ECM: Energy-Based Reconstruction of Binding Contributions in Unknown Gas Mixtures

Abstract

Most gas-mixture experiments reveal only two macroscopic quantities: the externally injected energy E_{in} and the outgoing energy E_{out}. All microscopic internal effects collapse into the measurable residual:

$$E_{\text{res}} = E_{\text{in}} - E_{\text{out}}.$$

This paper develops the Energy Composition Method (ECM), a general formalism for reconstructing or estimating intrinsic binding contributions using only this residual. Because internal structure is never directly observable, ECM classifies all possible information states (Case 0–4) and determines which parts of the composite residual are recoverable, partially identifiable, or fundamentally inseparable.

The expanded formulation demonstrates how ECM extracts the maximum possible information allowed by physics, without any molecular-level access, spectroscopy, or species identification.

1 Introduction

Determining binding contributions within a gas mixture is an indirect problem: microscopic interactions cannot be isolated experimentally. Instead they manifest only through the macroscopic energy difference

$$E_{\text{res}} = E_{\text{in}} - E_{\text{out}}.$$

However, this residual is a composite quantity originating from multiple internal sources:

$$E_{\text{res}} = E_{\text{bind}} + E_{\text{cross}} + E_{\text{loss}},$$

where E_{bind} reflects intrinsic species-level binding, E_{cross} reflects cross-species influence, and E_{loss} captures equipment dissipation.

The central challenge is identifiability: What internal quantities leave distinguishable footprints on E_{res}? Under which conditions can binding be reconstructed, and when is only partial estimation possible?

ECM establishes a complete classification (Cases 0–4) describing every logical information state and shows when reconstruction is exact, approximate, or non-unique.

2 Notation

Symbol	Meaning
E_{in}	Injected energy
E_{out}	Released energy
$E_{\text{res}} = E_{\text{in}} - E_{\text{out}}$	Energy residual
E_{bind}	Intrinsic binding contribution
E_{cross}	Cross-species influence
E_{loss}	Equipment dissipation
x_i	Mole fraction of species i
B_i	Binding coefficient of species i
n	Number of species in the mixture

3 Framework: Residual Decomposition

The central structural decomposition is:

$$E_{\text{res}} = E_{\text{bind}} + E_{\text{cross}} + E_{\text{loss}}.$$

If composition is known, binding is a linear mixture:

$$E_{\text{bind}} = \sum_{i=1}^{n} x_i B_i.$$

If composition is unknown, or if n is unknown, the problem becomes a mixture of parameter identification and inference. ECM provides systematic reconstruction rules for every information state.

4 Case-Based Reconstruction Framework

This section provides a complete, structured reconstruction logic for all information states—from fully known systems to fully unknown mixture structures. ECM treats each Case as a distinct identifiability regime. The goal is always the same:

$$E_{\text{bind}} \text{ from } E_{\text{res}} = E_{\text{in}} - E_{\text{out}}.$$

Because the observable dimension is only one scalar (E_{res}), the internal recoverability depends entirely on how much prior information is available.

The five Cases below form the full ECM reconstruction chart.

Case 0: Single-Species Mixture (Baseline Identifiable Case)

This is the only regime in which internal binding can be recovered exactly.

$$E_{\text{cross}} = 0.$$

Residual is

$$E_{\text{res}} = E_{\text{bind}} + E_{\text{loss}}.$$

If the equipment loss is calibrated:

$$E_{\text{bind}} = E_{\text{res}} - E_{\text{loss}}.$$

Why this Case is special. There are no multiplicative unknowns and no structural ambiguity. No composition, no species interaction, no hidden dimensions. Thus, ECM recovers E_{bind} uniquely.

Case 1: Multi-Species System with Known Composition

When the composition vector x_i is known, the binding term

$$E_{\text{bind}} = \sum_i x_i B_i$$

becomes directly accessible in structure. Under this condition, the residual

$$E_{\text{res}} = \sum_i x_i B_i + E_{\text{cross}} + E_{\text{loss}}$$

still contains two uncontrolled components, but the knowledge of x_i substantially reduces the overall uncertainty. Because the composition is fixed, controlled variations in external conditions (e.g., pressure, temperature, or injection energy) introduce measurable changes that help constrain or separate the remaining terms.

As a result, Case 1 provides a *more stable and better-conditioned* reconstruction setting compared to cases with unknown composition. Although the recovery is not fully exact in the strict mathematical sense, the structural information from x_i allows the binding contribution to be estimated with *moderate but dependable accuracy* in typical experimental environments.

Case 2: Species Count Known, Composition Unknown

Unknowns: - x_i (composition) - B_i (binding coefficients)

Residual:

$$E_{\text{res}}^{(k)} = \sum_{i=1}^{n} x_i B_i^{(k)} + E_{\text{cross}}^{(k)} + E_{\text{loss}}^{(k)}.$$

Why this Case is difficult. The term $\sum_i x_i B_i^{(k)}$ contains multiplicative unknowns $(x_i B_i)$. A single experiment cannot separate them.

