that isolates the subject, admitting a lot of light but holding only a thin slab in focus. Right: a mountain scene at f/22, the narrow aperture keeping the near wildflowers and the distant fjord and peaks all sharp at once, at the cost of much less light (a longer shutter or higher ISO). The same *f*-number is both a brightness setting and a depth-of-field setting. Both photographs © Frédo Durand.

Shutter speed: freeze the motion, or let it blur. **Click a photo to enlarge.**



fast shutter, $1/3200$ s — motion frozen



slow shutter, 1 s — motion blurs to silk

Photos credited in the book's image credits.

Figure 5.2.3 (reproduced). Shutter speed versus motion blur. Left: a child bursting up through the water at $1/3200$ s, fast enough to freeze every flung droplet in mid-air. Right: a waterfall at 1 s, long enough for the moving water to smear into a silky veil while the rock stays sharp. A fast shutter records a single instant; a slow one integrates the subject's motion over its whole open time, and admits far more light doing so (here held back by a small aperture). The same shutter time is both a brightness setting and a motion-rendering choice. Both photographs © Frédo Durand.

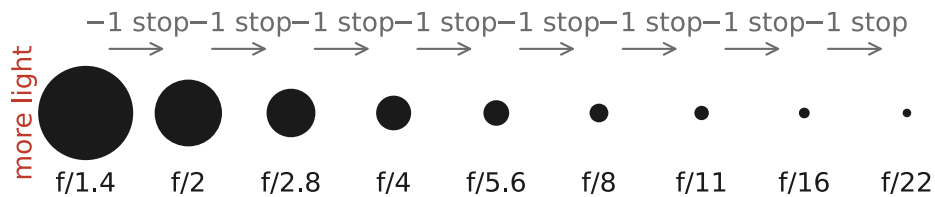


Figure 5.2.4 (reproduced). The f-stop progression as powers of two. The standard series $f/1.4, f/2, f/2.8, f/4, f/5.6, f/8, f/11, f/16$ drawn as aperture disks: each f-number is $\sqrt{2}$ times the previous, so the disk area, and the light admitted, halves at every step. The irregular-looking numbers are exactly the $\sqrt{2}$ spacing that makes each step one stop (a factor of two in light).

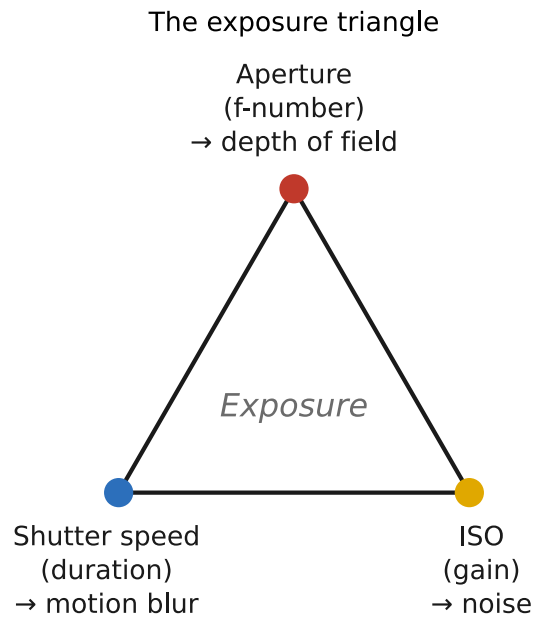


Figure 5.2.5 [\(reproduced\)](#). *The exposure triangle. Three controls set how bright a photo is, each in power-of-two stops (one stop = $\times 2$ light = 1 EV): shutter time (longer = brighter, but motion blur), aperture / f-number (wider = brighter, but shallower depth of field), and ISO (higher = a brighter rendering of the same light; the extra noise comes from the smaller exposure it lets you use, not from the setting itself). A stop of one can be traded for a stop of another to keep brightness constant while changing the side effects.*

The exposure triangle



Scene light



starlight → full moon → night → dark office → home → overcast → hazy sun → bright sun

Shutter



slower → more light + motion blur

Aperture



wider (small f) → more light + shallow depth of field

ISO



higher → brighter, not cleaner: gain scales noise with signal

Subject Sensor Auto-set (keep exposure constant):

2nd person depth



Subject motion

 [copy link](#)

Figure 5.2.6 [\(reproduced\)](#). *Try it yourself (web edition)*. A live exposure-triangle simulator: two 3D subjects stand at different depths and walk in opposite directions while you set shutter, aperture, ISO, scene brightness, and sensor size. It renders the scene physically by supersampling over time for real motion blur, over the aperture for real depth of field, and through a photon-shot-plus-read-noise model for real ISO grain. The noise model is illustrative, not a particular camera's. An "auto-set triangle" holds the rendered brightness constant so you feel each control's trade-off (freeze vs blur, deep vs shallow focus, clean vs noisy) rather than just its brightness. In the printed edition this stands in as a screenshot.

Metering to a middle tone. A reflected-light meter cannot tell what lightness you intend, so it renders what it measures as a middle tone. That suits an average scene; snow wants positive compensation and a dark subject negative. The 18% gray card is a practical reference, not the exact reflectance every meter is calibrated to. **Click any photo to enlarge.**



snow field — high-key (meter under-exposes)



old-town rooftops — average (meter is right)



dark cat — low-key (meter over-exposes)

Photos credited in the book's image credits.

Figure 5.2.7 [\(reproduced\)](#). *The 18% gray assumption and where it fails. The meter renders the average scene to mid-gray (~18%). For an ordinary scene this is right. For a snowfield (high-key), forcing the bright average to gray underexposes; the snow turns dull; the fix is to add +EV. For a black subject (low-key), forcing the dark average to gray overexposes; the fix is -EV. Matrix metering uses extra cues to guess intent.*

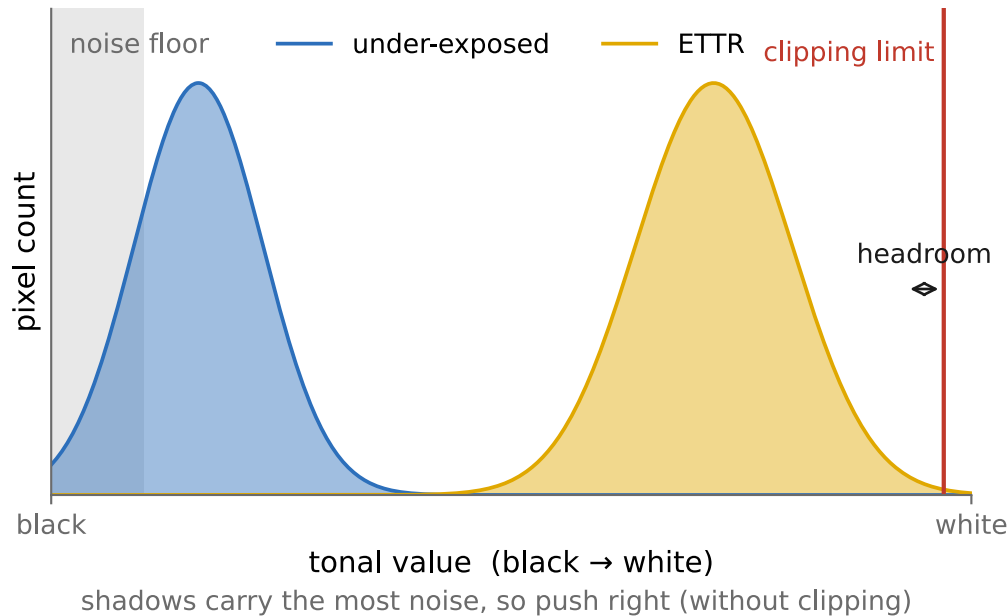


Figure 5.2.8 (reproduced). *Expose to the right (ETTR)*. Two captures of the same scene: (left) a timid exposure with the histogram bunched at the noise floor, later brightened in software, which amplifies shadow noise; (right) an exposure pushed as far right as possible without clipping the highlights, then pulled down in software, capturing more photons everywhere, especially in the noisy shadows. Brighter-but-not-clipped is the lower-noise capture, when motion and depth of field allow the longer exposure or wider aperture. Raising ISO alone adds no photons, and the comparison is made at the same final brightness.

💡 BIG LESSON (L5.1) — PHOTOGRAPHERS MEASURE LIGHT IN STOPS

A stop is a factor of two. Time steps by two and f-number by $\sqrt{2}$. ISO and a software exposure slider can change rendered brightness by a stop without collecting another photon.

The f-number normalizes the physical opening by focal length. At the same f-number, a longer lens has a proportionally larger entrance pupil. For an extended scene of equal radiance, that compensates for the change in image scale so the on-axis irradiance is approximately unchanged, apart from transmission and close-focus effects. Aperture diameter alone would not give that useful comparison across focal lengths.

The familiar series $f/1.4$, $f/2$, $f/2.8$, $f/4$, $f/5.6$, $f/8$, $f/11$, $f/16$ follows from area. Each full-stop step multiplies the f-number by about the square root of two and therefore halves optical exposure at fixed time. Doubling time compensates for that step. For example, $f/4$ at $1/250$ s and $f/5.6$ at $1/125$ s collect approximately the same light per sensor area, with different motion blur and depth of field. Equal collected light does not make them interchangeable photographs.

Viewpoint, lenses, and focus

Working with light

Developing and editing a photograph

Photographic representation and its limits

The contrast is limited: dynamic range and gamut

OPTIONAL
in the book [↗](#)

Two output constraints are easy to confuse. A print cannot emit the luminance of a sunlit scene, and its ratio between darkest and lightest regions may also be smaller. Scaling all values addresses the first mismatch but not the second. For a sense of scale, a sunlit scene with deep shade can span ten thousand to one or more, while a print under room light spans a few hundred to one at best. Tone mapping, local adjustments and lighting choices affect which distinctions survive.

A bright source can be suggested by a glow or surrounding dark tones even when the display cannot reproduce its luminance. That is a perceptual representation, not an absolute photometric measurement. Likewise, gamut mapping chooses a reproducible color when the target cannot display the original one. [Chapter 18.10, HDR displays and display mapping](#) and [Chapter 4.7, Synthesizing color](#) describe the output limits; [Chapter 5.5, Developing and editing a photograph](#) shows the photographer's choices within them.

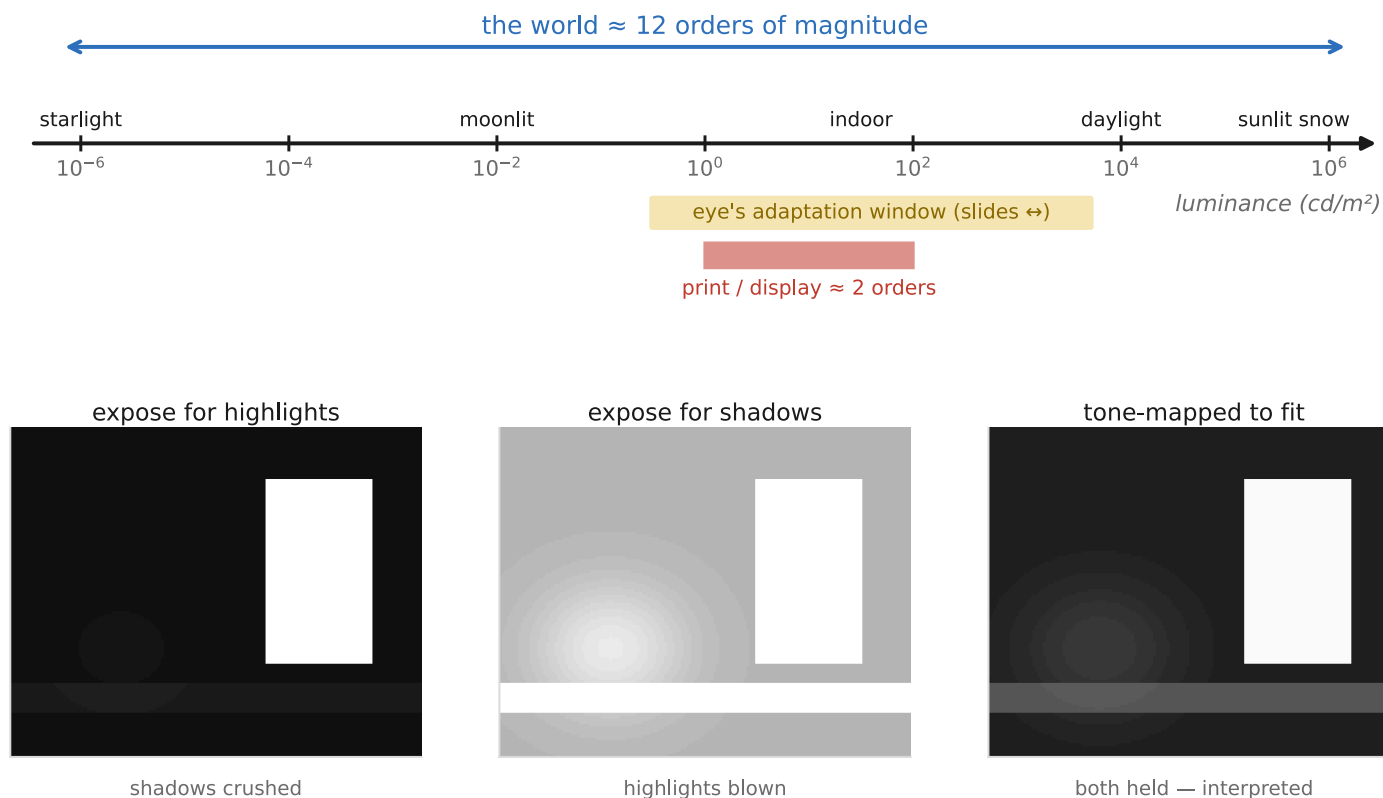


Figure 5.8.16 (reproduced). *The dynamic-range gap. A logarithmic luminance number line from 10^{-6} to 10^6 cd/m^2 shows the real world's $\sim$ 12-order span (starlight to sunlit snow); the eye's adaptation window slides along it (huge over time, far smaller at any instant); and the medium's $\sim$ 2-order print/display band sits as a narrow stripe. Beside it, one high-dynamic-range scene rendered three ways, exposed for the highlights (shadows crushed), exposed for the shadows (highlights blown), and tone-mapped to fit, annotated with the historical compensations: Brunelleschi's silver sky and an added flare/glow standing in for a brightness no print can emit.*

The history of depiction makes the intensity mismatch tangible. The lecture contrasts a painted sky with Brunelleschi's use of a reflecting silver surface for the sky in his perspective demonstration. A pigment and a reflection have different access to the surrounding illumination. Neither comparison can be reduced to an image's normalized pixel range.

The Zone System coordinates exposure, negative development and printing around planned tonal placements [Adams, *The Negative*](#). Dodging and burning modify those placements locally. Characteristic curves compress the ends of the range, while perceptual tone-reproduction methods can also model visibility, glare or adaptation [Ferwerda et al. 1996](#). These are distinct responses to limited output, not interchangeable names for a gamma curve. Retinex provides a related account of appearance through spatial relationships [Land & McCann 1971](#), but it should not be read as proof that the eye simultaneously measures the entire luminance span across all environments.

The darkroom: tone, dodging, and burning

OPTIONAL
in the book [↗](#)

Two prints from one negative can give different weight to a sky or a face. Dodging and burning change local emphasis, and a tone curve changes contrast relationships. The negative-to-print examples make that

interpretation visible. For factual uses, keep track of whether the operation merely changes visibility or hides a relevant feature; the name of the tool does not settle the question.

Same negative, two different photographs — which is the “objective” one?

