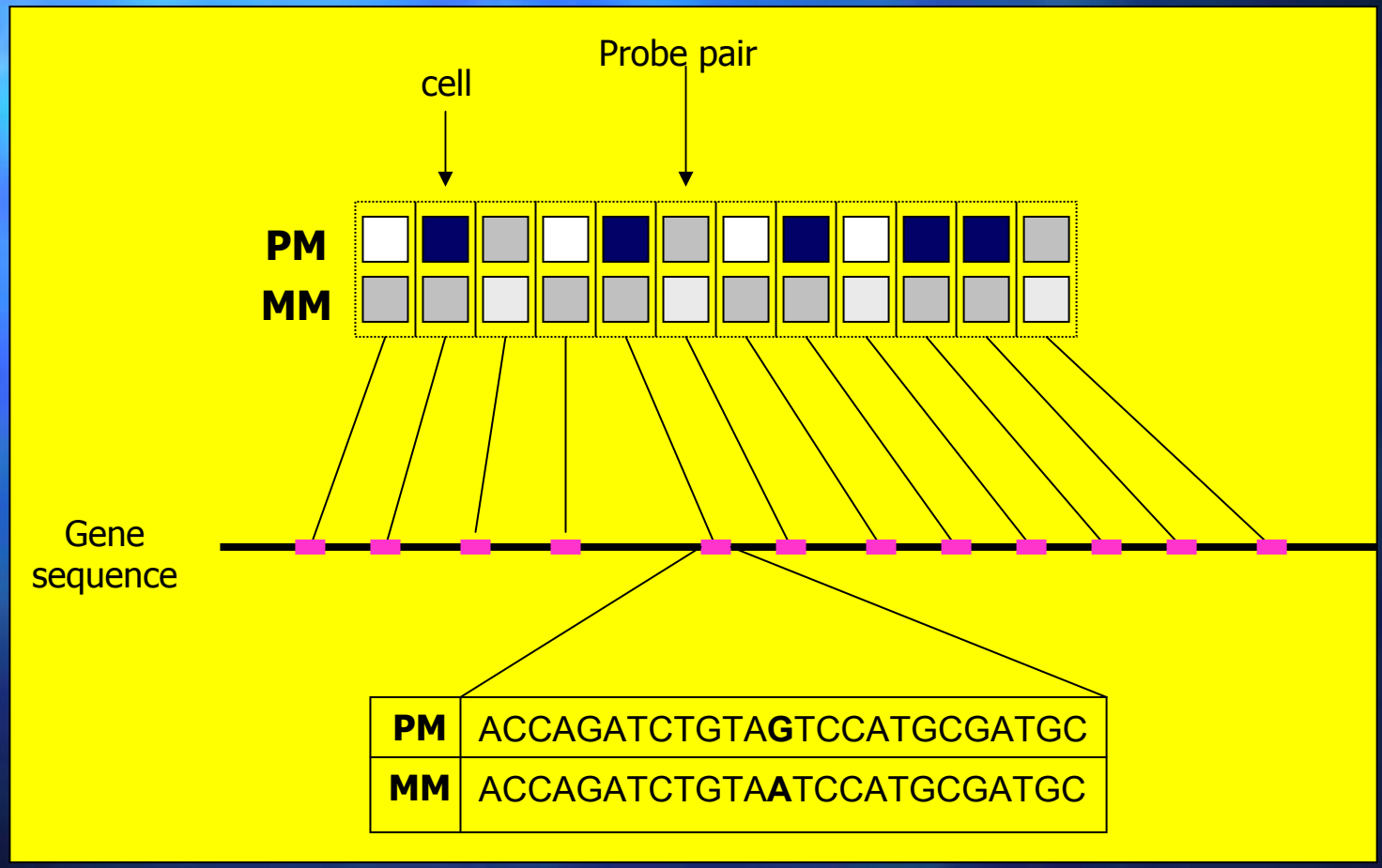
Probe set (Affymetrix)



Extracted from Affymetrix Manual

MAS 5.0 output

	Stat Pairs	Stat Pairs Used	Signal	Detection	Detection p-value
37984_s_at	16	16	92.2	P	0.000218
32102_at	16	16	59.5	P	0.000218
37900_at	16	16	72.6	P	0.000219
31697_s_at	16	16	664.2	P	0.000219
40567_at	16	16	502.3	P	0.000219
35808_at	16	16	212.6	P	0.000219
34819_at	16	16	143.0	P	0.000219
35787_at	16	16	295.7	P	0.000219
35758_at	16	16	301.0	P	0.000219
34817_s_at	16	16	339.6	P	0.000219
34644_at	16	16	723.9	P	0.000219
34608_at	16	16	3313.0	P	0.000219

Detection *p*-value which is evaluated against user-definable cut-offs to determine the **Detection** call. This call indicates whether a transcript is reliably detected (Present) or not detected (Absent).