Requirement. ECM requires a set of K independent experimental conditions:

- differing temperature, - differing input intensity, - differing process timing,

which produces variation in $B_i^{(k)}$.

Then the system becomes:

$$\mathbf{E}_{\text{res}} = \mathbf{A}(x_1, x_2, ..., x_n)^T + \varepsilon$$

with $\text{rank}(\mathbf{A}) \geq n$ required.

Outcome. Exact recovery is impossible, but ECM produces **stable estimates** for both x_i and B_i.

Case 3: Species Count Unknown

Here even the dimensionality of the system is hidden.

Residual set:

$$\{E_{\text{res}}^{(1)}, \ldots, E_{\text{res}}^{(K)}\}$$

contains patterns caused by different species.

Goal: infer n.

ECM uses structural features:

1. **Distinct response modes** of E_{res} (temperature slopes, pressure curvature)

2. **Singular value decay** (elbow detection)

3. **Stability across repeated conditions**

Once $\hat{n}$ is inferred, the system reduces to Case 2 and the usual estimation applies.

Outcome. Not exact, but **structurally consistent estimation** of x_i, B_i, and $\hat{n}$ becomes possible.

Case 4: Fully Unknown System

No species count, no composition, no coefficients, no cross/loss models.

Residual:

$$E_{\text{bind}}^{\text{est}} = E_{\text{res}} - \widehat{E}_{\text{cross}} - \widehat{E}_{\text{loss}}.$$

Estimation tools used:

- non-negativity constraints

$$E_{\text{bind}} \geq 0, \quad x_i \geq 0$$

- smoothing across conditions - physically reasonable monotonicity - consistency filters - minimizing residual mismatch

Outcome. Case 4 does not allow exact recovery, but ECM still yields the **closest possible estimate** permitted by macroscopic information theory.

Estimation Principle and Theoretical Justification

Whenever the internal degrees of freedom exceed the dimension of the observable residual, exact reconstruction becomes mathematically impossible. This follows from the basic identifiability condition of linear algebra:

$$\text{rank}(\mathcal{A}) < \text{number of unknowns} \quad \Rightarrow \quad \text{no unique solution.}$$

In these regimes (Cases 2–4), ECM does not perform heuristic guessing. Instead, it applies a constrained inverse-problem formulation that selects the physically admissible solution consistent with all observations.

The residual satisfies

$$E_{\text{res}}^{(k)} = \sum_i x_i B_i^{(k)} + E_{\text{cross}}^{(k)} + E_{\text{loss}}^{(k)}.$$

Thus the estimation step solves the constrained optimization problem:

$$\min_{x_i, B_i, E_{\text{cross}}, E_{\text{loss}}} \sum_k \left(E_{\text{res}}^{(k)} - \sum_i x_i B_i^{(k)} - E_{\text{cross}}^{(k)} - E_{\text{loss}}^{(k)} \right)^2,$$

subject to the physically required constraints:

$$x_i \geq 0, \qquad \sum_i x_i = 1, \qquad E_{\text{bind}},\, E_{\text{cross}},\, E_{\text{loss}} \geq 0.$$

These constraints ensure that ECM chooses solutions that are both mathematically consistent and physically meaningful. The method is therefore not a heuristic but a standard application of inverse-problem theory and constrained least-squares estimation.

As a result, even when exact decomposition is impossible, ECM produces the closest consistent estimate permitted by macroscopic information, achieving theoretical soundness across all partial-information regimes.

5 Discussion

It is important to emphasize that the estimation steps in ECM are not heuristic. They are dictated by the mathematical structure of the residual equation and follow standard principles from linear algebra, inverse problem theory, and constrained optimization. Whenever the number of internal unknowns exceeds the dimension of observable energy measurements, exact reconstruction becomes mathematically impossible. In such regimes, ECM performs the theoretically correct action: it selects the physically admissible solution that minimizes residual inconsistency under non-negativity and compositional constraints. Therefore, all estimation results are grounded in established identifiability theory rather than empirical guessing.

ECM provides a complete classification of what can and cannot be recovered from energy-only measurements. Reconstruction hinges not on precision,
but on **structural identifiability**. Even perfect measurement cannot reveal more information than the structure allows.

Case 0 and Case 1 permit exact or near-exact recovery. Case 2–4 allow only partial recovery, depending on dimensionality.

6 Conclusion

ECM shows that energy residuals encode significantly more information than previously recognized. The framework identifies precise reconstruction limits, clarifies when binding is recoverable, and provides estimation methods when it is not.

Future extensions include: temperature-dependent binding, nonlinear cross-interactions, and dynamic (time-resolved) reconstruction.

7 Numerical Examples (Appendix)

To illustrate how ECM behaves under different structural assumptions, we provide three numerical examples. All examples use the same baseline definitions:

$$E_{\text{res}} = E_{\text{in}} - E_{\text{out}}, \qquad E_{\text{res}} = E_{\text{bind}} + E_{\text{cross}} + E_{\text{loss}}.$$

Each example demonstrates how ECM reconstructs or estimates the binding contribution depending on the available information (Cases 0–4).