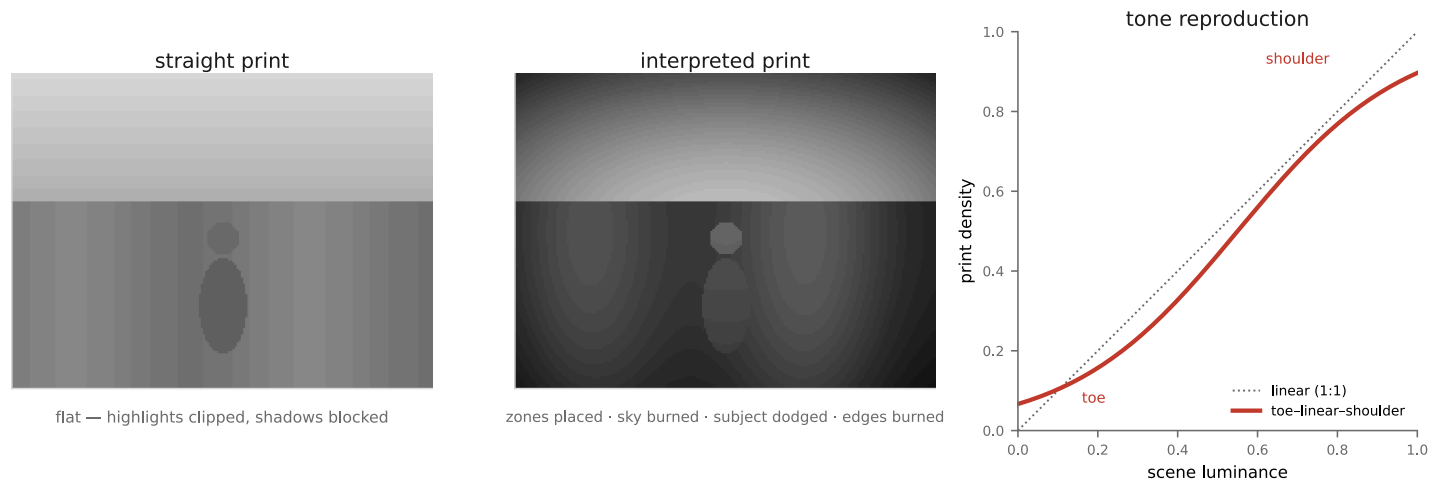


Figure 5.8.22 (reproduced). *The darkroom as authorship. One negative rendered two ways: a “straight” print (flat, the highlights clipped, the shadows blocked) and a fully interpreted print (zones placed, the sky burned down, the subject dodged up, the edges darkened), with a toe–linear–shoulder tone curve annotated alongside. Same latent image, two different photographs. Which is the “objective” one?*

In a photographic print, dodging holds back exposure to make a region lighter; burning adds exposure to make it darker. The printer can use those controls to separate a face from its background, give a sky more weight, or reduce a distracting bright border. These are local decisions even when the negative is unchanged. Compare two prints of the same negative by following where the eye goes first and which details become legible, rather than asking only whether the global brightness matches. The operation changes the interpretation through visibility and emphasis (Figure 5.8.22). The printers of the twentieth century treated this as the work itself: W. Eugene Smith reportedly spent days printing a single photograph of Albert Schweitzer, and Ansel Adams burned the sky of *Clearing Winter Storm* heavily to give it weight. The same negative yields as many prints as there are printers and intentions.

Tone point operations

High dynamic range capture

OPTIONAL
in the book ➤

In some scenes the range of luminance is too high to be captured, let alone displayed, in a single picture: a sunlit window and the dim room around it, a street at night with its lamps, a landscape with the sun in the frame. The brightest parts that matter can be thousands of times brighter than the darkest ones, while a sensor records perhaps twelve to fourteen stops between clipping and noise, and a screen or a print shows fewer still. Expose for the highlights and the shadows sink into noise; expose for the shadows and the highlights clip to white. This part is

about capturing such scenes anyway, by recording more than one exposure, or more than one kind of measurement, and combining them; the next part, [Part 18, Tone mapping and display](#), is about showing the result.

The dynamic range of a sensor is the ratio between the largest signal it records before clipping and the smallest it can tell apart from noise, usually counted in stops (log base 2); [Chapter 2.8, Noise, signal-to-noise ratio and dynamic range](#) defines it and shows how noise and ISO set it. A display's range is defined the same way, between its brightest white and its darkest distinguishable black. High dynamic range has no quantitative definition. A scene, an image or a display is called HDR when its range exceeds what a standard sensor can record in one exposure or a standard display can show, so the threshold moves as the technology improves.

Modern sensors can deliver excellent dynamic range, especially with large sensors at low ISO, and single-capture HDR is often a reality. However, in many scenarios, a single snapshot is not sufficient.

HDR is one of the paradigmatic examples of a central idea of computational photography: when the technology cannot record what we want in one picture, capture several pictures that each record part of it, and combine them computationally. The same idea extends depth of field by merging photographs focused at different distances ([Chapter 23.3, Focal stacks and depth of field extension](#)), extends the field of view by stitching photographs aimed in different directions ([Part 20, Field of view extension and panorama stitching](#)), and reduces noise by averaging several frames of the same scene ([Chapter 13.2, Denoising by averaging](#)), or recovers resolution from their sub-pixel shifts ([Chapter 14.2, Multi-frame super-resolution](#)). HDR is also one of the first techniques to which the name computational photography was attached: Steve Mann, who coined the term in the mid-1990s, used it for work that included combining differently exposed pictures to extend dynamic range [Mann & Picard 1995](#).

The idea of combining multiple exposures for dynamic range expansion is almost as old as photography. In the 1850s, Gustave Le Gray combined separate negatives for sea and sky to make seascapes that photographic materials struggled to record in one exposure. The Getty's examples even show the same sky negative used with different foregrounds (Getty exhibition). Combination printing was therefore both a response to capture limitations and a compositional choice. Modern HDR merging shares the idea of combining complementary exposures, while adding the requirement that registered measurements describe a common scene. [Chapter 18.9, High dynamic range in traditional photography](#) follows that craft further: lighting to reduce the scene's range, graduated filters, the Zone System, and dodging and burning under the enlarger.

[Figure 17.0.1](#) shows two of his prints taken from the same spot at Sète, each printed from two glass negatives, under two different skies.

Click either print to enlarge it.

One viewpoint, two skies: each print made from two glass negatives



Mediterranean Sea, Sète, 1857



Mediterranean with Mount
Agde, 1857

Gustave Le Gray. The Metropolitan Museum of Art, CC0.

Figure 17.0.1 ([reproduced](#)). Gustave Le Gray, *Mediterranean Sea, Sète* (left) and *Mediterranean with Mount Agde* (right), both 1857, albumen prints. Both show the same view, with Mount Agde on the horizon and the jetty on the right, and the museum catalogs each as printed from two glass negatives, one for the sea and one for the sky. The skies are clearly different. The seas differ too, sunlit in one and dark in the other, so this pair does not by itself show a negative reused; the Victoria and Albert Museum reports that prints of this view in other collections share one sea negative under several different skies. (The Metropolitan Museum of Art, CC0.)

HDR capture estimates a wider range of scene intensities than one exposure can usefully record; tone mapping renders that range for a particular display or print. Calibration and merging should preserve measurement relationships, while rendering chooses brightness and contrast for the viewer, which is why they are two parts. This one ends with a radiance map and the knowledge of where it is trustworthy. [Part 18, Tone mapping and display](#) takes it from there, along with the display encodings and gain maps that determine how the result survives distribution.

The chapters follow the capture pipeline. [Chapter 17.1, Dynamic range and multiple-exposure capture](#) sets out the gap between the scene, the sensor and the display, then how to capture a bracket: which exposure parameter to vary, what each choice costs, and how some sensors combine exposures on the chip. [Chapter 17.2, Camera response and radiance calibration](#) recovers the camera's response so that differently exposed frames can be compared in linear units, a step raw files make nearly free. [Chapter 17.3, Merging exposures](#) combines the calibrated frames into one radiance estimate, weighting each frame by how well it measures each pixel, and [Chapter 17.4, Ghosts and motion](#) deals with whatever moved between frames. [Chapter 17.5, Single-shot range and inverse tone mapping](#) covers sensors that deliver extra range in one frame and methods that estimate the missing range from a single image. [Chapter 17.6, HDR+ and burst imaging on phones](#) protects the highlights with a burst of short, identical exposures and averages them to recover the shadows. [Chapter 17.7, Encoding HDR images](#) closes the part with the file formats that store a radiance map and deliver an HDR photograph. Rendering the result, including exposure

fusion, which skips the radiance map, and the darkroom craft that met the same problem before digital sensors, is the subject of [Part 18, Tone mapping and display](#).

Dynamic range and multiple-exposure capture

[in the book](#) ➤

Problem set

PS4 merges an exposure bracket into a radiance map and tone-maps it. → [Chapter 1.5, Problem sets](#) (appendix).

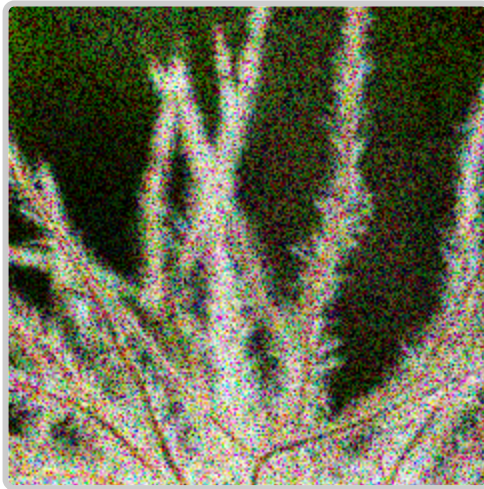
Dynamic range is the ratio between the largest and smallest signal a sensor records. At the bright end, charge capacity or the readout chain limits the recorded signal. At the dark end, noise sets a gradual loss of precision rather than a hard black threshold. The usable dynamic range depends on the minimum signal-to-noise ratio (SNR) accepted for the task ([Chapter 2.8, Noise, signal-to-noise ratio and dynamic range](#)). That chapter develops both ends: full-well capacity at the top (the maximum number of electrons a pixel can hold), read noise and photon noise at the bottom. The engineering definition places the floor where the signal equals the noise (0 dB), but that is too low for photographic purposes, where 30 dB is considered close to perfect and even 15 dB is severely noisy. [Figure 17.1.1](#) shows examples. Like depth of field, quantitative dynamic range is a question of tolerance, to noise in this case. Recall that bright scene values that saturate the photoreceptor may still be read below the maximum because of electronic noise. The ability to capture high dynamic range can further be limited by the veiling glare inside the lens, which lifts the darkest tones before the sensor sees them.

Click any picture to enlarge it. SNR in dB is $20 \log_{10}(\text{signal} / \text{noise})$; 20 dB is a noise of one tenth of the signal.

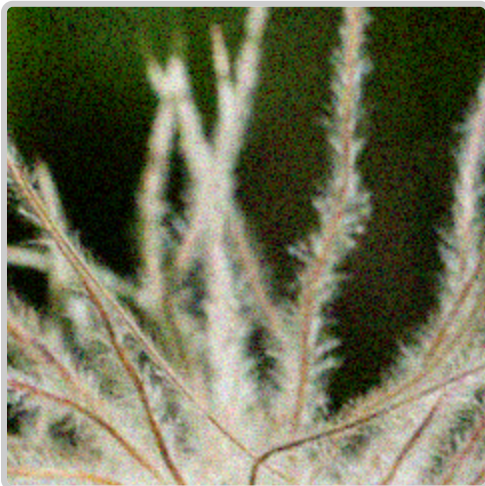
A 240 × 240 crop at 1:1, at six signal-to-noise ratios (stated at an 18% gray)



no added noise



0 dB: SNR 1, the engineering floor



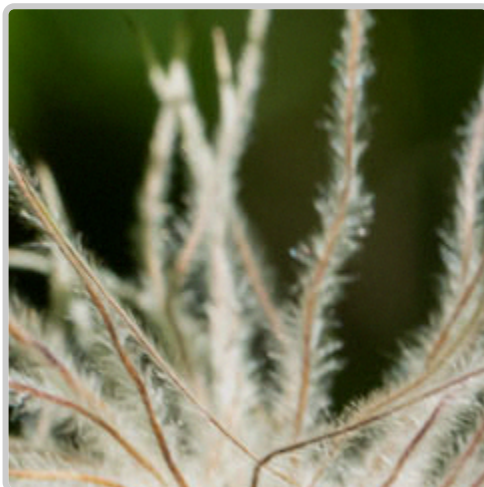
10 dB



15 dB: severely noisy



20 dB



30 dB: close to clean



40 dB

Noise added in linear light, growing with the square root of the signal. Photograph © Frédo Durand.

Figure 17.1.1 (reproduced). What a signal-to-noise ratio looks like: a 1:1 crop of a photograph, one image pixel per screen pixel, with noise at 0, 10, 15, 20, 30 and 40 dB, where the SNR in decibels is $20 \log_{10}$ of the signal over the noise, so 20 dB is a noise of one tenth of the signal. The noise is added in linear light and grows with the square root of the signal, like photon noise, and the stated SNR is that of an 18% gray: the bright seed head is cleaner than the figure says and the dark background noisier. At 0 dB, the engineering floor of dynamic range, the picture is barely usable; at 15 dB it is severely noisy; at 30 dB the noise is hard to see.

When the dynamic range of a scene exceeds that of the sensor, we capture multiple exposures to span the range. We may need to register these images in space (similar to denoising). [Chapter 13.2, Denoising by averaging](#) aligns and averages a burst of equal exposures, [Part 12, Matching pixels across space and time](#) covers the registration itself, and [Chapter 17.4, Ghosts and motion](#) deals with what moves between the frames of a bracket. We also need to calibrate the pixel values photometrically to bring them into the same absolute linear range proportional to scene-radiance values $L(x, y)$, which may require the calibration of the film or digital camera response curve (the beauty curve in particular, [Why scene-linear looks dull](#) in [Part 8, Basic image processing and ISP](#); [Chapter 17.2, Camera response and radiance calibration](#) recovers it). Motion, clipping in every frame, or inadequate shadow exposure can leave parts of that estimate uncertain.

💡 BIG LESSON (L2.46)

A common sensor dynamic-range measure divides the saturation signal by the read-noise standard deviation, in consistent units. A stricter quality threshold gives a smaller usable range. Bracketing extends coverage when its exposures provide reliable, registered measurements of the same scene.

The gap between the scene, the sensor and the screen

Distinguish the range within one scene from the much larger range encountered across daylight, interiors and night. Likewise, visual adaptation across time is different from simultaneous contrast sensitivity. A sensor's capture range and a display's visible contrast also depend on noise criteria, viewing conditions and hardware; one fixed ratio does not describe all of them.

With that caveat, some orders of magnitude help. A view from a dim interior through a window onto a sunlit street can span around 100,000:1, about 17 stops, between the brightest diffuse surface and the darkest shadow that still shows detail, before counting the sun or specular highlights.

Following the light from its source shows where such a number comes from. The disk of the sun has a luminance of about 1.6×10^9 cd/m², far beyond anything else in a photograph. A clear blue sky is around 5,000 to 10,000 cd/m², some two hundred thousand times dimmer. A white wall in direct sun receives about 100,000 lux and, reflecting most of it, has a luminance around 25,000 cd/m². In the shade, a surface sees the whole sky but not the sun: a uniform sky of luminance L over a full hemisphere delivers an illuminance of πL , about 15,000 lux, so a mid-gray surface there is around 2,500 cd/m², ten times dimmer than the sunlit wall. Indoors, a surface near a window sees only the part of the sky the window frames. A one-square-meter window three meters away subtends about 0.1 steradian, so the illuminance drops to a few hundred lux, some forty times less than in the open shade, and a mid-gray surface is around 60 cd/m². Many points in the room do not see the window at all and are lit only by light bounced off the walls and ceiling, a few lux to a few tens of lux; a dark fabric there, reflecting five percent, is around 0.2 cd/m². From the sunlit wall outside to that fabric is about 25,000 / 0.2, or 125,000:1: the 100,000:1 above, reached by a chain of ordinary factors, each of ten or forty, and without the sun or the sky ever appearing in the frame.

No single exposure of such a scene holds both ends: [Figure 2.8.14](#) develops one scene at three exposures.

The dynamic-range problem: one high-contrast scene, no single exposure that holds both ends



Figure 2.8.14 ([reproduced](#)). *The dynamic-range problem. One high-contrast scene (a sunset over water) developed at three exposures: expose for the highlights and the foreground is crushed to black; expose for the shadows and the sky blows out to white; the compromise loses both ends. No single exposure holds the full range: the motivation for high-dynamic-range merging of multiple exposures.*

The eye covers a far wider range, but not at once ([Chapter 3.3, Adaptation and constancy](#)). Across its states of adaptation, vision works from about a millionth of a candela per square meter, the threshold of fully dark-adapted vision, to sunlit snow at 10^4 to 10^5 cd/m², some ten or eleven orders of magnitude, well over thirty stops. At any one operating point it distinguishes far less, roughly four orders of magnitude, and it sees a high-contrast scene

well because it moves: each fixation adapts to the part of the scene it lands on. A camera has to record the whole scene in one exposure, which is why 100,000:1 is a problem for it and not for the eye.