Signal value assigns a relative measure of abundance to the transcript.

Present & absent call

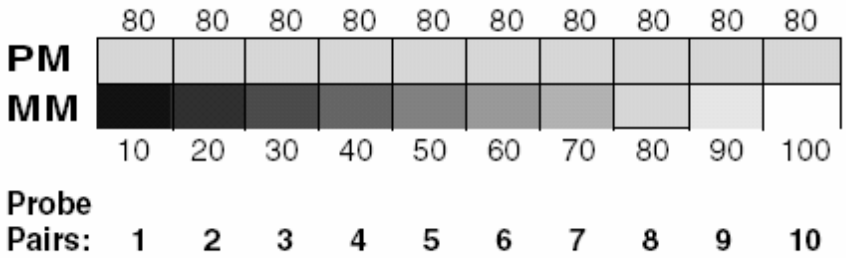
- A two-step procedure determines the *detection p-value* for a given probe set:
  - Calculate the *discrimination score* [R] for each probe pair.
  - Test the *discrimination scores* against the user-definable threshold Tau.
- The *discrimination score* describes the ability of a probe pair to detect its intended target.
  - it measures the target-specific intensity difference of the probe pair (PM-MM) relative to its overall hybridization intensity (PM+MM):

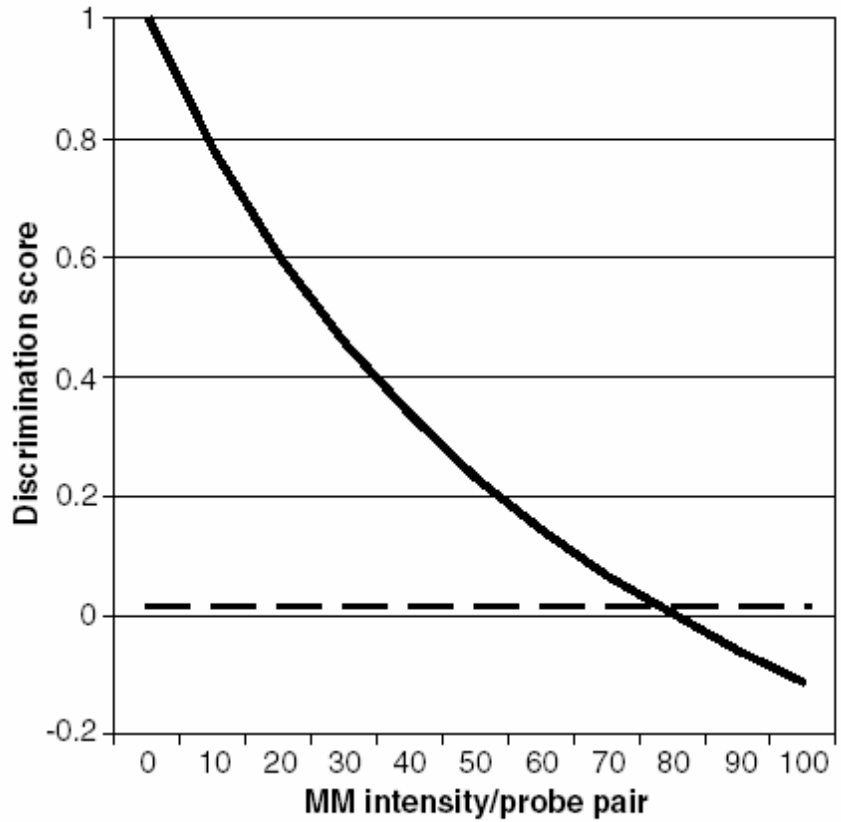
$$R = (PM - MM) / (PM + MM)$$

N.B. a probe pair is rejected from further analysis when a Mismatch (MM) probe cell is saturated

Present & absent call

- The next step toward the calculation of a Detection *p*-value is the comparison of each Discrimination score to the user-definable threshold Tau.
  - Tau is a small positive number that can be adjusted to increase or decrease sensitivity
- The One-Sided Wilcoxon's Signed Rank test is the statistical method employed to generate the *detection p-value*.
  - It assigns each probe pair a rank based on how far the probe pair *discrimination score* is from Tau.