Example 1: Single-Species (Case 0)

Let the experiment inject

$$E_{\text{in}} = 12.0 \text{ J}, \qquad E_{\text{out}} = 7.6 \text{ J}.$$

Residual:

$$E_{\text{res}} = 12.0 - 7.6 = 4.4 \text{ J}.$$

If calibrated equipment loss is

$$E_{\text{loss}} = 0.9 \text{ J},$$

then the binding energy is uniquely recovered:

$$E_{\text{bind}} = E_{\text{res}} - E_{\text{loss}} = 4.4 - 0.9 = 3.5 \text{ J}.$$

No uncertainty or estimation is required because the system contains no hidden degrees of freedom.

Example 2: Two Species, Composition Known (Case 1)

Let

$$x_1 = 0.40, \qquad x_2 = 0.60.$$

Observed residual under a given condition:

$$E_{\text{res}} = 5.8 \text{ J}.$$

Assume cross and loss contributions are partially identifiable:

$$E_{\text{cross}} = 0.6, \qquad E_{\text{loss}} = 0.7.$$

Then:

$$E_{\text{bind}} = E_{\text{res}} - E_{\text{cross}} - E_{\text{loss}} = 5.8 - 0.6 - 0.7 = 4.5.$$

Binding coefficients follow:

$$x_1 B_1 + x_2 B_2 = 4.5.$$

If an additional condition provides:

$$x_1 B_1 + x_2 B_2 = 4.2,$$

the two linear constraints bracket the true value. This is an example of ECM's **structured estimation** approach.

Example 3: Unknown Species Count (Case 3)

Suppose three experiments yield:

$$E_{\text{res}}^{(1)} = 4.1, \qquad E_{\text{res}}^{(2)} = 6.8, \qquad E_{\text{res}}^{(3)} = 5.3.$$

Stacking residual responses and applying SVD:

$$\mathbf{E}_{\text{res}} = \begin{bmatrix} 4.1 \\ 6.8 \\ 5.3 \end{bmatrix} \quad \Rightarrow \quad \sigma = (10.63,\ 0.41,\ 0.02).$$

The sharp spectral drop indicates an effective **rank of 1**, implying the mixture behaves as if dominated by a **single effective binding mode**, even if multiple species exist.

ECM therefore reconstructs:

$$\widehat{E}_{\text{bind}} \approx E_{\text{res}} - (\widehat{E}_{\text{cross}} + \widehat{E}_{\text{loss}})$$

with rank-1 consistency constraints.

This demonstrates how ECM extracts structure even in fully unknown systems.

8 Advantages of ECM

ECM offers several advantages over conventional mixture-analysis approaches, especially in systems where microscopic or molecular-level information is inaccessible.

1. Works with Only Macroscopic Energies

Most traditional mixture or binding-analysis methods require:

- spectroscopy,
- concentration measurement,
- species identification,
- direct mechanistic observability.

ECM requires only

$$(E_{\text{in}},\, E_{\text{out}})$$

making it applicable even in highly opaque, unknown, or complex systems.

2. Handles Unknown Numbers of Species

Many frameworks **assume** the number of species is known. ECM does not. It infers structure directly from:

- variation of the residual,
- rank patterns,
- spectral curvature,
- consistency across conditions.

3. Provides the Maximum Identifiable Information

ECM never overclaims: it reconstructs only what is *mathematically possible* to extract from macroscopic energy.

In underdetermined systems, it switches to:

consistent estimation (structural constraints + non-negativity)

instead of pretending to recover unobservable quantities.

4. Decomposition Separates Binding From Other Effects

Residual decomposition:

$$E_{\text{bind}},\ E_{\text{cross}},\ E_{\text{loss}}$$

makes explicit what most models hide under empirical constants.

ECM imposes:

- structural separation,
- theoretical justification,
- identifiable partitions,
- estimation under constraints.

5. First Framework to Formalize Energy-Based Reconstruction

To the best of our knowledge, ECM is the first systematic, theory-backed method to quantify binding contributions **from energy alone**, without molecular information.

This conceptual novelty is one of the strongest points of the framework.

References

[1] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*, 2nd ed., Wiley (2007).

[2] P. Atkins and J. de Paula, *Physical Chemistry*, 10th ed., Oxford University Press (2014).

[3] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed., Wiley (1985).

[4] G. Strang, *Linear Algebra and Its Applications*, 4th ed., Brooks/Cole (2006).

[5] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM Monographs on Mathematical Modeling and Computation (1998).

[6] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM (2005).

[7] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press (2003).

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press (1985).

[9] J. Nocedal and S. Wright, *Numerical Optimization*, Springer (2006).

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004).

[11] E. Kreyszig, *Advanced Engineering Mathematics*, 10th ed., Wiley (2011).

[12] A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, Prentice Hall (1997).

[13] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill (2003).