A current large-sensor camera at base ISO (the lowest ISO, i.e. the lowest gain) records roughly 12 to 14 stops between clipping and a noise floor at a signal-to-noise ratio of one, and a phone sensor fewer. That is the engineering definition, with the floor at 0 dB, where the signal equals the noise, which nobody would call a usable shadow. Raising the floor costs range quickly. With photon noise and read noise, the dimmest signal reaching an SNR of q is $N_{\min} = (q^2 + \sqrt{q^4 + 4q^2\sigma_r^2})/2$ electrons ([Chapter 2.8, Noise, signal-to-noise ratio and dynamic range](#)). For a full-frame pixel with a full well of 50,000 electrons and 3 electrons of read noise, the range is 13.8 stops at 0 dB, 11.6 stops at 10 dB (an SNR of about 3) and 8.9 stops at 20 dB (an SNR of 10). For a phone pixel with 6,000 electrons and 1.5 electrons of read noise, the same three floors give 11.5, 9.0 and 5.9 stops. At 20 dB the floor is set almost entirely by photon noise, about 100 electrons whatever the read noise, so the range is little more than the full well divided by 100, and what buys clean shadows is a deeper well, which means a larger pixel. [Figure 17.1.2](#) shows how it falls with ISO for ten cameras.

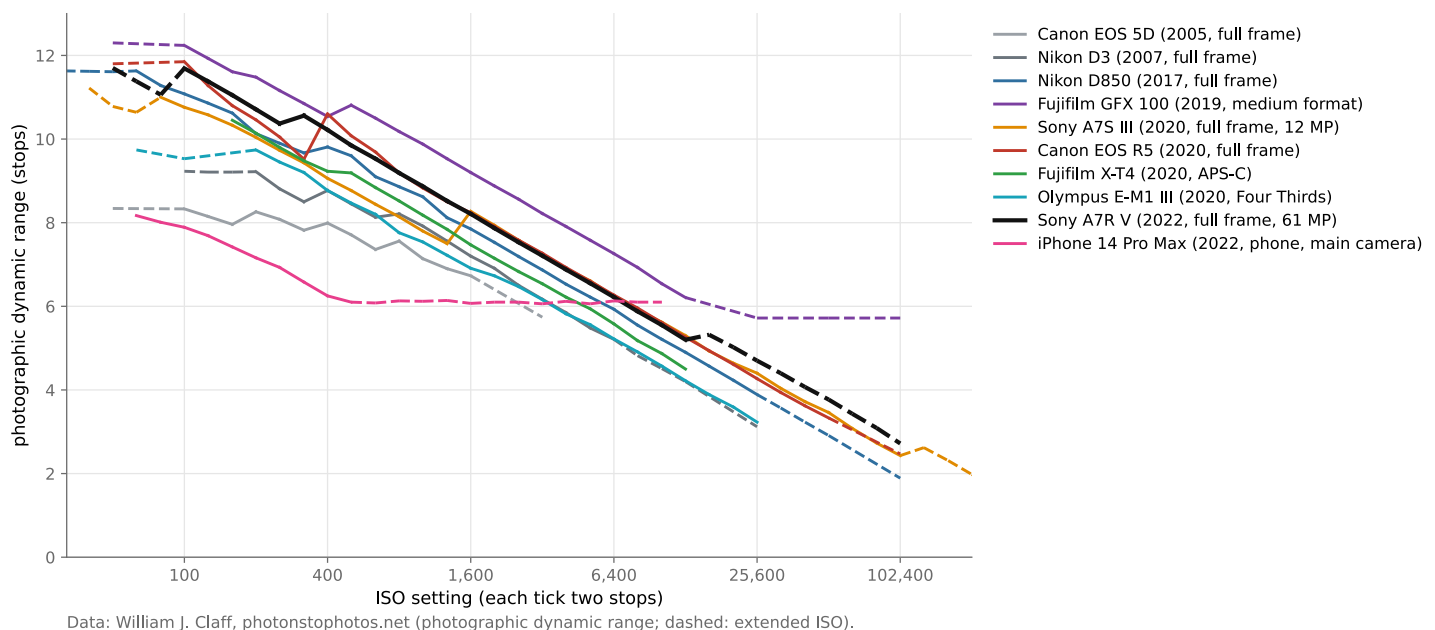


Figure 17.1.2 (reproduced). Dynamic range against ISO for ten cameras, from a 2005 full-frame DSLR to a 2022 phone, measured by William Claff as photographic dynamic range: from clipping down to the level where noise becomes objectionable in the whole image seen at a fixed size, rather than to a signal-to-noise ratio of one. That lower floor is why the values sit below the 12 to 14 stops of the engineering definition, and the normalization to the whole image is what lets sensors of different sizes be compared. Each doubling of ISO costs close to a stop once read noise stops dominating, since it halves the light behind each value. The medium-format Fujifilm GFX 100 leads by about half a stop at base ISO; the 2005 Canon EOS 5D starts more than three stops lower than current full-frame cameras; the steps near ISO 400 or 1600 on several recent cameras are where the sensor switches to a higher conversion gain. The phone's curve flattens above about ISO 400, probably because its raw files are already processed. Dashed: ISO settings outside the camera's native range. Data: William J. Claff, photonstophotos.net.

The most widely used sensor measurements are DxOMark's (DxOMark camera sensor database). They report dynamic range in stops (EV), from saturation down to the level where the signal equals the noise, as a function of ISO, in two modes. Screen mode measures each pixel as it is. Print mode normalizes every camera to the same output size, 8 megapixels, as if the image were downsampled for a print of fixed size; averaging pixels lowers the

noise, so a sensor with more pixels gains about half a stop in Print mode for every doubling of its pixel count, at DxOMark's floor where signal equals noise; at a floor high enough that photon noise dominates, the gain approaches a full stop per doubling. Screen mode compares pixels, Print mode compares pictures, and the two can rank cameras differently. The best full-frame sensors reach close to 15 stops in Print mode at base ISO, and above the lowest ISOs a sensor loses close to a stop of range for each doubling of ISO. Bill Claff's Photons to Photos charts use a stricter floor, set by what is visible in a print of a given size, and so report smaller numbers for the same cameras.

[Figure 17.1.3](#) computes such curves from the noise model alone, for four typical sensors, and leaves the floor and the normalization to the reader.

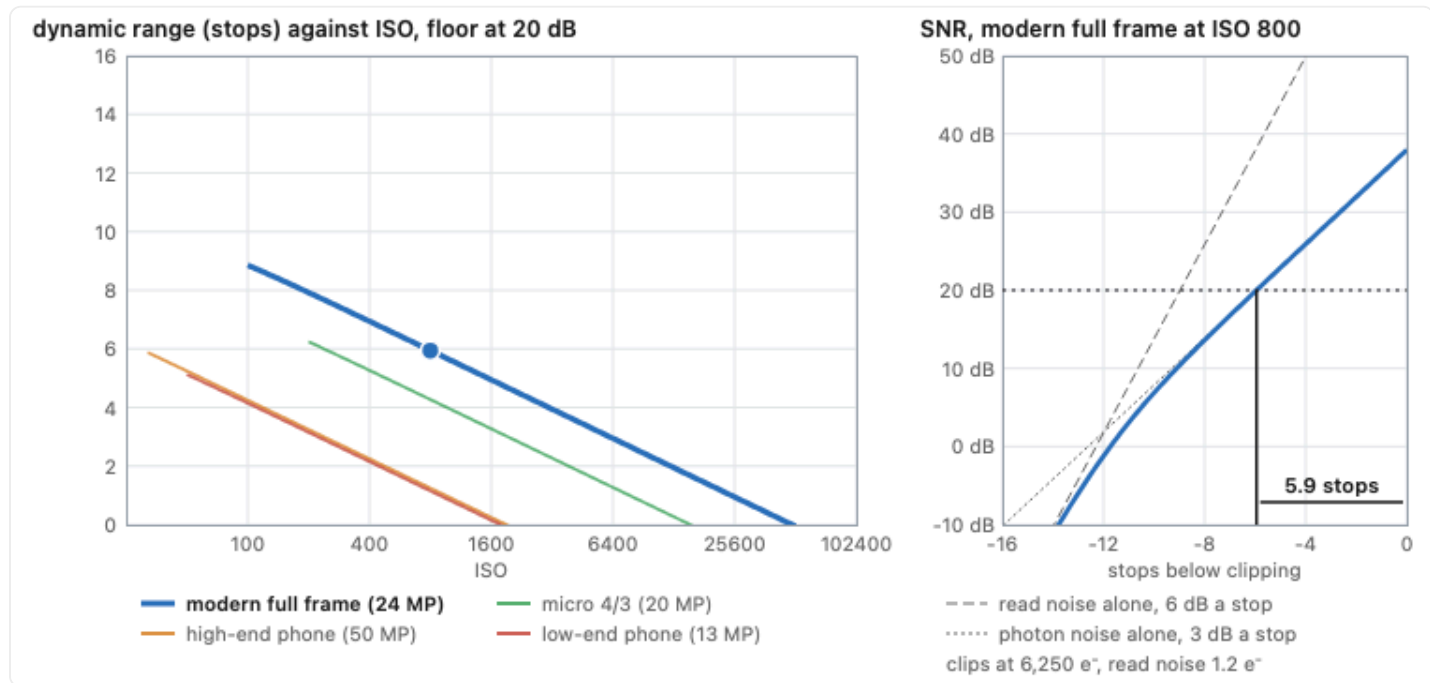


Figure 17.1.3 (reproduced). *Dynamic range against ISO from the noise model, for four typical sensors rather than particular cameras: a modern full frame (24 MP), micro 4/3 (20 MP), a high-end phone (50 MP on a 1/1.3-inch sensor) and a low-end phone (13 MP on 1/3 inch). All four are given the same sensitivity per unit area: an ISO fixes the exposure at clipping, so at a given ISO every sensor clips at the same number of electrons per square micron, and a pixel's electrons at its base ISO follow from its area (50,000 for the full frame at ISO 100, 7,800 for micro 4/3 at ISO 200, 6,000 for the high-end phone at ISO 32, 3,700 for the low-end phone at ISO 50). At the same ISO the full frame therefore leads micro 4/3 by its area, a factor of 3.8 or about two stops, once photon noise sets the floor. A pixel's noise is photon noise plus Gaussian read noise, $\sqrt{N + \sigma_r^2}$ electrons. Each doubling of ISO halves the electrons behind the clipping level, and lowers the read noise referred to the input until only the part before the gain is left, after which it costs a full stop. The range runs from clipping down to $N_{\min}$, the signal whose SNR equals the chosen floor. Left: the four sensors. Right: SNR against signal for the chosen sensor and ISO, following read noise at 6 dB a stop in the shadows and photon noise at 3 dB a stop above. Normalized to an 8-megapixel picture, as in DxOMark's Print mode, pooling pixels gains half a stop per doubling of the pixel count when read noise sets the floor and close to a full stop when photon noise does, as at 20 dB. Per pixel at base ISO, the full-frame and high-end phone sensors give the values in the text. Claff's measurements in [Figure 17.1.2](#) pool over a print's circle of confusion, more pixels than an 8-megapixel picture, so they sit higher at the same floor; and no conversion-gain switch is modeled here, so the measured curves' steps are absent.*

The ranges in this chapter move between stops, decibels and plain ratios. The converter of [Chapter 2.8, Noise, signal-to-noise ratio and dynamic range](#) is reproduced here for going from one to another ([Figure 2.8.5](#)).

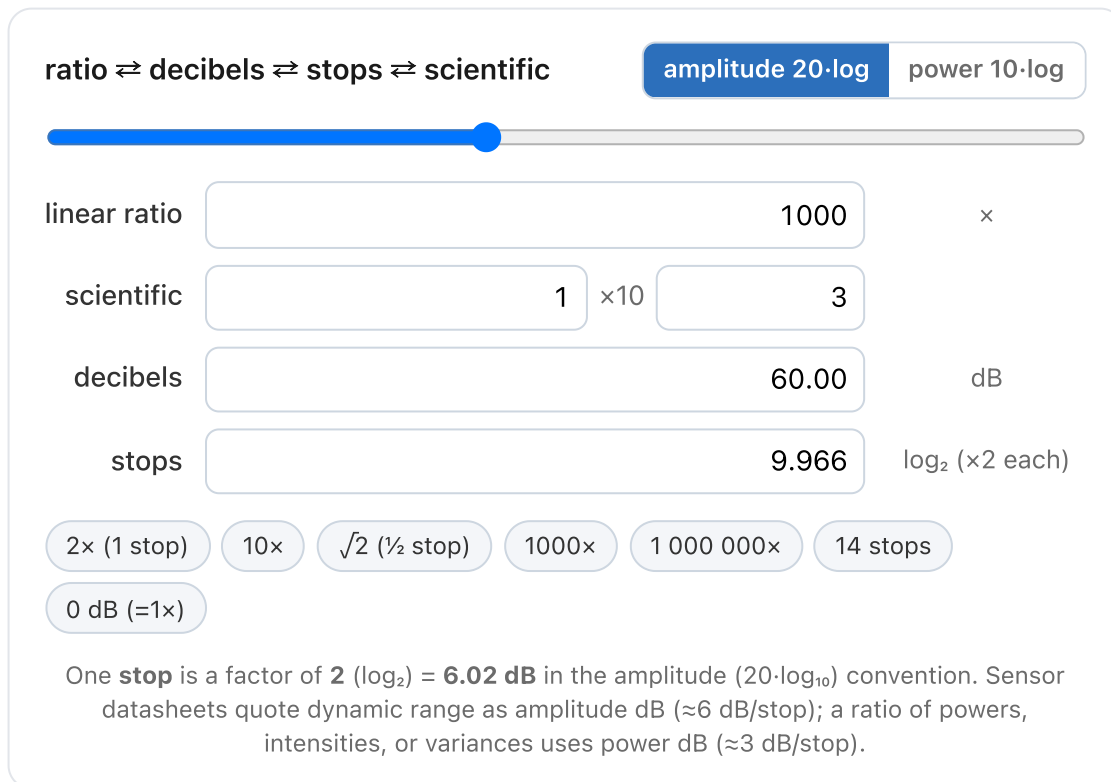


Figure 2.8.5 ([reproduced](#)). A units converter for ratios (interactive). Decibels, photographic stops ($\log_2$), a plain linear ratio, and scientific notation are four ways of writing one multiplicative factor. Type into any field, or drag the log-progression slider, and the others update to match; a toggle switches the decibel convention between amplitude ($20 \log_{10}$, about 6 dB per stop, the convention sensor datasheets use for dynamic range) and power or energy ($10 \log_{10}$, about 3 dB per stop, for ratios of intensities or variances). Preset chips jump to the common landmarks: one stop, ten-fold, a million-to-one, a full-frame sensor's roughly fourteen stops.

A standard-range display in good conditions (dim room, no glare) shows perhaps 1,000:1, about 10 stops, and much less in daylight, when reflected room light raises its black. Video projectors in a classroom can be as low as 10:1. A glossy print reflects roughly 100:1 to 250:1 if you avoid the reflection angle. Each number depends on the criterion used to define its dark end, which is why they should be compared with care.

Use a simplified linear observation model after black-level subtraction. The exposure factor k_i includes shutter time, aperture transmission and calibrated gain; it converts scene radiance to the chosen signal units:

$$I_i(x, y) = \text{clip}(k_i L(x, y) + n). \quad (7)$$

Each frame scales the same radiance, adds noise and clips at its recording limits. This approximation omits spatially varying optics, motion during exposure and nonlinear processing. Gain changes the signal units and noise model; it does not collect more photons.