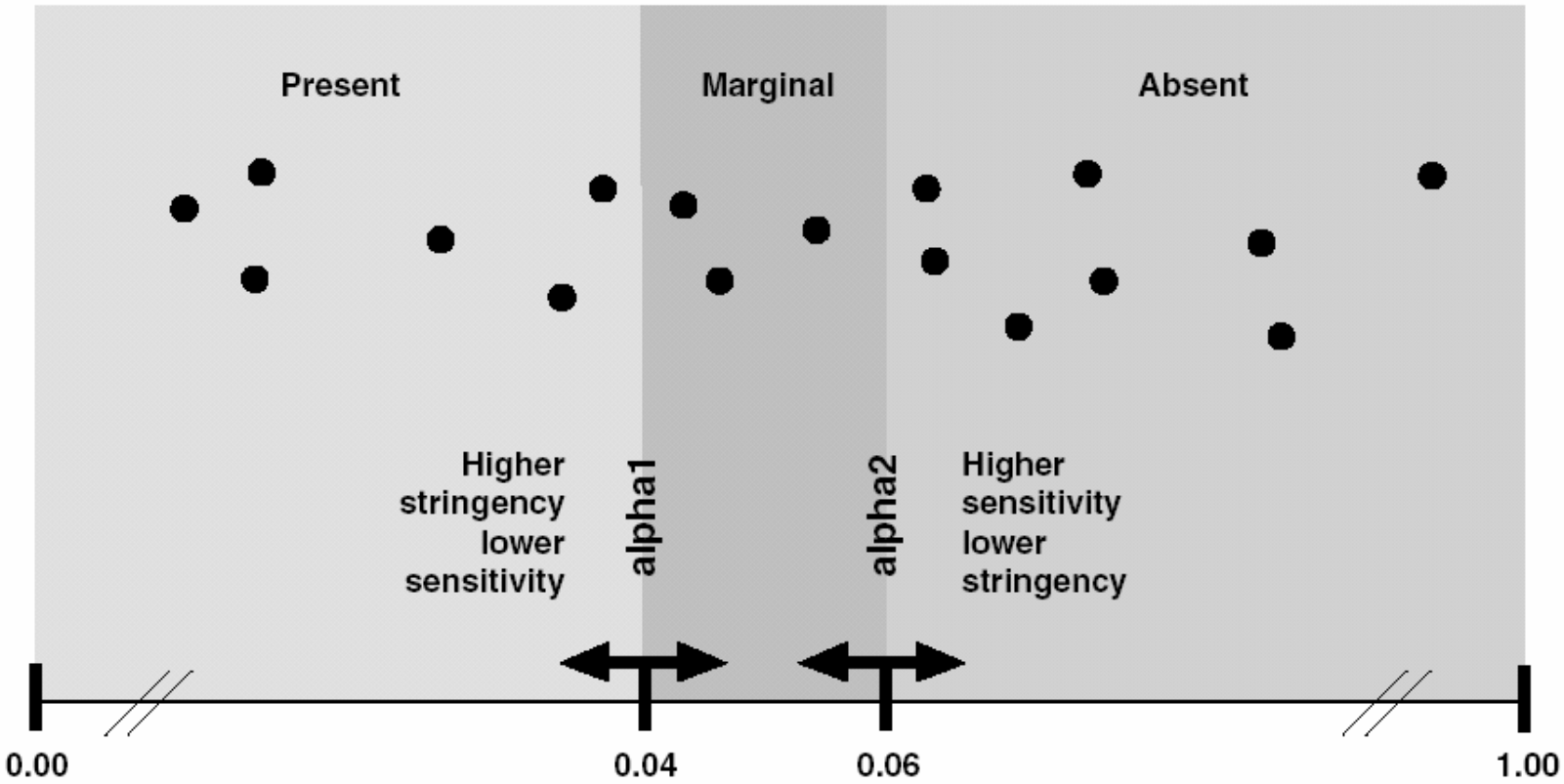




In this hypothetical probe set, the Perfect Match (PM) intensity is 80 and the Mismatch (MM) intensity for each probe pair increases from 10 to 100. The probe pairs are numbered from 1 to 10. As the Mismatch (MM) probe cell intensity, plotted on the x-axis, increases and becomes equal to or greater than the Perfect Match (PM) intensity, the Discrimination score decreases as plotted on the y-axis. More specifically, as the intensity of the Mismatch (MM) increases, our ability to discriminate between the PM and MM decreases. The dashed line is the user-definable parameter Tau (default = 0.015).

