Logical Explanation of HGGA for 1D BPP

I. Fundamental Axioms & Assumptions

This derivation rests on three core hypotheses accepted as the foundation for the algorithm's design:

1. **The Building Block Hypothesis (Holland):** A Genetic Algorithm optimizes by identifying, preserving, and recombining short, low-order, high-fitness partial solutions called "schemata" or "building blocks".
2. **The Grouping Structure Assumption:** In the Bin Packing Problem (BPP), the cost function depends solely on the composition of the groups (bins). Therefore, the only meaningful building block is a specific, well-filled group (bin), regardless of its position in the chromosome.
3. **The Dominance Conjecture:** A bin that is locally optimal (maximally filled) dominates other possible bins formed from the same items. We assume that constructing a global solution requires the accumulation of these "Dominant Bins".

II. Definitions

To formalize the proof, we define the following variables:

- H **(The Schema):** A *Dominant Bin.* Defined as a specific subset of items that fills a bin capacity C efficiently (representing a "group-schema of order one").
- $m(H, t)$**:** The number of individuals in the population at generation t containing bin H.
- $f(H)$**:** The fitness contribution of bin H, defined by the algorithm's cost function as $(F_H/C)^k$ where $k = 2$.
- $\bar{f}$**:** The average fitness of the entire population.
- $P_{disruption}$**:** The probability that the reproductive operators (Crossover/Mutation) destroy the composition of bin H during transmission.

III. The Derivation

Theorem: The HGGA satisfies the conditions required for the expected count of Dominant Bins, $E[m(H, t + 1)]$, to grow exponentially over successive generations.

Step 1: Guaranteed Generation (The Existence Proof)

Standard GAs rely on random chance to generate good schemata. HGGA ensures $m(H, t) > 0$ through hybridization.

- **Mechanism:** The algorithm applies a local optimization heuristic (inspired by the Dominance Criterion) that iteratively swaps items to maximize bin fill before evaluation.
- **Implication:** The algorithm acts as a localized search engine that manufactures Dominant Bins (H), ensuring valid building blocks exist in the gene pool.

Step 2: Selection Pressure (The Growth Proof)

We examine the Selection Ratio $\frac{f(H)}{\bar{f}}$.

- **Mechanism:** The cost function raises the bin fill ratio to the power of $k = 2$.
- **Derivation:**

$$\text{If } Fill_H \approx C \implies f(H) \approx 1.0$$
$$\text{If } Fill_{avg} \ll C \implies f(avg) \ll 1.0$$

  Consequently, the ratio $\frac{f(H)}{\bar{f}}$ is significantly greater than 1.

- **Implication:** Individuals containing H are selected as parents with high probability via Tournament Selection.

Step 3: Zero-Disruption Transmission (The Survival Proof)

We examine the survival probability $(1 - P_{disruption})$.

- **Mechanism:** The HGGA alters the encoding such that **One Gene = One Bin**. The Crossover operator (BPRX) transmits randomly selected genes (whole bins) from parent to child.
- **Derivation:** Since H is treated as an atomic unit, the crossover point cannot fall "inside" the bin definitions.

$$P_{disruption}(H) \approx 0 \implies (1 - P_{disruption}) \approx 1$$

- **Implication:** Once selected, the building block H is transmitted intact to the next generation.

IV. Conclusion

Combining these steps into the Schema Theorem inequality:

$$E[m(H, t+1)] \geq \underbrace{m(H, t)}_{\substack{\text{Generated by} \\ \text{Step 1}}} \cdot \underbrace{\left[\frac{f(H)}{\bar{f}}\right]}_{\substack{\text{Step 2:} \\ \gg 1}} \cdot \underbrace{(1 - P_{disruption})}_{\substack{\text{Step 3:} \\ \approx 1}}$$

Since the Growth Factor is strictly greater than 1 and the Disruption probability is negligible, the expected number of Dominant Bins grows geometrically. The population will progressively converge toward a state composed entirely of Dominant Bins—which constitutes the global optimum.

Q.E.D.