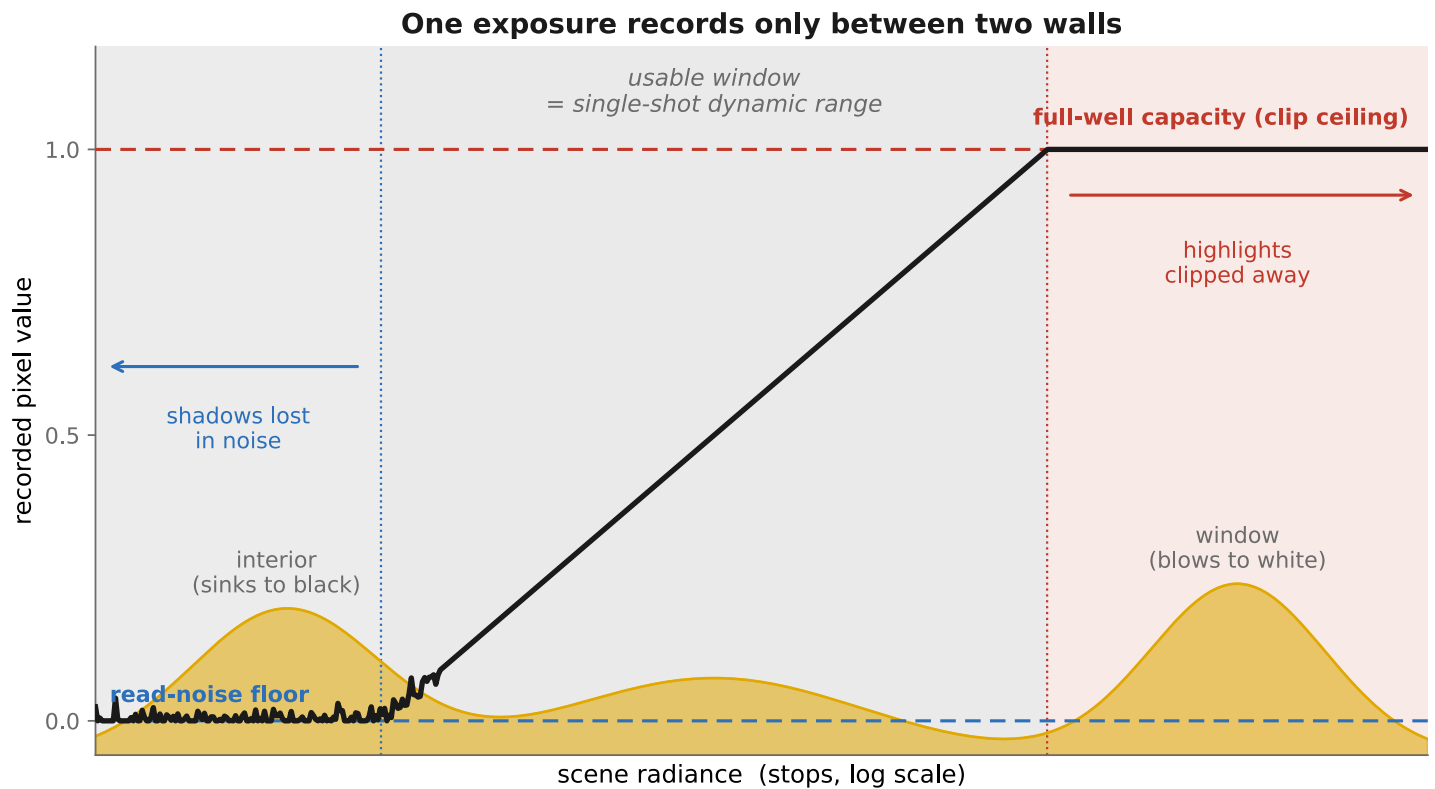


Figure 17.1.4 (reproduced). One scene, one exposure, and the two walls of the single-shot dynamic range. A high-contrast scene (interior plus bright window) is shown with its radiance histogram laid against the photosite's response: highlights past the full-well capacity clip to a flat ceiling (the window blows to white), and shadows beneath the read-noise floor drown in noise (the interior sinks to black). The usable window between the two walls (the single-shot dynamic range) is shaded; the scene's true range overflows it on both ends. This is [L2.46](#) [show](#) drawn for one capture: dynamic range is the ratio of that ceiling to that floor.

Bracketing, and which parameter to vary

[in the book](#) [↗](#)

The principle of multiple-exposure high-dynamic-range photography is simple. Choose exposure factors with overlapping usable ranges, take photos, then merge into a floating-point radiance estimate. Its scale is usually relative unless an absolute radiometric calibration is supplied. Formats such as OpenEXR can preserve scene-linear values for later rendering ([Chapter 8.11, File formats and compression](#)).

On a logarithmic radiance axis, changing exposure slides each frame's useful interval. One stop moves it by a factor of two. Overlap matters because neighboring exposures need common usable signal for calibration and a smooth change in which measurement is most precise. A gap leaves intensities that are clipped in one frame and poorly measured in the next. The merged floating-point values can exceed one; that is normal for a scene-linear representation whose scale is relative.

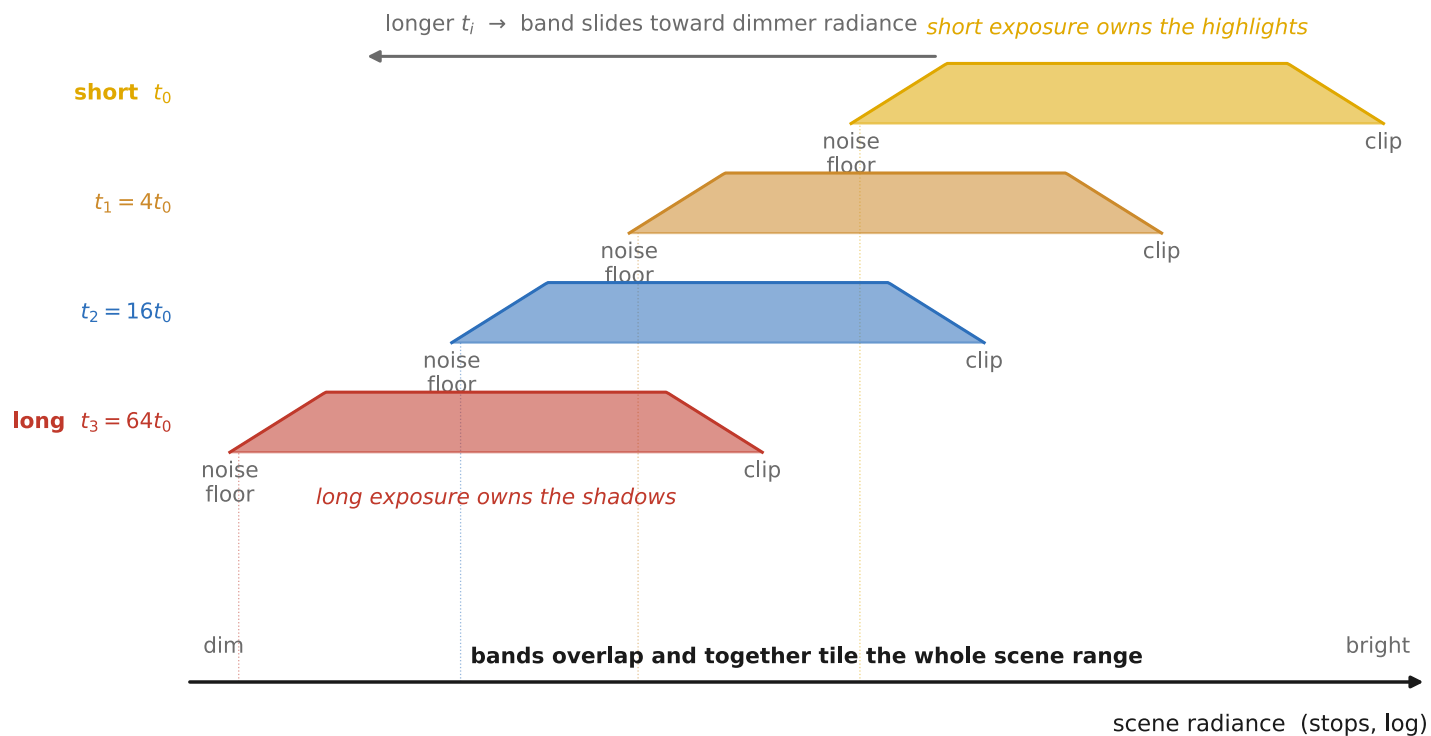


Figure 17.1.5 ([reproduced](#)). A bracket tiles the radiance axis. The horizontal axis is scene radiance (log scale, in stops); each exposure I_i in the bracket is drawn as a usable band (above its noise floor, below its clip ceiling), and increasing the shutter time t_i slides the band leftward (toward dimmer radiance). A short exposure owns the highlights, a long one owns the shadows, and the bands overlap so that together they cover the full scene range. The merge keeps the reliable part of each band and stitches them into one radiance map spanning the whole axis.

A bracket: no single exposure holds both the dark interior and the bright window
— each frame captures a different band of scene radiance (→ merge)



Figure 17.1.6 (reproduced). The same idea on a real bracket: a dim interior with a blown-out window onto Boston harbor, a textbook high-dynamic-range scene. Four frames of a longer sequence, from the shortest exposure (the window and skyline are correctly exposed, but the room is lost in noise-black) to the longest (the room opens up, but the window clips to featureless white). No single frame holds both ends; each captures a different band of scene radiance, exactly the bands the previous figure drew, which is why the scene must be bracketed and merged.

The exposure factor k_i can change in several ways, but each control changes something else as well. A useful capture sequence considers those side effects before choosing how widely to space its exposures.

For a conventional bracket, keep aperture, focus and gain fixed and vary shutter duration. This preserves the lens's focus blur and usually provides predictable exposure ratios. It also changes motion blur and capture timing, so the frames are not identical apart from brightness.

Under steady illumination and away from clipping, expected photon count is proportional to exposure duration. A longer exposure improves photon statistics but integrates more subject or camera motion and accumulates more dark current. Flickering light and rolling-shutter timing can violate a simple global exposure ratio.

Changing aperture changes depth of field, diffraction and often vignetting. Changing ISO changes gain and possibly input-referred read noise or saturation headroom, without changing the photons captured at fixed shutter and aperture. A neutral-density filter attenuates light, including flash, but its transmission and color should be calibrated. These controls can be useful in a designed capture sequence; they are not interchangeable with shutter duration.

Set the shortest exposure to preserve important highlights and the longest to reach acceptable shadow SNR without excessive motion blur. Choose intermediate exposures with enough overlap for calibration and merging. Stop adding frames once they provide little useful coverage or noise reduction for their time cost.

A tripod reduces camera motion but does not freeze subjects, foliage or illumination. Handheld brackets need registration; even tripod brackets need consistency checks. Ward's median-threshold bitmap (MTB) aligns differently exposed images using thresholded intensities and excludes values near the median ([Ward 2003 \(MTB\)](#)); [Chapter 17.4, Ghosts and motion](#) shows it at work. Clipping, noise and changing content can still corrupt the comparison.

The exposure triangle

rendering...

Scene light



starlight → full moon → night → dark office → home → overcast → hazy sun → bright sun

Shutter



slower → more light + motion blur

Aperture



wider (small f) → more light + shallow depth of field

ISO



higher → brighter, not cleaner: gain scales noise with signal

Subject Sensor Auto-set (keep exposure constant):

2nd person depth



Subject motion

[copy link](#)

Figure 5.2.6 [\(reproduced\)](#). *Interactive: drive the three clean exposure knobs and watch each side effect appear. Hold the brightness fixed and trade one stop of shutter for a stop of aperture or ISO, and the picture's tone barely moves while its character changes: longer shutter brings motion blur, wider aperture (smaller f-number) shallower depth of field, higher ISO more noise. For bracketing, this is the figure's lesson made tangible: only shutter changes nothing but the exposure (so it is the knob you bracket with), while aperture would smear the focus across frames and ISO would amplify noise. (The fourth fallback knob, an ND filter, is the one this demo cannot show: it darkens the world but is not perfectly neutral, adds a color cast, and forces you to touch the camera.)*

HDR exposure bracketing & merge

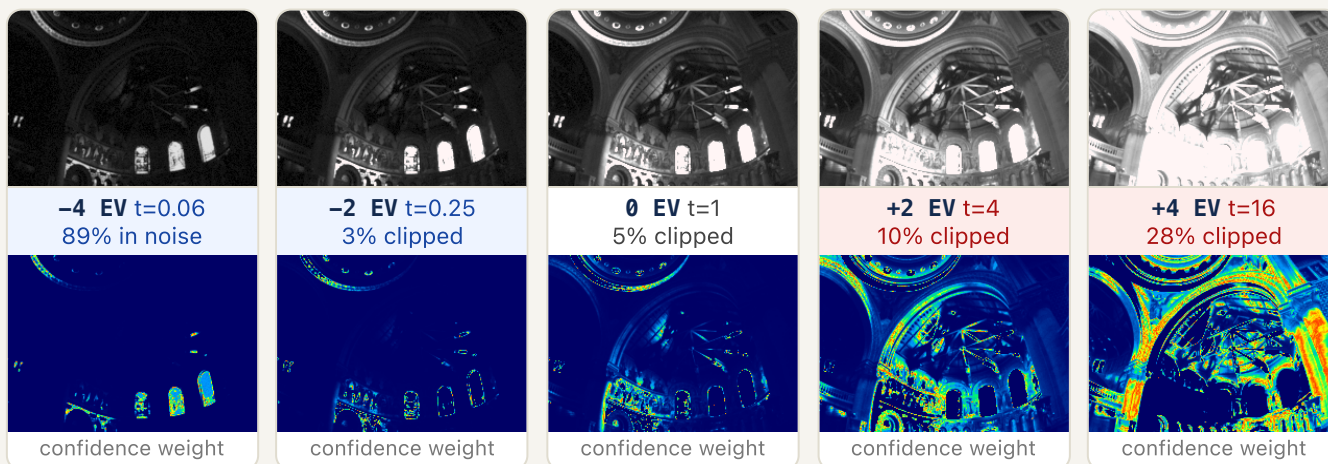
SCENE Memorial church (real) ▾ upload bracket

BRACKET shots 5 spacing 2.0 EV

CAMERA noise 1.50 RANGES IN stops (log₂) ▾

TONE MAP Reinhard (global) ▾ key 0.32

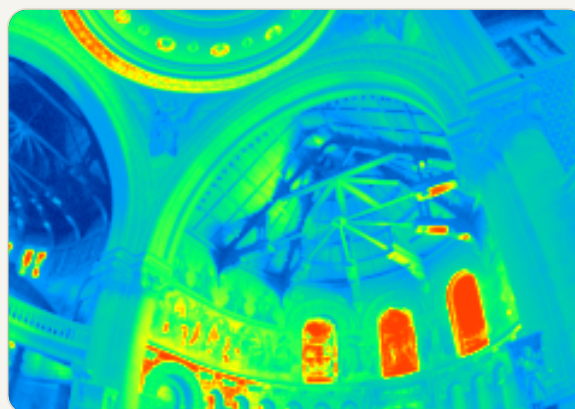
☒ weight maps ☐ MTB align UPLOAD EV STEP stops 1.0



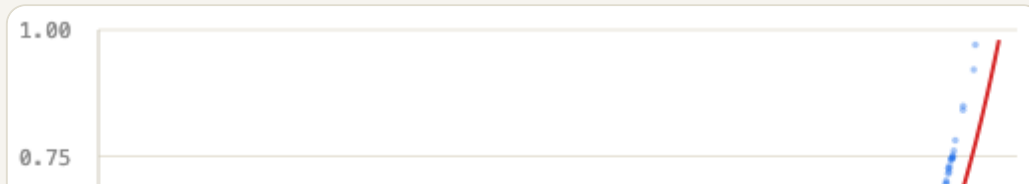
Merged HDR — confidence-weighted radiance, then tone-mapped

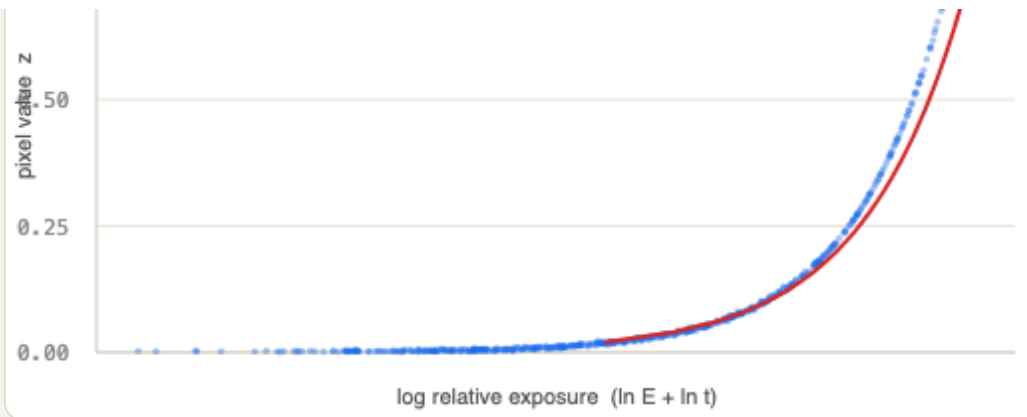


Recovered radiance — log false-color (blue = dark, red = bright)



Camera response curve — sampled pixels (value vs log relative exposure) + recovered $g(z)$





585 samples from 140 scene points across 5 exposures collapse onto one curve. Fitted effective response: $z \propto (\text{relative exposure})^\gamma$ with $\gamma \approx 0.92$ (red curve; our sensor is near-linear in radiance, so γ reflects the display gamma baked into the samples).