Extracted from Affymetrix Manual

Present & absent call



Significance levels α_1 and α_2 define cut-offs of *p*-values for Detection calls. Please note that these cut-offs are for probe sets with 16–20 probe pairs.

Intensity signal calculation

- Signal is a quantitative metric calculated for each probe set, which represents the relative level of expression of a transcript.
- Signal is calculated using the One-Step Tukey's Biweight Estimate which yields a robust weighted mean that is relatively insensitive to outliers, even when extreme.

Intensity signal calculation

- Each probe pair in a probe set is considered as having a potential vote in determining the Signal value.
- The vote, in this case, is defined as an estimate of the real signal due to hybridization of the target. The mismatch intensity is used to estimate stray signal.

- The real signal is estimated by taking the log of the Perfect Match intensity after subtracting the stray signal estimate.
- The probe pair vote is weighted more strongly if this probe pair Signal value is closer to the median value for a probe set.

- Once the weight of each probe pair is determined, the mean of the weighted intensity values for a probe set is identified.
- This mean value is corrected back to linear scale and is output as Signal.

Intensity signal calculation

- One-Step Tukey's Biweight (implemented in Affy):
  - tukey.biweight <- function(x, c=5, epsilon=0.0001)

```
{
    m <- median(x)
    s <- median(abs(x - m))
    u <- (x - m) / (c * s + epsilon)
    w <- rep(0, length(x))
    i <- abs(u) <= 1
    w[i] <- ((1 - u^2)^2)[i]
    t.bi <- sum(w * x) / sum(w)
    return(t.bi)
}
```

The probe pair vote is weighted more strongly if this probe pair Signal value is closer to the median value for a probe set.

Intensity signal calculation

- One-Step Tukey's Biweight (implemented in Affy):
  - tukey.biweight <- function(x, c=5, epsilon=0.0001)

```
{
  m <- median(x)
  s <- median(abs(x - m))
  u <- (x - m) / (c * s + epsilon)
  w <- rep(0, length(x))
  i <- abs(u) <= 1
  w[i] <- ((1
  t.bi <- sum(w * x) / sum(w)
  return(t.bi)
}
```

- Once the weight of each probe pair is determined, the mean of the weighted intensity values for a probe set is identified.
- This mean value is corrected back to linear scale and is output as Signal.

Intensity signal calculation

- MAS 5.0 computes probe set intensity signal as the anti-log of a robust average of the values $\log(PM_{ij} - CT_{ij})$.
  - CT is defined as a quantity equal to MM when MM<PM, but adjusted to be less than PM when MM $\geq$ PM.
- A model for MAS 5.0 probe set intensity measure is

$$\log(PM_{ij} - CT_{ij}) = \log(\theta_i) + \varepsilon_{ij}, \text{where } j = 1,\ldots,J.$$

  - The expression quantity on array i is represented with θ_i and ε_{ij} is the error term which is equal to the variance for j=1,...,J.