SCENE RANGE
13.0 stops

1 EXPOSURE COVERS
5.3 stops

5-SHOT BRACKET COVERS
13.3 stops

MERGED VS TRUE ERROR
0.03 stops

How to read this. The top row is the **bracket**: each shot is $\text{clip}(E \cdot t_i + \text{noise})$ through the camera, shown as it would look. Walk it left $\rightarrow$ right: short (dark) exposures keep the window's structure but the shadows are grainy; long (bright) exposures clean the shadows but the window goes flat white — **clipped**, its radiance lost. Under each shot is its **confidence weight** (the hat function): near-black and near-white pixels get ~ 0 weight, mid-tones get ~ 1 . The **merge** takes, per pixel, the weighted average of pixel_i / t_i over the shots — that division by exposure time is what turns each shot back into a common **radiance** scale, and the weights throw away the clipped/noisy samples. The recovered radiance (false-color) spans the scene's full range; the **tone-map** compresses it to a display. Drop the bracket to **1 shot** and the readout shows the error jump — a single exposure simply cannot hold the range.

Two kinds of scene. The **real HDR** maps are genuine measured radiance — Paul Debevec's classic **memorial church** (~ 13 stops, bright stained glass over a dark nave) and **vine sunset**, each merged from a real exposure bracket and shipped here as a luminance radiance map; the demo re-captures synthetic brackets *from that true radiance* and merges them back. The **synthetic** scenes generate the radiance procedurally, so the recovered radiance can be checked against an exactly-known ground truth. Either way the "merged vs true error" readout compares the recovery against the radiance the brackets were drawn from.

Figure 17.1.7 (reproduced). *Interactive: why you bracket, and how the merge works. Pick a scene, either a real measured radiance map (Paul Debevec's classic memorial church, ~ 13 stops of bright stained glass over a dark nave, or his vine sunset) or a synthetic scene with an exactly-known true radiance, and the demo captures it as an EV bracket (sliders for shot count and spacing); each frame shows real highlight clipping and shadow noise, so no single exposure holds the whole scene. The demo then merges them, a confidence-weighted per-pixel average of pixel_i / t_i that drops clipped and noisy samples (the Debevec/Mertens hat weight [Debevec & Malik 1997](#)), into an HDR radiance map, and a tone-curve slider maps it to display. Watch the per-shot weight maps, recover a blown highlight from the short exposure and a noisy shadow from the long one, and narrow the bracket to see the error return. (Real HDR maps courtesy Paul Debevec.)*

Stops make the sequence easy to reason about: $1/1000$, $1/250$ and $1/60$ second form an approximately two-stop progression. The same exposure increase could be obtained by opening the aperture two stops, but then depth of field and the blur operator change. Flash gives another reason to distinguish controls. Within ordinary flash-sync operation, changing shutter duration mainly changes the continuous-light contribution; a neutral-density filter attenuates both ambient and flash light. The desired bracket must therefore account for which source the control actually changes.

Doing it in the sensor instead

OPTIONAL
in the book [↗](#)

Sensors can combine measurements with different integration intervals, conversion gains or sensitivities into one delivered HDR frame. A single output frame does not necessarily mean all its measurements cover the same time interval. Check the sensor's acquisition and readout design ([Part 34, Computational sensors](#)).

Time multiplexing combines shorter and longer integrations. Digital-overlap HDR interleaves their readout; it should not be described as simply reading one unchanged exposure twice. Different integration windows can produce motion or illumination inconsistencies even when packaged as one output frame.

Gain multiplexing can read charge collected during a common integration through paths with different conversion gains. One path favors low input-referred read noise and another highlight capacity. This avoids the inter-exposure mismatch of a bracket when the integrations are shared. Ordinary selectable dual-conversion-gain operation does not by itself imply that both gains are delivered simultaneously.

Spatial multiplexing assigns different sensitivities or collection areas to neighboring sites or subpixels. Nayar and Mitsunaga's spatially varying exposure array reconstructs a wide-range image from such a pattern ([Nayar & Mitsunaga 2000](#)). The sampling and noise tradeoffs depend on the layout. Overflow-capacitor designs instead add charge storage; logarithmic or companding responses allocate output precision nonuniformly across intensity. None removes the need for calibration or a noise model.

Split-pixel sensors, common in automotive cameras, are a concrete spatial design. Each pixel holds a large photodiode and a small one. For the same light the small one collects far fewer electrons, so it saturates much later and records the highlights while the large one records the shadows, and the two are merged into one value per pixel. The motivating scene is a car leaving a tunnel: dark walls and daylight at the exit in one frame, with no time to bracket, plus LED signs and headlights driven by pulses that a short exposure can miss entirely. Because the small diode can integrate over a long window without saturating, it can record those pulses where a short exposure of the large diode would not. The two diodes differ in noise, response and color, so the merge needs its own calibration.

Compare integration timing, spatial sampling, read noise and saturation capacity separately. Shared integration avoids one source of ghosts, but does not eliminate motion blur or rolling-shutter distortion. A software burst adds repeated readout, computation and alignment costs.

On-chip HDR: three ways one sensor captures high dynamic range

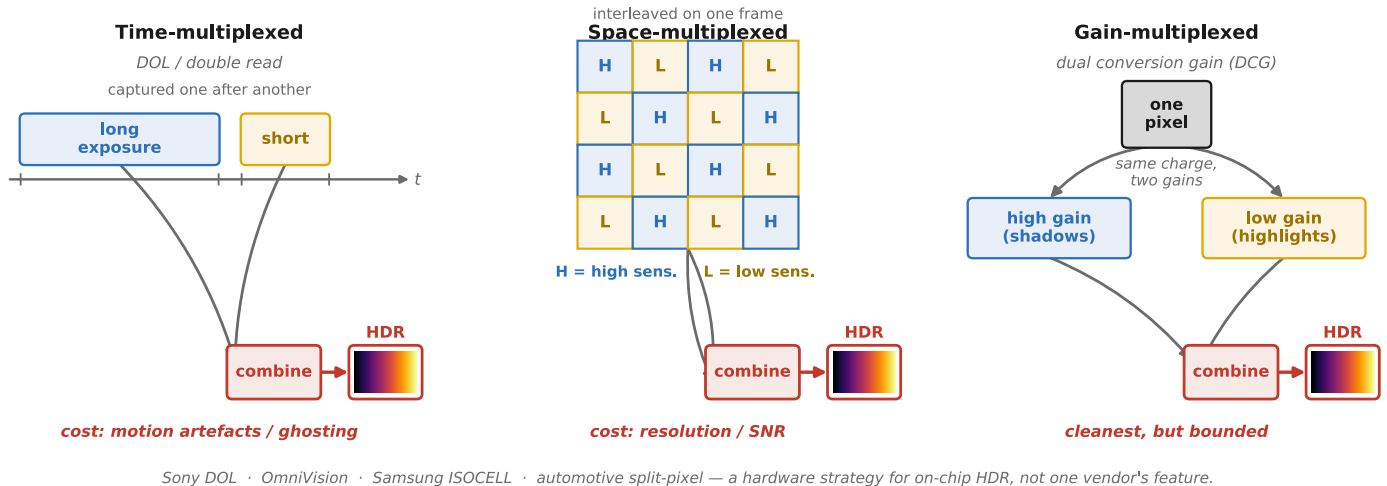


Figure 17.1.8 (reproduced). On-chip ("direct") HDR sensors, organized by where they multiplex the dynamic range. Time-multiplexed (DOL "double read"): long and short reads of the same frame are combined; the schematic shows the two interleaved reads and a ghosted moving object flagging the motion artifact. Space-multiplexed (split-pixel large+small sub-photodiode, and SVE / assorted-pixels exposure mosaic): neighboring sites carry different sensitivities, single-shot but at a resolution/SNR cost. Gain-multiplexed (dual-conversion-gain): the same instant read at two readout gains: cleanest, but its added range is bounded. Each cell labels its cost (motion / resolution+SNR / bounded range), and a fourth panel contrasts the software HDR+ burst [Hasinoff et al. 2016](#): maximal flexibility, paid in compute and alignment.

Two caveats: single-shot range, and where this is used

OPTIONAL
in the book [↗](#)

First test a single raw file at the intended output size. Bracketing helps when it supplies cleaner shadows or unclipped highlights that the single capture lacks. Fill flash and graduated neutral-density filters can also reduce scene contrast at capture; dodging and burning change its rendering afterward.

These choices have photographic precedents. Le Gray combined negatives to retain sea and sky, while graduated neutral-density filters and fill flash change the contrast presented to the camera. Dodging, burning and the Zone System help control how recorded tones become a print; [Chapter 18.9, High dynamic range in traditional photography](#), in the tone-mapping part, treats each of them.

Camera response and radiance calibration

in the book [↗](#)

Merging multiple exposures requires the conversion of pixel values into the same meaningful units proportional to scene radiance. Calibration establishes that relationship. When working with raw files, where the pixel values are linearly related to radiance, only pairwise ratios are needed. In the general case, the non-linear response curve from radiance to pixels must be calibrated. The unknown scene can help calibrate the camera because each scene point

appears at several exposures. We assume the images in a bracket are pixel-aligned and the scene does not contain movement. The model $I_i = f(k_i L)$ assumes a fixed pointwise response shared across frames, where k_i characterizes the exposure for frame i . In the general case f must be calibrated. We assume I_i values have been black-level corrected (recall that some cameras add a constant offset to measure the negative component of readout noise, [Chapter 2.8, Noise, signal-to-noise ratio and dynamic range](#)). Throughout, $I_{p,i}$ is the value of pixel p in frame i , L_p the radiance at that pixel, and $w(I)$ a reliability weight that is zero for clipped values and small near the noise floor.

The linear case: recovering the exposure ratios

[in the book](#) ↗

In the linear case, f is simply a global multiplication factor and the problem boils down to recovering the k_i exposure factors up to a multiplicative factor. But even for linear RAW files, linearity breaks down for dark values, and especially for bright ones. We call a pixel reliable if it is not too dark and not too saturated (between ϵ_0 and $1 - \epsilon_1$ in normalized pixel values). For exposure ratio estimation, the values of the thresholds are not critical and anything significantly above the read-noise level will do.

For linear, black-level-corrected reliable observations, $I_{p,i} = k_i L_p$ in the noiseless unsaturated model. At a pixel that is reliable in two images, the radiance cancels:

$$\frac{I_{p,i}}{I_{p,j}} = \frac{k_i L_p}{k_j L_p} = \frac{k_i}{k_j}, \quad (8)$$

which is the same number at every good pixel, if linearity holds. Estimate it robustly with the median over the pixels where both frames are reliable,

$$\frac{k_i}{k_j} = \text{median}_{p: w(I_{p,i}) > 0, w(I_{p,j}) > 0} \left(\frac{I_{p,i}}{I_{p,j}} \right). \quad (9)$$

Pairwise ratios recover relative exposure factors, but small denominators and noisy correspondences can bias them. Use many reliable pixels and validate residuals; jointly fitting exposure scales can avoid drift from chaining. Metadata supplies a useful initial estimate and can be accurate enough for some cameras. The image only determines relative radiance without an absolute calibration.

To avoid statistics over ratios, which is always dangerous when denominators can be close to zero, you can use a linear regression where you fit the scalar $\frac{k_i}{k_j}$ according to the model

$$I_{p,i} = \frac{k_i}{k_j} I_{p,j} \quad (10)$$

which naturally weights the more reliable evidence from larger pixel values higher. But in practice, the median works fine, especially if you exclude ratios with low denominators.

If pixels exhibit inconsistent values between normalized frames, check black level, response, timing, illumination and registration before merging.

Validate exposure factors

Metadata records nominal capture settings. Check them against unsaturated, registered image regions when radiometric accuracy matters ([Burggraaff et al. 2019](#)). Pixel-based calibration also has failure modes: motion, flickering light, noise and local processing can make a single global ratio inappropriate.

The non-linear case: recovering the curve

OPTIONAL
in the book [↗](#)

In the non-linear case, we need to recover the response curve. Real cameras (and film, and JPEG) apply a non-linear response f , such as a gamma, a tone curve, or a film characteristic, including the beauty curve that gives a camera its look ([Why scene-linear looks dull](#), in [Part 8, Basic image processing and ISP](#)). Then $I_i = f(k_i L)$, and you must recover f (equivalently its log-inverse $g = \ln f^{-1}$) before you can merge. The trick of [Debevec and Malik \(1997\)](#) is that a pixel imaged at *several* exposures gives several $(I_{p,i}, t_i)$ pairs that all constrain the *same* radiance L_p , and that over-determination pins down the curve. Here aperture and calibrated gain are fixed, so k_i is the shutter time t_i times a common scale, which is absorbed into L_p . Working in the log domain, each observation says

$$g(I_{p,i}) = \ln L_p + \ln t_i. \quad (11)$$

The unknowns are the curve $g(\cdot)$, one value per possible pixel level, say 256 of them, and one $\ln L_p$ per pixel. Stack one such equation per pixel per exposure and you have a linear least-squares system in g and the $\ln L_p$'s, solved jointly, exactly the kind of inverse problem of [Chapter 6.7, Image formation as a linear system](#). Two riders make it well-behaved: a smoothness penalty on g'' (a real response curve is smooth, not jagged), and the same reliability weight $w(I)$ we will use in the merge, down-weighting samples near the clip and near the noise floor. The recovered curve is fixed only up to one additive constant, again the global scale ([Figure 17.2.2](#)).

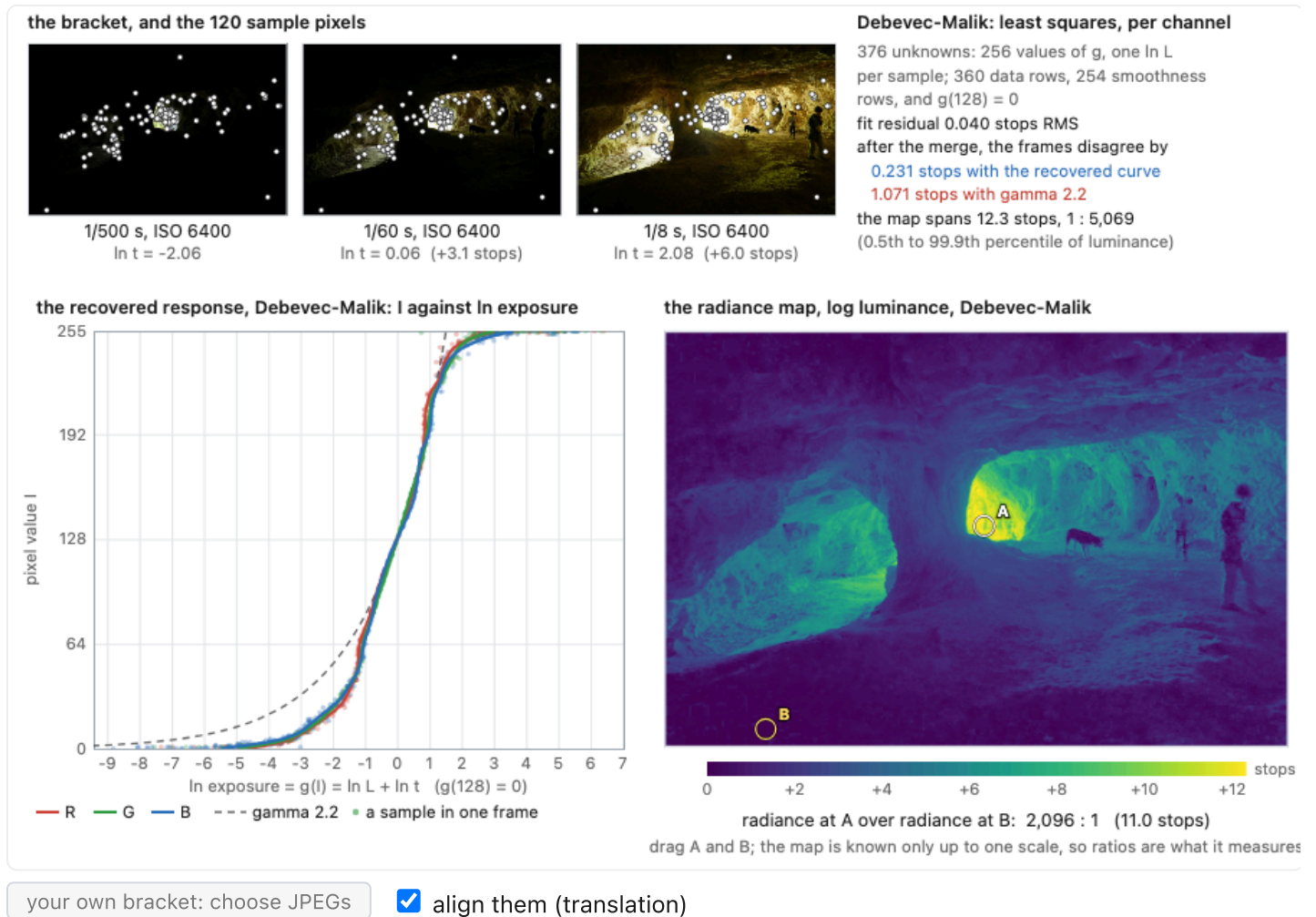
With sample pixel p and frame i , one explicit fit is

$$\min_{g, \ell_p} \sum_{p,i} w(I_{p,i})^2 [g(I_{p,i}) - \ell_p - \ln t_i]^2 + \lambda \sum_I w(I)^2 [\Delta^2 g(I)]^2, \quad g(I_{\text{mid}}) = 0. \quad (12)$$

Here $\ell_p = \ln L_p$, the second sum runs over the possible pixel values I , Δ^2 is a discrete second difference, and I_{mid} is a mid-range value whose g is fixed. Fixing one curve value removes the additive-scale ambiguity. The written

weights multiply residuals before squaring; using weights outside the square is a different convention. Smoothness alone does not enforce a monotone response. Check the recovered curve and held-out exposure residuals.

The system is small. Each ℓ_p appears only in its own sample's rows, so it can be eliminated exactly, leaving a symmetric 255 by 255 system for g per channel that solves in milliseconds ([Figure 17.2.1](#)). Run it on camera JPEGs and try three things. Set the weight to flat, and the clipped samples at 0 and 255 bend the ends of the curve. Lower λ and the curve follows the noise; raise it and it can no longer turn at the toe. Merge with an assumed gamma of 2.2 instead of the recovered curve, and the frames of a bracket disagree about a pixel's radiance by about a stop, against a few tenths of a stop with the curve.



Two or more shots of one still scene at different exposures. Exposure times come from EXIF; edit them if missing.

Figure 17.2.1 (reproduced). Debevec and Malik, and Mitsunaga and Nayar, calibration on camera JPEGs (interactive). Top: a bracket of three JPEGs with shutter time and ISO from EXIF, and the sample pixels, drawn at random or one per band of intensity from flat neighborhoods. Left: the recovered log-inverse response $g(I)$ for R, G and B, solved per channel by least squares with the smoothness weight λ and the weighting $w(I)$ chosen below the figure; each sample's observations sit at $\ln L_p + \ln t_i$, and the dashed curve is a gamma of 2.2 anchored at the same code value. Right: the bracket merged with the recovered curve, or with the gamma for comparison, into a radiance map shown as false-color log luminance or tone-mapped with global Reinhard; drag the probes A and B to read a radiance ratio. The readout gives the fit residual and how far the frames still disagree after the merge. The method control switches to Mitsunaga and Nayar's self-calibration [Mitsunaga & Nayar 1999](#): a polynomial of degree D (set by hand, or chosen on held-out pixels) fitted together with the exposure ratios, which can be started off EXIF; the panel at top right then plots the ratios and the error at each iteration, and the curve is drawn as $\ln f^{-1}(I/255) - \ln f^{-1}(128/255)$ so it shares the axes of g : the unknown scale of f^{-1} , which $f^{-1}(1) = 1$ fixes, is the same unknown as the constant in g , which $g(128) = 0$ fixes. Either method's curve can be overlaid on the other's. Where the camera changed ISO within a bracket, t_i is shutter time times ISO/100. You can load your own bracket; exposure times are read from EXIF and can be edited. Photographs by Frédo Durand, the camera's own JPEGs (Sony $\alpha 1$ II and $\alpha 7R$ V).

[Mitsunaga and Nayar \(1999\)](#) assume even less. They model the response as a low-order polynomial and recover it, together with the unknown exposure ratios, by radiometric self-calibration, so you need not even trust the EXIF exposure ratios. The spirit is identical: pixels seen at multiple exposures over-determine the curve, so fit a smooth response to them. ([Mann & Picard 1995](#)'s "comparametric," or Wyckoff, work is the earlier multiple-exposure precursor to both.)

The figure runs their method as a second option on the same samples. Their model is the inverse response itself rather than its logarithm g : a polynomial of degree D , $f^{-1}(I) = \sum_{n=0}^D c_n I^n$, on pixel values normalized to $[0, 1]$, with $f^{-1}(1) = 1$. Two steps alternate: with the exposure ratios k_i/k_{i+1} of consecutive frames fixed, every pixel seen in both gives a linear equation $f^{-1}(I_{p,i}) = (k_i/k_{i+1}) f^{-1}(I_{p,i+1})$ in the c_n , solved by least squares; with the curve fixed, each ratio is re-estimated as the average of $f^{-1}(I_{p,i})/f^{-1}(I_{p,i+1})$. Start the ratios a stop off EXIF, alternately up and down, and the iteration brings them back to within a few tenths of a stop. Start them all off the same way and it cannot. If f^{-1} and the ratios k_i/k_{i+1} explain the images, so do $(f^{-1})^\gamma$ and $(k_i/k_{i+1})^\gamma$ for any exponent γ , so the pictures fix the ratios only relative to one another, as Grossberg and Nayar showed [Grossberg & Nayar 2003](#), and the polynomial model pins γ only weakly. Run exactly as published on these JPEGs, the iteration drifts along that family, so the figure holds the bracket's total span in stops at its starting value, which removes that one direction and nothing else.

Joint fitting uses many observations while imposing a smooth response. It still requires sufficient exposure overlap and a suitable observation model. Debevec–Malik's conventional intensity weights are reliability heuristics, not automatically the inverse variances of log-transformed noise.

The log-domain estimate in Debevec–Malik is a weighted average of log exposures. Exponentiating it gives a geometric mean of radiance estimates, which differs from a linear-light arithmetic mean. Both can be useful under their respective error models; neither justifies averaging gamma-encoded values as radiance.

Debevec–Malik: recover the response and its log-inverse from multiply-exposed pixels

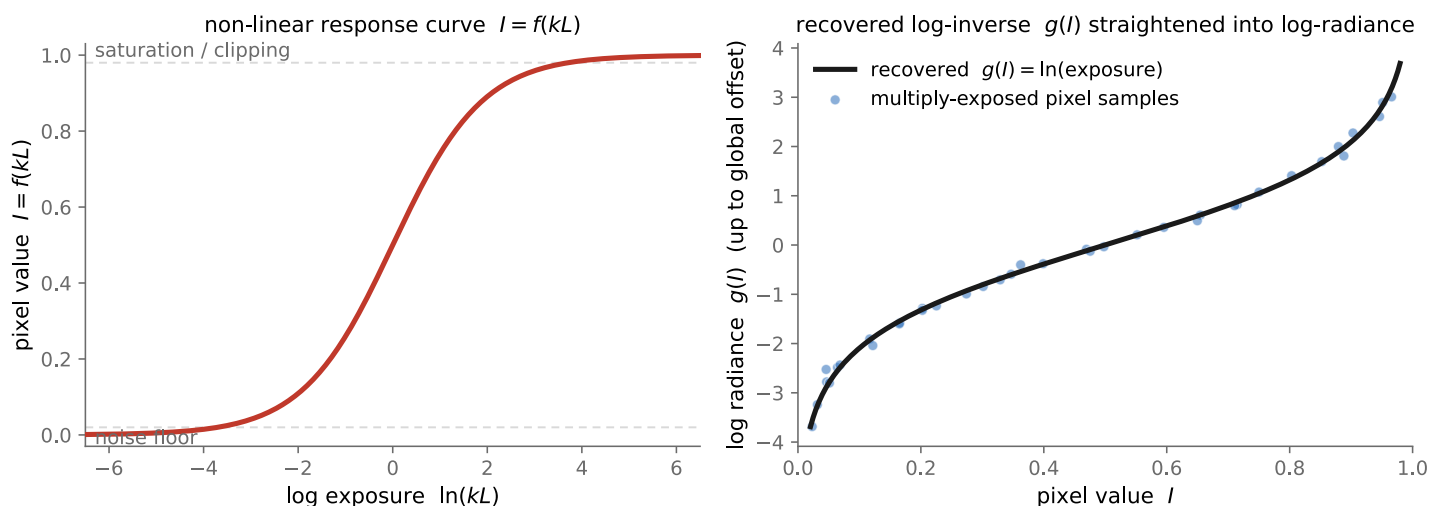


Figure 17.2.2 (reproduced). *The recovered camera response (Debevec–Malik). Left: the camera's non-linear response curve $I = f(kL)$ (pixel value against log exposure), an S-shaped film/JPEG curve. Right: its log-inverse $g(I) = \ln(kL)$, recovered by least squares from pixels seen at many exposures and straightened into log-radiance. Faint clusters of sample points (one cluster per pixel, spread across exposures and shifted by $\ln t_i$) are shown pinning the curve down, up to one global vertical offset; a smoothness penalty keeps g from wiggling between samples.*

The pipeline end to end

OPTIONAL

in the book [↗](#)

With calibration and combination both in hand, run the whole pipeline end to end, once, on one bracket: alignment, response recovery, merge, and (a preview of) tone mapping in a single interactive ([Figure 17.2.3](#)).

The HDR pipeline, end to end

SCENE Memorial church upload bracket

BRACKET shots 5 EV step 2.0

CAMERA response sRGB noise 1.00 ☒ hand-held jitter

1 · THE BRACKET

Each frame captures the same scene through a different exposure and the camera's non-linear response; none holds the whole range, and (with jitter on) each is shifted a little.



2 · ALIGN — WARD'S MEDIAN-THRESHOLD BITMAP

Threshold each frame at its **median** brightness to get a 1-bit image that barely changes with exposure, then slide it over the reference (coarse-to-fine) to find the integer shift that best lines them up — robust to the huge brightness differences a bracket has.

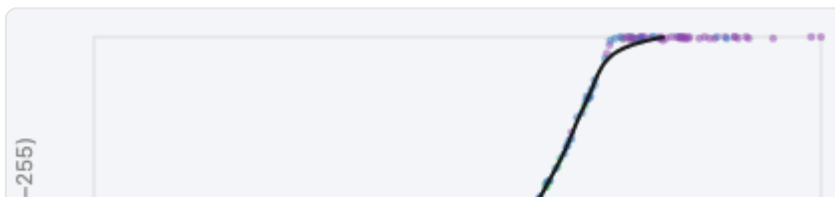
median-threshold bitmaps

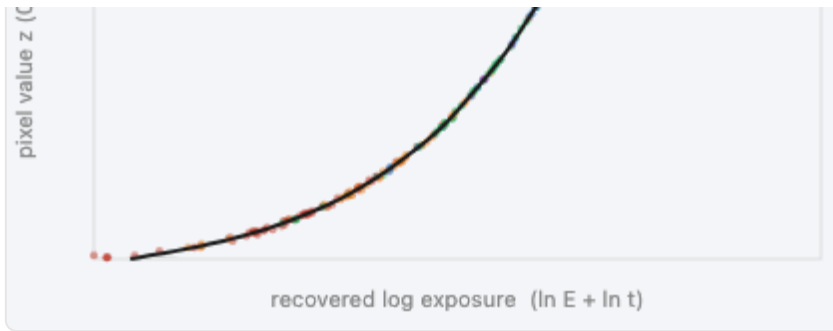


Recovered the integer shift on **4 / 4** non-reference frames (exact match to the injected jitter). The median-bitmap barely changes with exposure, so a bright and a dark frame still align.

3 · RECOVER THE CAMERA RESPONSE CURVE

A pixel's value is **response(exposure)**, and the response is unknown and non-linear. Sample pixels across all frames and solve (Debevec–Malik) for the curve **g** that makes every pixel's recovered log-exposure consistent across frames. The scatter shows each sample's value against its log exposure; the curve threads through it.

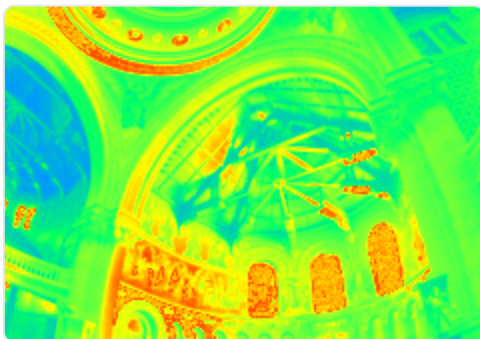




Each dot is one sampled pixel in one frame, placed at its value (vertical) and its recovered log exposure (horizontal); the dots from all exposures fall along **one** curve — the camera response **g** (black). Frames are colour-coded; they overlap into a single curve exactly when the recovered response is consistent. (Built-in scenes apply a known srgb response, which the regression reconstructs.)

4 • MERGE INTO A RADIANCE MAP

With the response known, convert every frame back to radiance ($g(\text{value}) - \log t$) and take a confidence-weighted average. The result spans the scene's full range — shown here in **false colour on a log scale** (blue = dark, red = bright).



recovered radiance (log false-colour)

Merged radiance spans about **9.9 stops** — far more than any single 8-bit frame. The false-colour map reads the log luminance: deep blue in the nave/shadows, red in the windows/sky. This radiance map is the input to tone mapping.

5 • TONE-MAP FOR THE DISPLAY

Squeeze that range onto a screen. **Linear** just clips; **Reinhard** rolls highlights off globally; **bilateral-grid** and **fast local Laplacian** are *local* operators that compress range while preserving local contrast (the difference shows in the deep shadows and bright windows).

OPERATOR **linear (clip)** Reinhard global bilateral grid fast local Laplacian

exposure key 0.30 strength 0.60





fast local Laplacian

Fast local Laplacian (Paris/Aubry): manipulate detail per-scale on a Laplacian pyramid, discretized over intensity for speed. The strongest local-contrast preservation of the four — the deep shadows open up without halos.

Figure 17.2.3 (reproduced). *The HDR pipeline, end to end (interactive). A hand-held bracket is taken through every classical stage on one screen: (1) the differently-exposed frames, each shifted a little by hand tremor; (2) Ward's median-threshold-bitmap alignment, which thresholds each frame at its median into an exposure-invariant 1-bit image and slides it coarse-to-fine to recover the integer shift; (3) Debevec–Malik response recovery, with the full scatter of (value, log-exposure) samples from every frame falling onto the single recovered curve g ; (4) the merge into a radiance map, shown in false color on a log scale; and (5) tone mapping to the display, switchable between linear (clip), Reinhard global [Reinhard et al. 2002](#), bilateral-grid [Chen et al. 2007](#), and fast local Laplacian [Aubry et al. 2014](#) so the global-vs-local difference is visible in the windows and shadows. Built-in scenes are real radiance maps (Debevec's memorial and vine sunset) from which a realistic bracket is synthesized; you can also upload your own exposures. The tone-mapping stage is a preview; its operators are developed in [Part 18, Tone mapping and display](#) and [Chapter 10.1, Bilateral filtering](#).*

Merging exposures

[in the book](#) ➤

Calibration puts the exposures into common units. The next decision is how much to trust each estimate at each location. We start with a simple method based on thresholds. But we then observe that long exposures can supply precise shadows and clipped highlights at the same time, so a more advanced merge must combine both measurement values and their reliability.