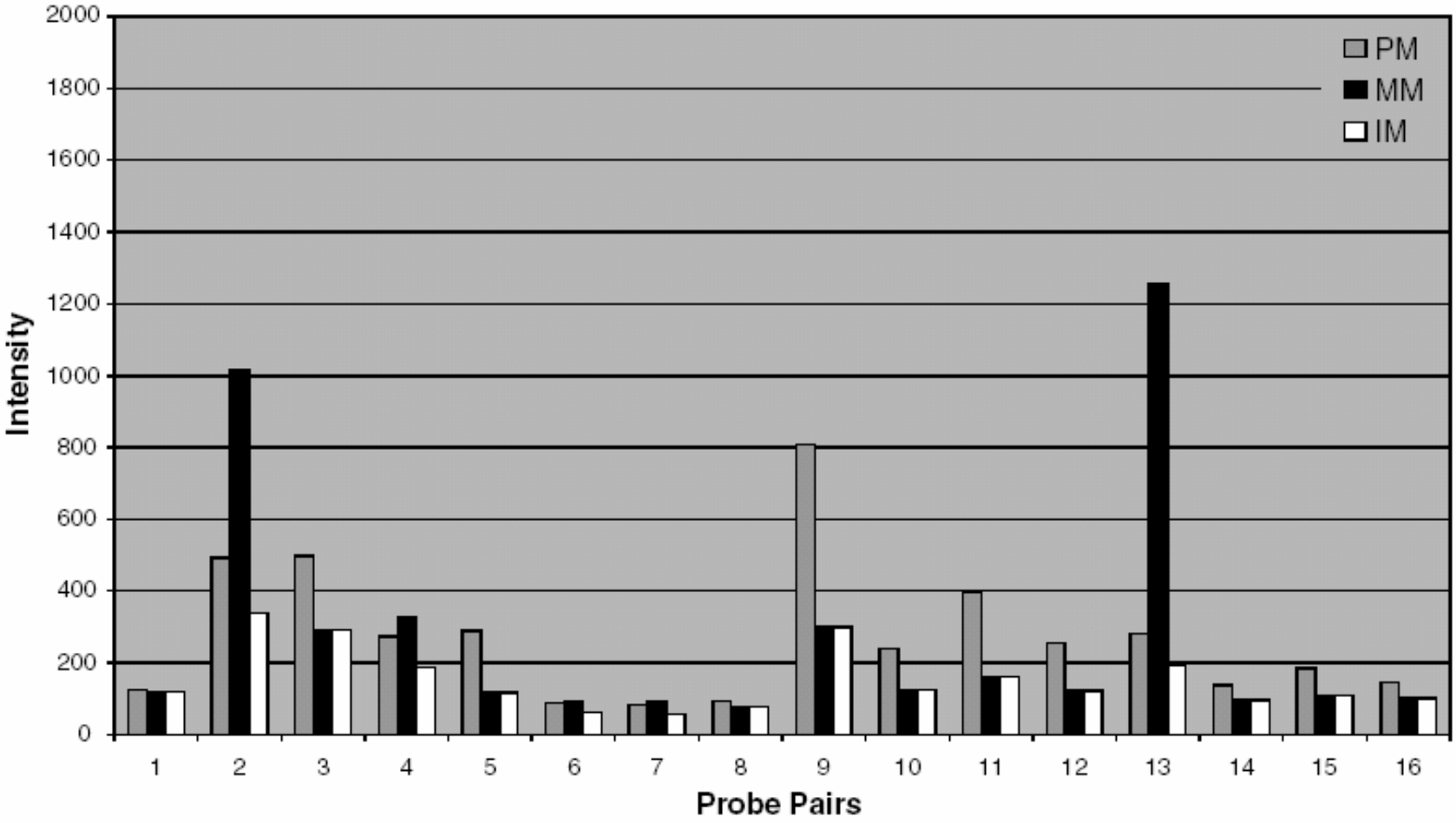
Intensity signal calculation

- When the Mismatch intensity is lower than the Perfect Match intensity, then the Mismatch is informative and provides an estimate of the stray signal.
- Rules are employed in the Signal algorithm to ensure that negative Signal values are not calculated.
  - Mismatch values can be higher than Perfect Match values
    - If the Mismatch is higher than the Perfect Match, the Mismatch provides no additional information about the estimate of stray signal.
    - Therefore, an imputed value called Idealized Mismatch (IM) is used instead of the uninformative Mismatch.

Intensity signal calculation

- **Rule 1:**
  - If the Mismatch value is less than the Perfect Match value, then the Mismatch value is considered informative and the intensity value is used directly as an estimate of stray signal.
- **Rule 2:**
  - If the Mismatch probe cells are generally informative across the probe set except for a few Mismatches, an adjusted Mismatch value is used for uninformative Mismatches based on the biweight mean of the Perfect Match and Mismatch ratio.
- **Rule 3:**
  - If the Mismatch probe cells are generally uninformative, the uninformative Mismatches are replaced with a value that is slightly smaller than the Perfect Match. These probe sets are generally called Absent by the Detection algorithm.

Extracted from Affymetrix Manual



The grey bars illustrate the Perfect Match (PM) intensities and black bars the Mismatch (MM) intensities across a 16-probe pair probe set. The white bars, Idealized Mismatch (IM), are the intensities of the Mismatch based on the Signal rules. In this example, most of the Perfect Match intensities are higher than the Mismatch intensities and therefore Mismatch values can be used directly (e.g., probe pair 9). When the Mismatch is larger than the Perfect Match (e.g., probe pairs 2, 4, and 13) the IM value is used instead of the Mismatch.

MAS 5.0 scaling

- Global scaling strategy sets the average signal intensity of the array to a target Signal of 100.
- The key assumption of the global scaling strategy is that there are few changes in gene expression between the arrays being analyzed.