The weighted merge

[in the book](#) ➤

We work with registered, black-level-corrected linear values I_i and calibrated exposure factors k_i . We divide by k_i to put the observations in common relative-radiance units, then combine the valid estimates. For encoded inputs, we first invert the appropriate response.

Validity masks reject saturated, defective or unregistered samples. Noise weights reduce the influence of uncertain measurements. We choose clipping margins in the input's calibrated units; thresholds such as 0.99 or 0.002 are not universal. Very dark samples can still provide information when their noise is modeled. We avoid selecting on noisy intensity more aggressively than necessary, since this can bias the merge.

We let w_i include both validity and reliability. The linear-light merge is

$$\hat{L}(x, y) = \frac{\sum_i w_i(x, y) I_i(x, y) / k_i}{\sum_i w_i(x, y)}. \quad (13)$$

In words, *the estimated radiance is the reliability-weighted mean of the rescaled observations*. In code this is one pass per pixel over the bracket: divide out each exposure factor, accumulate by reliability, normalize:

Algorithm 1

```
for each pixel p:
    num, den = 0, 0
    for each exposure i:
        if valid(i, p):
            q = linear_value(i, p) / exposure_factor[i]
            w = radiance_weight(i, p)
            num += w * q
            den += w
    if den > 0:
        L[p], measured[p] = num / den, True
    else:
        L[p], measured[p] = flagged_fallback(p), False
```

For a calibrated nonlinear response, Debevec–Malik [Debevec & Malik 1997](#) instead combines log-radiance estimates:

$$\ln \hat{L}(x, y) = \frac{\sum_i w(I_i) (g(I_i) - \ln t_i)}{\sum_i w(I_i)}, \quad (14)$$

Exponentiation produces a weighted geometric mean, not the preceding arithmetic mean. The error model and weighting should match the domain. Multiple independent, equally precise observations reduce standard deviation by $1/\sqrt{N}$; unequal weights, correlations and data-dependent selection require the more general variance calculation ([Chapter 13.2, Denoising by averaging](#)).

When every exposure is bad at a pixel

[in the book](#) ↗

If every exposure clips, the shortest exposure supplies the strongest lower bound on radiance. If all are too noisy, the longest usable exposure is usually the best fallback at fixed gain, provided motion and defects do not invalidate

it. We keep an uncertainty or validity flag, since neither case supplies a reliable point estimate. With variable gain, we select by calibrated radiance-domain uncertainty rather than by exposure duration alone.

[Chapter 18.4, Exposure fusion](#) directly blends rendered inputs using appearance weights. It is useful when a finished picture is the objective; its output is not a calibrated radiance measurement.

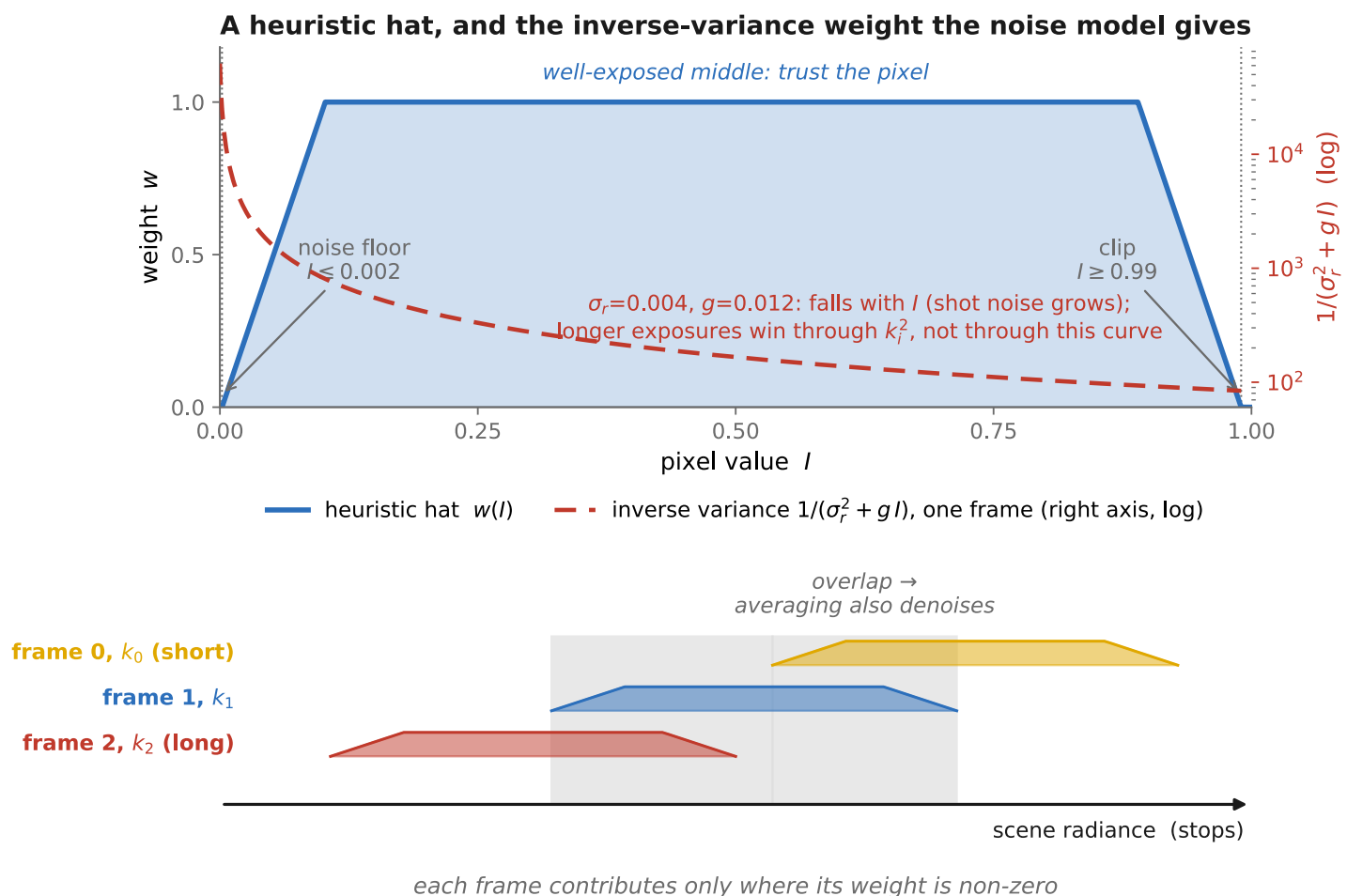


Figure 17.3.1 (reproduced). The reliability weight $w(I)$ as a function of pixel value. The weight is zero at the clipped end (e.g. $I \geq 0.99$) and zero at the dark, noisy end (e.g. $I \leq 0.002$), rising to a plateau in the well-exposed middle, a hat shape. A dashed overlay shows the SNR-optimal version of the weight, skewed toward the brighter, less-noisy samples rather than symmetric. Below, a sketch of three bracket frames shows each contributing only over the band where its weight is non-zero, with overlap regions (where averaging also denoises) shaded.

HDR merge: each pixel is a weighted average of the exposures in linear radiance (weights favour well-exposed mid-tones), then tone-mapped for display

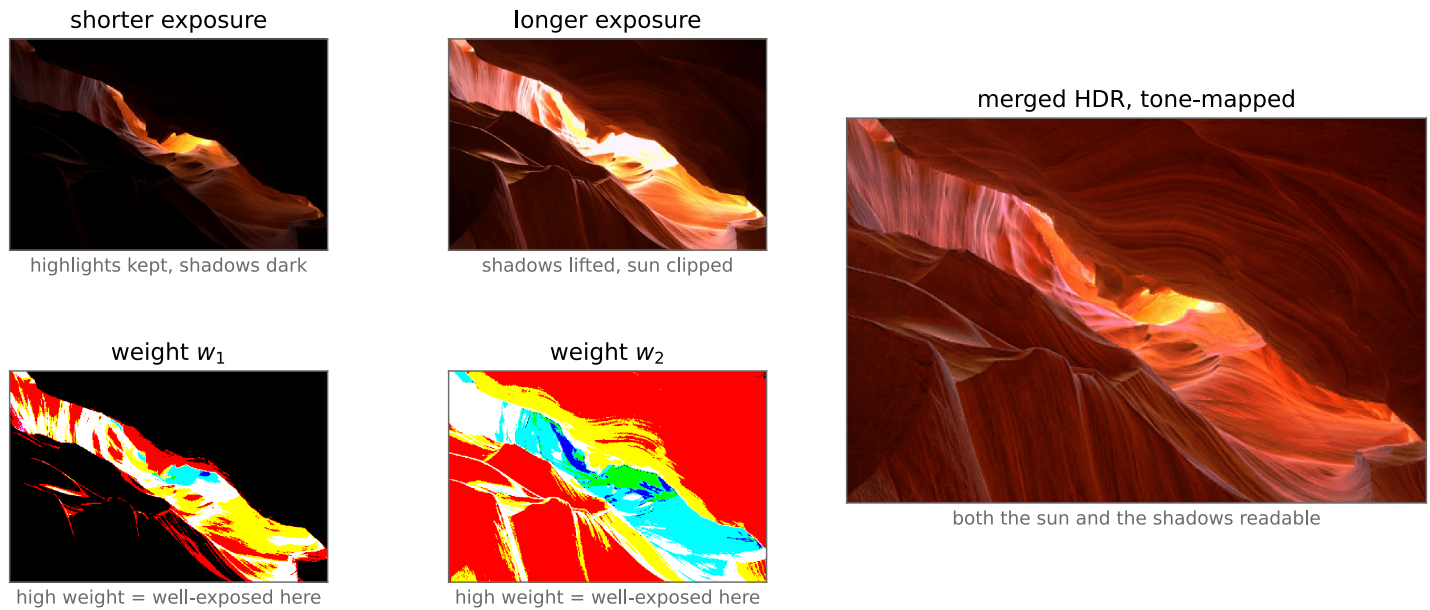


Figure 17.3.2 (reproduced). The merge on a real bracket (Antelope Canyon). A shorter exposure keeps the bright sunlit slot but buries the rock in noise-dark shadow; a longer exposure lifts the shadows but clips the sun. Each frame's weight map (rainbow, bright means high weight) marks where it is well-exposed, neither clipped nor noisy. The reliability-weighted average in linear radiance keeps the good band from each, and the float radiance map is then tone-mapped for display, so both the sun and the deep shadows read at once.

HDR+ and burst imaging on phones

Why a phone shoots a burst, and why it underexposes

OPTIONAL
in the book [➤](#)

A handheld camera faces both limited light collection and motion. Longer exposures improve photon statistics but can blur, while shorter exposures require more frames to collect the same total light and pay read noise repeatedly. Capture policy therefore depends on scene motion, illumination and sensor characteristics.

HDR+ uses equal-exposure raw frames to reduce noise while preserving the highlight range of its chosen short exposure. It improves usable shadow range; it does not extend the saturation threshold by measuring brighter highlights at a second exposure. This differs from a conventional bracket, whose frames cover different radiance intervals.

An exposure chosen to protect highlights may leave noisy shadows. Independent repeated measurements can reduce that uncertainty, whereas a highlight clipped in every frame remains only bounded from below. In a normalized noiseless model, protecting a chosen maximum radiance requires

$$k L_{\max} < 1, \quad (19)$$

Noise and uncertain metering require a margin; transient or deliberately unprotected highlights may still clip. Extra frames do not guarantee usable shadows under a fixed time budget. Choose which highlights matter rather than imposing an unconditional no-clipping rule.

💡 **BIG LESSON ((L13.1) · CAPTURE THE FULL SET, DECIDE LATER)**

The burst is [L13.1](#) [show](#) on a phone: capture many identical frames now and defer the noise-versus-range tradeoff to a downstream merge. ([L13.1](#) [show](#) first appears in this part's introduction; the light-field/plenoptic camera ([Part 33, Computational optics and coded imaging](#)) is the same idea for the focus/viewpoint axis, the special case.)

Noise-optimal bracket design ([Hasinoff et al. 2010](#)) and HDR+ both account for noise, but they solve different capture problems. A uniform short burst is not the universal optimum of the earlier shutter/ISO optimization. Motion, latency and available buffering also influence the deployed design.

Bracket varies exposure across few frames; burst holds it fixed across many

classic HDR bracket

few frames · DIFFERENT exposures

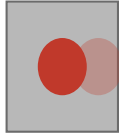
**motion blur +
ghost**



short
(highlights)



mid



long
(shadows)

needs: static scene · steady tripod

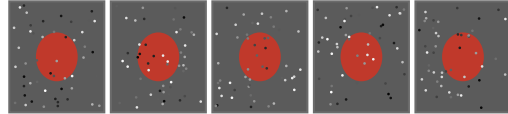
long frame blurs; subjects ghost between
the differently-exposed frames

exposure varies across FEW frames

HDR+ burst

many frames · ONE identical short exposure

N frames, every one the same → easy alignment, motion frozen



all identical · short · deliberately under-exposed

hand-held · align + merge

short-exposure noise beaten down $\sigma/\sqrt{N}$ by merging

exposure HELD FIXED across MANY frames

Figure 17.6.2 (reproduced). Burst, not bracket. Left, classic HDR bracket: a few frames at different exposures (short for highlights, long for shadows); demands a static scene and a steady mount, and the long frame blurs while moving subjects ghost between the differently-exposed frames. Right, HDR+ burst: many frames at one identical, short, deliberately under-exposed setting; hand-held, motion frozen, every frame the same so alignment is easy, and the noise of the short exposure is beaten down $1/\sqrt{N}$ by merging. The bracket varies exposure across few frames; the burst holds exposure fixed across many.

Clipped highlight: gone forever. Noisy shadow: only noisy. → bias down ($kL_{\max} < 1$), recover by computation

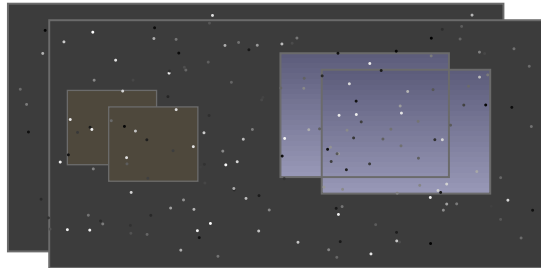
single “correctly” exposed frame



highlight clipped → flat white

detail above the full-well ceiling is
DESTROYED — unrecoverable

under-exposed burst → merge → lift



$\sigma/\sqrt{N}$

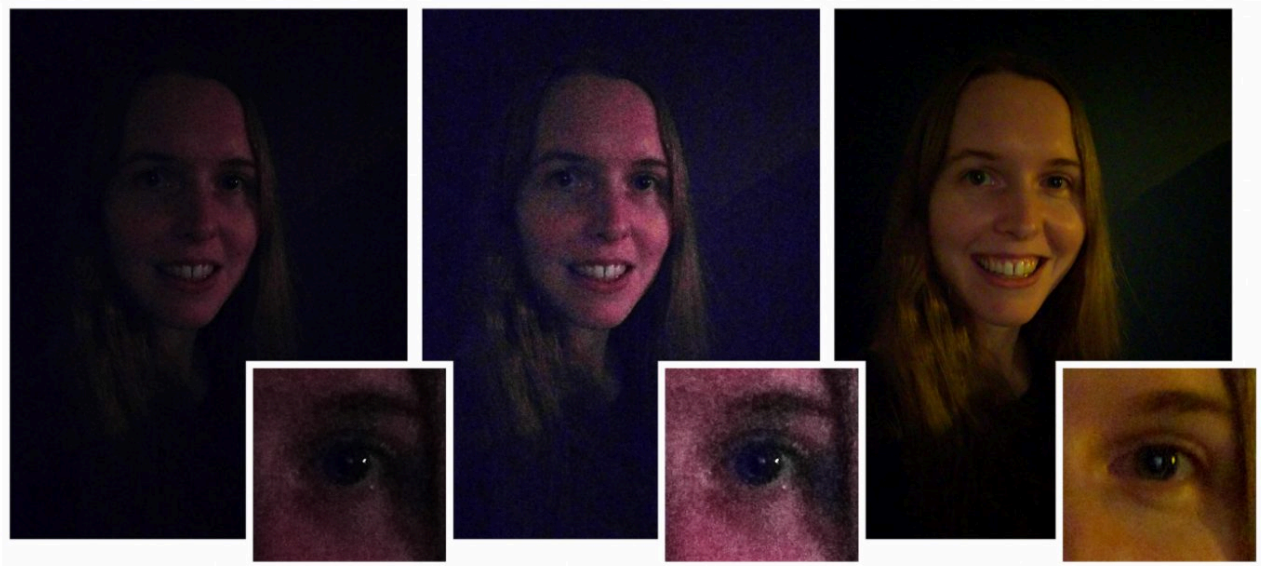
highlight detail survives

shadow noise cleaned by averaging,
then lifted — RECOVERABLE

N dark, noisy raws (window unclipped)

Figure 17.6.3 (reproduced). *The irreversibility asymmetry. Top, a single “correctly” exposed frame: the bright window clips to flat white; the detail above the full-well ceiling is destroyed and unrecoverable. Bottom, an under-exposed burst: the same window stays below clipping in every frame, so its detail survives, while the now-darker, noisier shadows are cleaned up $1/\sqrt{N}$ after the merge. A clipped highlight is gone forever; a noisy shadow is only noisy, so bias the exposure down ($kL_{\max} < 1$) and recover the shadows by computation.*

The paper's own teaser shows what the underexposed burst buys at the hardest extreme: a near-dark, hand-held portrait.



Reproduced from Hasinoff et al. (2016), with permission of S. Hasinoff.

Figure 17.6.4 (reproduced). *HDR+ in the dark. A near-black, hand-held selfie that a single short frame renders as noisy murk (left) becomes a clean, well-exposed portrait (right) after a burst of identical underexposed raws is aligned and merged; the eye crops show grain beaten down and detail restored, with no motion blur. The whole multiple-exposure part, shipped to a phone (Hasinoff et al. 2016).*

Compare a shadow patch with a lamp that is saturated in every frame. The shadow measurements can vary around a useful underlying signal, so combining independent observations can make that signal more precise. The clipped lamp values do not vary with how far the true signal exceeded the ceiling; averaging the ceiling repeatedly supplies no measurement of that excess. This asymmetry motivates reserving highlight headroom and paying some of the cost as shadow noise. It does not make arbitrarily short exposures desirable: read noise, total capture time, rejected frames, and values lost below the usable floor still constrain what the burst can recover.

The HDR+ pipeline: align and robust-merge in raw

OPTIONAL
in the book [↗](#)

Choose a reference, estimate local alignment and combine consistent raw samples. Maintain CFA phase, black-level and gain conventions. After raw merging, demosaic and render the result. The exact interpolation and alignment precision depend on the implementation; the original HDR+ and later burst-super-resolution pipelines should not be conflated.

Raw merging retains access to sensor-domain measurements and a calibrated noise model. Interpolation and merging can still introduce correlations. Joint reconstruction or a suitably modeled RGB pipeline are alternatives; pre-demosaic processing is a design choice rather than a universal optimality result.

A simplified spatial-weighted model illustrates exposure normalization, although the actual HDR+ merge operates on tile transforms:

$$\hat{L}(x, y) = \frac{\sum_i w_i I_i(W_i^{-1}x)}{k \sum_i w_i}, \quad (20)$$

HDR+ uses a transform-domain merge on aligned tiles, with shrinkage informed by estimated noise and differences from the reference. This differs from simply assigning a per-pixel inverse-variance weight. Large inconsistencies reduce temporal averaging; a reference fallback sacrifices noise reduction to limit motion artifacts. Equal averaging provides a useful independent-noise reference case:

$$\sigma_{\text{merged}} \approx \frac{\sigma_{\text{single}}}{\sqrt{N}}, \quad (21)$$

This approximation assumes equal noise and independent, correctly registered samples. Rejection, adaptive filtering and correlated errors change it.

Local alignment handles some motion that a global transform misses. Occlusions and varying exposure blur remain difficult, so the merge needs confidence and fallback even after alignment. For a merged image, inspect both detail and the spatial variation in noise rather than reporting burst length as the number of useful samples everywhere.

HDR+ pipeline — align and robust-merge in raw

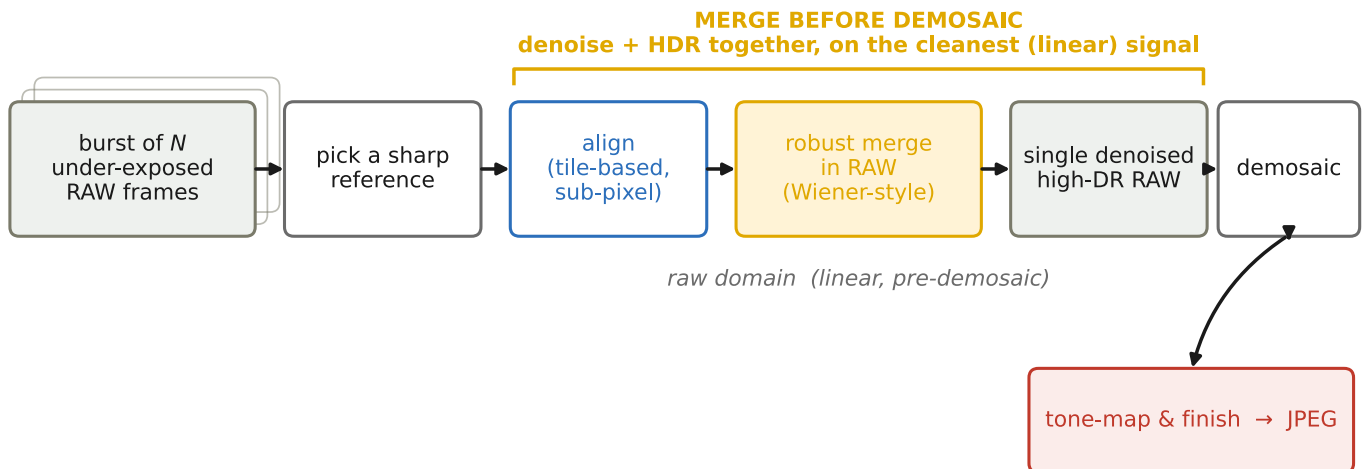
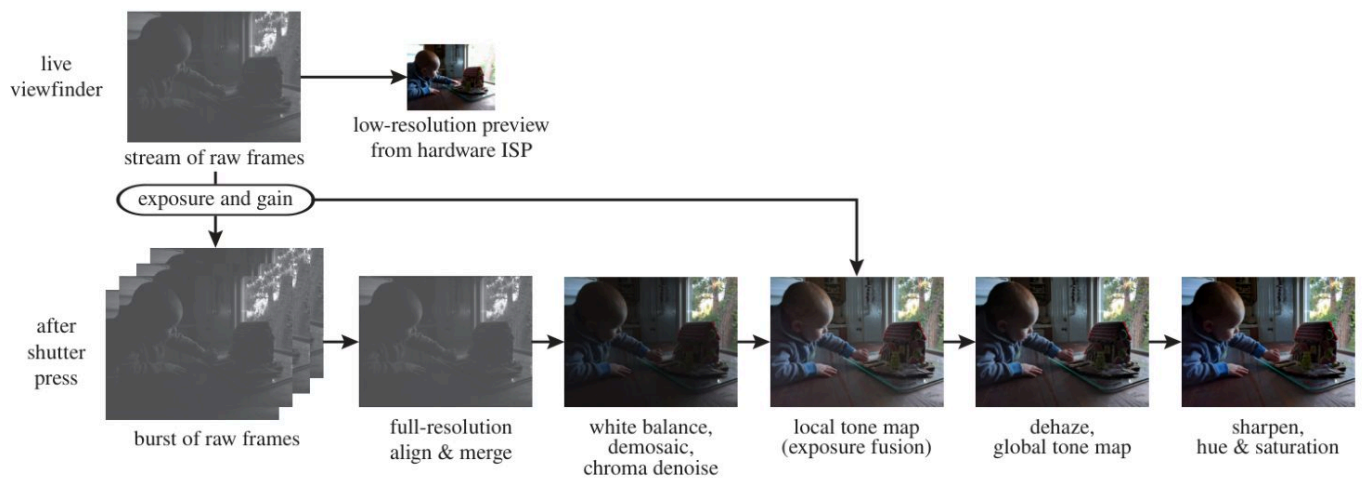


Figure 17.6.5 (reproduced). The HDR+ flow. Left to right: a burst of N identical under-exposed raws → pick a sharp reference frame → align every other frame to it (tile-based, sub-pixel) → robust merge in raw (Wiener-style; reject mis-aligned or moving tiles) → a single denoised, high-DR raw → demosaic → tone-map and finish → JPEG. The decisive choice is that the merge, which does denoising and HDR together, happens in raw, before demosaicking.

The original paper draws the same flow on a real scene, and shows both why alignment is decisive and how the finished pipeline compares to a conventional ISP.



Reproduced from Hasinoff et al. (2016), with permission of S. Hasinoff.

Figure 17.6.6 [\(reproduced\)](#). *The HDR+ pipeline as the authors drew it. From a live viewfinder fed by the hardware ISP, the shutter press grabs a burst of raw frames at one chosen exposure-and-gain; these are aligned and merged at full resolution, then run through white balance, demosaic and chroma denoise, a local tone map (exposure fusion [Mertens et al. \(exposure fusion\)](#)), dehaze and global tone map, and a final sharpen and saturation pass. The decisive step, align-and-merge, sits before demosaicking, so denoising and HDR happen together on the raw signal ([Hasinoff et al. 2016](#)).*

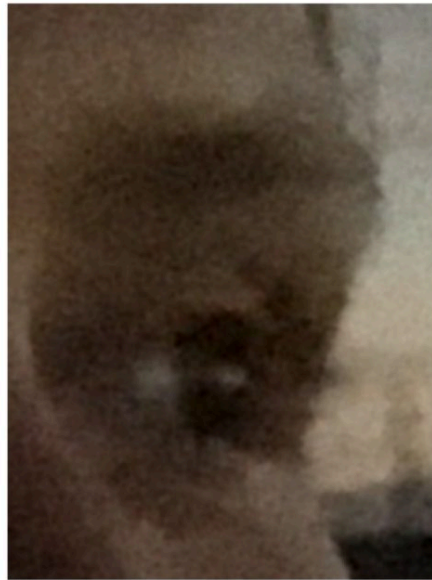


Reproduced from Hasinoff et al. (2016), with permission of S. Hasinoff.

Figure 17.6.7 (reproduced). Why alignment decides the result, on a night street. The same moving scene merged four ways: a single reference frame (noisy), a naive temporal mean (motion smears the sign and the passing vehicle into a ghost), the mean with alignment (the static sign sharpens, but residual motion still ghosts), and the robust merge with alignment (the sign is crisp and the moving vehicle is handled without a double image). Alignment plus the outlier-rejecting weight is exactly what turns a helpful average into a clean result ([Hasinoff et al. 2016](#)).



(a) *ghosting*



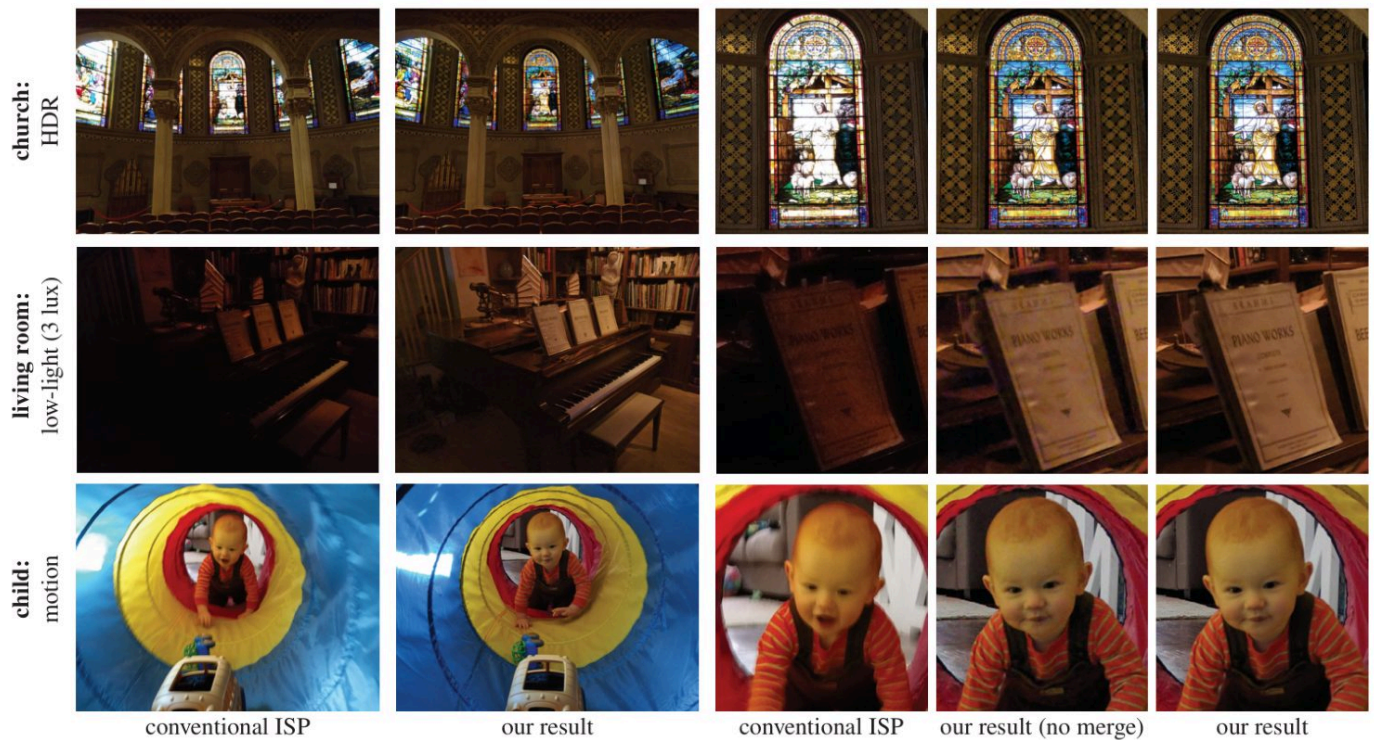
(b) *excessive denoising*



(c) *excessive range compression*

Reproduced from Hasinoff et al. (2016), with permission of S. Hasinoff.

Figure 17.6.8 [\(reproduced\)](#). Three ways a burst merge goes wrong, named. (a) *Ghosting*, a moving subject left a double image where the merge trusted misaligned frames; (b) *excessive denoising*, over-aggressive averaging smeared away real texture into a waxy blur; (c) *excessive range compression*, the tone map pushed local contrast flat and hazy. The failure gallery that the robust weight, a light hand on denoising, and a restrained tone curve each exist to avoid ([Hasinoff et al. 2016](#)).



Reproduced from Hasinoff et al. (2016), with permission of S. Hasinoff.

Figure 17.6.9 (reproduced). *Conventional ISP versus HDR+, across hard scenes. A dim church interior, a low-light living room, and a moving child are each shown as a conventional single-frame ISP result beside the HDR+ result (and, for the child, the cost of not merging). The burst pipeline holds highlights, cleans shadows, and freezes motion where the conventional pipeline leaves noise, blur, or clipping (Hasinoff et al. 2016).*

Tile alignment makes a local motion approximation. A tile on a stationary wall and a tile on a moving arm can choose different displacements, whereas one global translation must compromise between them. But a tile straddling the arm boundary contains two motions. No single displacement aligns both, and a newly revealed background patch has no corresponding foreground sample to average. The robust merge handles some of these failures by reducing inconsistent contributions. Where fewer observations survive, more noise remains. The alignment and merge therefore need to be understood together: locally different noise levels can be the visible consequence of preserving moving structure.